



Working Paper No. 955

Class Size, Cognitive Abilities, Bullying, and Violent Behavior: Evidence from West Bank Schools

by

Sameh Hallaq*

Levy Economics Institute of Bard College
and
Al-Quds Bard College

May 2020

* The analysis presented in this paper is part of the project “Determinants of Cognitive Development in Deprived Environments: Evidence from the West Bank,” funded by the German Research Foundation (DFG) under grant number JU 2769/2. We are grateful to the Palestinian Ministry of Education and Higher Education, test administrators, and the students who participated in our study for their time and effort. All errors are our own.

The Levy Economics Institute Working Paper Collection presents research in progress by Levy Institute scholars and conference participants. The purpose of the series is to disseminate ideas to and elicit comments from academics and professionals.

Levy Economics Institute of Bard College, founded in 1986, is a nonprofit, nonpartisan, independently funded research organization devoted to public service. Through scholarship and economic research it generates viable, effective public policy responses to important economic problems that profoundly affect the quality of life in the United States and abroad.

Levy Economics Institute
P.O. Box 5000
Annandale-on-Hudson, NY 12504-5000
<http://www.levyinstitute.org>

Copyright © Levy Economics Institute 2020 All rights reserved

ISSN 1547-366X

ABSTRACT

This study uses rich administrative and survey data to investigate the effects of class size on students' cognitive tests as well as bullying and violent behavior. I use the maximum class size rule to create a regression discontinuity (RD) relation between cohort enrollment size and class size in the public and the United Nations Relief and Works Agency (UNRWA) school system in the West Bank. In addition, I provide evidence that there is no violation of the RD assumptions resulting from discontinuities in the relationship between enrollment and students' household background at cutoff points induced by a maximum class size rule. The main findings suggest that class size has no direct impact on students' cognitive skills except for those in grade six. However, class size reduction improves the quality of life for children by mitigating the bullying and violent behavior among pupils that may negatively affect their achievements. Finally, I point to peer relations and mental health problems as a potential mechanism through which class size affects children's self-reported bullying–victim instances and violent behavior.

KEYWORDS: Class Size; Cognitive Abilities; Bullying; West Bank

JEL CLASSIFICATIONS: I20; I12

1. INTRODUCTION

Educationalists, economists, and policymakers recognize the importance of schooling in shaping human capital. Economists emphasize several inputs in the educational production function as indicators of school quality, such as teacher qualifications and school resources. Obviously, class size is one of these indicators. The impact of class size on the outcome of a child's cognitive development is still one of the major concerns of researchers due to the adverse results found in the economics literature. Class size can have a direct (short-run) effect on students' achievement, well-being, and cognitive and noncognitive outcomes (Dee and West 2011; Jakobsson, Persson, and Svensson 2013). It might also lead to indirect (long-run) impacts on other outcomes, such as educational attainment and wages (Fredriksson, Öckert, and Oosterbeek 2012).

The observed class sizes may not be exogenous to students' outcomes, since the class size may depend on the choices made by parents or school administrators. Economists have suggested two approaches to dealing with the endogeneity problem associated with class size. The first one is a randomized experiment, in which students are randomly assigned to classes of different size to create exogeneity of the class size variation (Krueger 1999). The second approach is to implement quasiexperimental identification strategies that require the exploitation of exogenous variations in class size. These exogenous variations could come from the enrollment cohort (Angrist and Lavy 1999) or population variations in previous years (Hoxby 2000). Numerous studies have adopted the second approach, and their findings differ significantly regarding the estimates of class size effects, varying from no effects to significant and substantial ones. Policymakers who recommend class size reductions have emphasized that small classes, in general, are associated with increased student achievement, since the teacher can give more individualized attention in smaller classes. On the other hand, its opponents see class size reduction as an expensive policy, since teachers' salaries constitute the vast majority of schools' noncapital expenses and there is no guarantee of improved achievement.

Several empirical studies have highlighted the crucial role of schooling in the development of almost all cognitive abilities (Carlsson et al. 2015; Nisbett 2009). Cognitive skills, as measured by standard intelligence tests, have a remarkable impact on labor market and subsequent later-life

outcomes. However, while the literature has investigated the causal relationship between class size and students' achievements in general, policymakers are also interested in whether or not class size reduction helps to mitigate problems such as bullying and violent behavior, which are growing in schools (Eriksen, Nielsen, and Simonsen 2012; PCBS 2013).

Bullying and violent behavior among peers are observed phenomena in all cultures, and they are common in adolescent school students. These problems have costly consequences. The adverse circumstances in which education takes place in West Bank schools do not contribute to solving this problem.¹ Bullying among peers leads to psychological suffering, low self-esteem, and isolation, which are detrimental to learning and academic performance (Moura, Cruz, and Quevedo 2011). Moreover, recent studies have indicated an association between bullying and suicidality among adolescents (Abdeen et al. 2018; Sarzosa and Urzúa 2015). The effects of class size are stronger for ill-behaved children than for well-behaved children (Lazear 2001), and a common view holds that bullying problems are a consequence of large classes and schools (Smith et al. 1999).

However, previous studies measuring both the impact of schooling on cognitive abilities and the classroom behavioral problems have been conducted primarily in stable settings in Western countries. Other contexts of schooling may have different outcomes, such as schools in the West Bank. Public and United Nations Relief and Works Agency (UNRWA) schools in the West Bank represent distinct schooling contexts. For instance, UNRWA schools are designed to accommodate children from refugee camps until grade nine. In general, the class size in these schools (as well as other governmental schools) is large compared with those in Western countries. The UNRWA believes that the costs of reducing the class size exceed the benefits, while other educational policies, such as improving teacher training and the quality of education, would be a better investment (UNRWA 2014). Therefore, UNRWA schools have raised the maximum class size from 45 students per class to 50 students, while the maximum class size in governmental schools is still 40.

¹ More than 20 percent of students enrolled in West Bank schools are exposed to peers' and teachers' violent behavior inside schools (PCBS 2013).

This study attempts to investigate the effect of class size on students' standardized cognitive test scores, as well as bullying and violent behavior. In so doing, it adds to the sparse literature on class size effects from less developed countries and regions, particularly the West Bank. I employ the survey data collected in 2013 from approximately 6,000 students in grades 5–9 in West Bank schools in the context of a cooperative research agreement funded by the German Research Foundation (DFG). An important advantage of this database is that it provides a large amount of student-level observations rather than class-level observations, as well as other school and locality-level covariates. Moreover, some of these data are administrative, thereby giving rise to no or only slight measurement error in the actual average class size, enrollment, and students' age.

In this study, I use the maximum class size rule to create a regression discontinuity (RD) relation between cohort enrollment size and class size. Further, the econometric approach provides evidence for no discontinuities in the relationship between enrollment and students' household background at cutoff points induced by a maximum class size rule, which violates the RD assumptions. The results reported in this paper show that class size affects self-reported bullying–victim and violent behavior, while there is no evidence that class size affects students' cognitive abilities in primary schools, except students in grade six. Further, I point to peer relations and mental health problems, extracted from the impact supplement and follow-up questions in the Strengths and Difficulties Questionnaire (SDQ), as a potential mechanism through which class size affects children's self-reported bullying–victim instances and violent behavior.

The paper proceeds as follows. Section 2 reviews the literature on the class size effects on achievement. Section 3 describes the relevant institutional features of the Palestinian schooling system, while section 4 describes the data. Section 5 continues with the description of the identification strategy and the empirical approaches. The main results are reported in section (6); section (7) suggests the potential mechanism, while section (8) concludes the paper.

2. RELATED LITERATURE

Over more than a decade, a growing literature has attempted to estimate the link between class size and student achievements. Class size is an endogenous variable and could correlate with other unobserved factors. For example, parents from higher socioeconomic backgrounds may put their children in schools with smaller classes or school principals may put weaker students in smaller classes. Researchers have adopted a quasiexperimental approach to isolate the effect of class size from other determinants of pupil outcomes to account for the endogeneity problem in the class size effect. Despite substantial research on class size, much about this relation is still unknown. Below, I present a brief review of some of these studies according to their main conclusion.

The first stream of studies found, in general, that smaller classes are associated with increased students' achievements, usually measured by standardized tests in multiple subjects, such as mathematics and reading. Akerhielm (1995) found that class size has a significant negative impact on student achievement. However, this study was criticized, since the identification strategy that the researcher employed only addressed the potential bias due to sorting within schools and not that attributed to parental tastes for education. Krueger's (1999) study analyzed the Tennessee Student Teacher Achievement Ratio (STAR) as a form of controlled experiment and concluded that, in the first year of elementary schools, students who attended small classes (13–17 students) outperformed their classmates assigned to regular classes (22–25 students) by 4 percentile points. Therefore, the advantage of attending small classes expands by about 1 percentile point per year in subsequent years. He also found that class size has a larger effect on minority students and those enrolled in free-lunch programs.

Angrist and Lavy (1999) employed Maimonides' rule to create an exogenous variation in class size in Israeli public schools. According to this rule, if the size of an enrollment cohort in a school exceeds 40, an extra class should be created. This rule creates RDs in the relation between cohort enrollment size and class size. However, the actual class size does not exactly match that predicted by the rule, which provides a source of exogenous variation in class size. Angrist and Lavy found a significant negative class size effect on pupils' achievement in the fourth and fifth grades but not in the third grade, which indicates the issue of the cumulative nature of class size effects. Numerous

more recent studies have implemented this approach similarly in related contexts. Gary-Bobo and Mahjoub (2013) applied the maximum size rule to French junior schools and obtained a small but significant and adverse effect of class size on the probabilities of educational success in grades six and seven, but this effect seems to vanish in higher grades. Bonesrønning (2003) found significant effects in lower-secondary schools in Norway. In addition, the further examination recommended that the impact differs among student subgroups and that the advantages of small classes are larger in schools with a high proportion of pupils who come from intact families. In Denmark, Browning and Heinesen (2007) employed Maimonides' rule to investigate the impact of class size on years of education and the completion of upper-secondary school as an alternative outcome. They found a marginally significant negative effect of class size. Moreover, Bingley, Jensen, and Walker (2005) used the Danish data to compare siblings who experienced different class sizes to control for different unobserved outcomes, such as a family background effect; they found that class size reduction in junior-high school significantly increases the students' years of education, but the effect is too small to justify the public costs. In developing countries, other studies have applied the same technique and found a negative impact of class size on student outcomes. Urquiola (2006) investigated the class size effect in Bolivia by focusing on schools in a rural area with only one or two classes and found a negative effect of class size on test scores. Moshoeshe (2015) used the two-stage least-square (2SLS) and instrumental variable (IV) quantile regression methods on grade six students' math and reading test scores in Lesotho to estimate, respectively, the mean and the distributional effect of class size. He found strong evidence of a putative class size effect on reading but not on math achievement.

However, Urquiola and Verhoogen (2009) raised some concerns regarding Angrist and Lavy's regression discontinuity approach. They provided evidence that Chilean private schools, seeking to maximize their profits, manipulate enrollment to avoid adding a classroom, and households sort themselves across schools in response to this situation. Thus, their empirical evidence demonstrates that this behavior invalidates the RD design. Therefore, their data suggest that students after the cutoffs (smaller class size) consistently come from a stronger background, which means that RD estimates overestimate the effect of class size on students' achievements. These findings were extended by Danny, Gradstein, and Reuven (2009). After using the rich individual-level dataset for secondary public schools in Israel, they concluded that IV estimates of class size effect are likely to

be seriously biased. Thus, the success of the RD design hinges on the institutional details of the context in which it is applied.

The second stream of literature found that class size has either no or a positive effect on students' outcomes. Hoxby (2000) used demographic variation to identify the class size effect. She exploited the idea that, after correcting for a trend in school districts, cohort sizes can be larger or smaller in some years than others, creating a source of exogenous variation. Using data on elementary school pupils in the state of Connecticut, she did not find any statistically significant effect of class size on student achievement, and her estimates were precise enough to rule out even a modest effect. Leuven, Oosterbeek, and Rønning (2008) exploited a quasiexperimental method to estimate the effect of class size on student performance at the end of lower-secondary school in Norway. They found that the maximum class size rule and population variation identification give very similar estimates. Their results consistently point to a lack of any impact of class size on achievement; for instance, they found an effect as small as 1.5 percent of a standard deviation for a one-student change in class size during three consecutive years. Asadullah (2005) used the variation in the average pupil–teacher ratio in Bangladeshi schools as an IV. The policy there allows secondary school principals to recruit an extra teacher whenever the class size exceeds 60. He found strong positive class size effects on students' achievements. Denny and Oppedisano (2013) used the data collected in 2003 from the Program for International Student Assessment (PISA) for the United States and the United Kingdom. They applied the IV strategy and controlled for school fixed effects to address the potential bias due to between-school sorting. Their results demonstrate that an increase in class size leads to an improvement in students' mathematics score. Only the results for the United Kingdom are statistically significant. However, Moshoeshoe (2015) indicated major concern regarding the positive or lack of effect of class size on students' achievements suggested by Asadullah (2005), Denny and Oppedisano (2013), and Leuven, Oosterbeek, and Rønning (2008). In fact, there is an important common denominator in these studies in that they all use secondary school students' data, which could explain their results. These students, in general, have a longer attention span, are less disruptive in class, and can learn more on their own compared with younger primary school students. Therefore, the class size effects are likely to be swapped. Lazear (2001) explained the difficulty in finding class size effects as being due to the possibility that the optimal class size for well-behaved students may be larger than the optimal class size for ill-behaved

students. The observed relation of educational outcome to class size is small or even positive for well-behaved children, while the class size effects are larger for children with behavioral problems.

With regard to Palestinian schools, to the best of my knowledge, no particular study has been conducted to explore the effect of class size on students' achievements apart from that of Altinok and Kingdon (2012). This study used Trends in International Mathematics and Science Study (TIMSS) data in 45 countries. Its results show the statistically significant effect of class size for 16 countries, but the effect is negative only in 10 of them, including Palestine, and the effect size is slight in most cases. Their findings are in line with other studies that have found little evidence of an adverse effect of larger classes on student achievement. They concluded that the class size effects are smaller in developed countries than in developing countries. In addition, this impact is smaller in regions with higher teacher quality.

3. INSTITUTIONAL SETTINGS

Compulsory education in Palestine covers ten years, starting at age six, and it consists of two stages: the preparatory stage or lower basic level (grades 1–4), and the empowerment stage or basic level (grades 5–0). There are three supervising authorities covering education in the West Bank: the government, the UNRWA, and private schools. These types of schools differ in their individual and school characteristics. UNRWA schools are located in Palestinian refugee camps and provide primary education to refugees until grade nine. The Palestinian Ministry of Education and Higher Education (MoEHE) is the official body in charge of the educational process in schools and formulating the educational policies. Further, all schools follow the same national curriculum, provided by the MoEHE, and apply the same evaluation criteria (UNESCO 2011). More than 70 percent of Palestinian students are enrolled in governmental schools, 23 percent in UNRWA schools, and less than 3 percent in private schools (PCBS 2017). Most students attend the closest school to their residence. According to ministry regulations, registering children in schools that report to a different educational governorate from the parents' local address is not allowed. The school sample used in this study consists of both governmental and UNRWA schools. Both types lack resources, infrastructure, and adequate classroom space. Teachers suffer from overcrowded

classes, a lack of discipline, and low salaries. They also lack motivation and professional commitment, and many of them have a second job (Jabr and Cahan 2014). Each class is taught by a different teacher depending on the subject; the teacher rotates among classrooms (MoEHE 2016).²

In general, the average class size is large compared with that in other countries.³ Recent statistics show that the average class size in Palestinian elementary schools is 27.8 and 32 students per class in public and UNRWA schools, respectively (PCBS 2017). According to the MoEHE regulations, the school principal can open a new section for the same grade in the case that the number of enrolled students exceeds 40 per cohort. This rule is not strictly implemented in all public schools due to extra hiring costs and space availability. Unlike the public schools, and despite the fact that the UNRWA follows the same educational regulations, the maximum class size in UNRWA schools was 45 students and has recently been increased to 50 students. In addition, the teachers' workload in UNRWA schools is 28 classes per week, while it is 22 classes per week in governmental schools.

In September 2015, teachers in the West Bank led the largest mass protests in Palestine in recent years due to their pay scale staying the same when other governmental sectors' pay scale increased. The Palestinian Authority (PA) allocates approximately 20 percent of its annual budget to the MoEHE, with salaries representing the largest proportion of this budget (MoFP 2016). The MoEHE has suggested several solutions to improve students' performance rather than reducing the class size. At the beginning of 2014, the UNRWA announced the possibility of postponing the starting date of the first semester of the 2014/15 academic year due to the staff strike, which continued for more than two months as a result of the UNRWA's new policy of merging classes (UNRWA 2015). Many reports have documented that parents of UNRWA school students objected to this policy, which adds more students to each class. According to parents, this can weaken the quality of education. The UNRWA justified its merging policy by citing the fact that class size itself is not a key determinant of good quality. If the class size is around 40–45 students, reducing it is not

² Any academic year consists of two semesters, and marks are divided equally between the two semesters. All courses are graded based on a scale ranging from 0 to 100. The passing grade is 50. Students who obtain a score lower than 50 percent in three subjects or fewer are required to sit for a make-up exam; if they fail four subjects or more, they repeat the same grade based on the allowed repetition rate (MoEHE 2016).

³ For example, see Altinok and Kingdon (2012). They investigated the effect of class size in 45 countries. The average class size in Palestine (39.52) was among the largest sizes in their sample.

correlated with a significant improvement in students' achievements. Most importantly, a minimum reduction in class size has a substantial impact on the overall education costs (UNRWA 2014).

4. DATA

This study employs the survey data collected during the academic year 2012/13 on approximately 6,000 students enrolled in grades 5–9 in gender-segregated primary schools in the West Bank and East Jerusalem. This rich dataset was collected as part of a cooperative research agreement between Al-Quds University, Hebrew University, and the University of Wuppertal, and was funded by DFG. The survey covered 100 primary schools selected randomly and stratified according to region (north, center, and south) and school supervisory authority (60 government and 40 UNRWA). Then 12 students were selected randomly from each grade (5–9), which yielded 60 students/school. Those students were asked to complete standardized cognitive ability tests administered in class. The intelligence abilities test was designed to measure three major cognitive aspects (verbal, numerical, and figural). Each aspect had several subtests covering a wide range of item contents (analogies, series, vocabulary, and inference). The subtests were selected and adapted to the Arab culture.⁴ The tests consisted of 181 items computed as the percentage of correctly answered questions. The cognitive test score was the first outcome variable, and it has an arithmetic mean of 60.7 and a standard deviation of 16.7 score points.

Of particular importance in the context of this study are bullying and fighting inside the school (violent behavior). The second outcome—conduct problems proxied by self-reported bullying and violent behavior—may be influenced by class size. These data are drawn from the Health Behavior in School-Aged Children (HBSC) survey answered by students. According to the literature, bullying⁵ was measured by asking students two questions: 1) How often have you been bullied?; and 2) How often have you taken part in bullying another student at school in the past couple of

⁴The subtests were selected and adapted from established tests of general ability: the Cognitive Ability Test (CAT) (Thorndike, Hagen, and Lorge 1971); Milta, a Hebrew version of the Lorge–Thorndike Test (Ortar and Shachor 1980); Standard Progressive Matrices (Raven 1983); and the Cattell and Cattell (1960) Culture Fair Intelligence Test.

⁵A student is characterized as being bullied or victimized when he or she is exposed, repeatedly and over time, to negative actions on the part of one or more other students (Olweus 1997).

months? The response options ranged from never to 1–2 times, 2–3 times, once a week, and several times a week. Then, the final outcomes were converted into three binary variables: bullying others only, victim only, and bully–victim (those who are both bullies and victims). Violent behavior was measured by asking students the following question: During the past 12 months, how many times were you injured by other students and had to be treated by a doctor or nurse or paramedic? The answers to this question were provided on a scale from 0 to 4 times or more.

The dataset includes the administrative database provided by the MoEHE on student age, gender, and school grades in each subject for three subsequent academic years (2011–13). The main explanatory variable used in this study, class size, was obtained from these administrative records, which register all the schools participating in the survey and the number of classes per grade. For each student, I combine this information with the enrollment data to calculate the average class size at the grade level as:

$$\text{average class size} = \text{enrollment} / \text{no. of classes.}$$

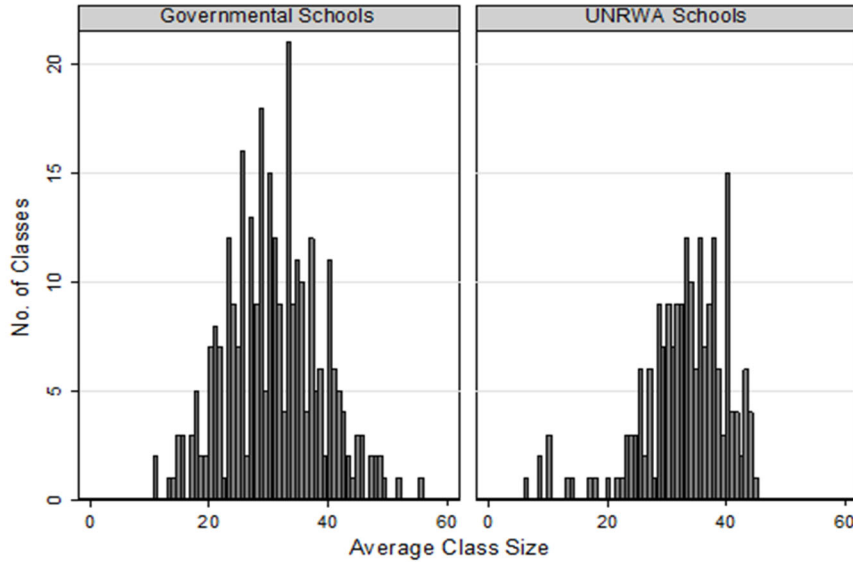
Thus, I have data on the average class size per grade and not the actual class size (except when schools have exactly one class in a grade). The advantage of the average class size is that it eliminates the biases resulting from within-school sorting, while there is a possibility of measurement error that might underestimate the impact of class size, which will be addressed in the identification strategy. Figure 1 shows the distribution of the average class size in grades 5–9 in both types of schools. The average class size in governmental schools is 30.7, while it is 33.4 in UNRWA schools.

Additionally, these administrative records allow me to include the students' grade point average (GPA) during the previous two academic years in the main analysis, which is computed as the average of six main subjects: Arabic, English, mathematics, science, national education, and religious education.

Further, this new dataset contains a large number of students' family background covariates, which are used as control variables (e.g., household structure, household monthly net income, parental

education, books at home, and living standard). The students’ parents provided these data through a paper-and-pencil questionnaire.

Figure 1: Distribution of the Average Class Size



Source: Own calculations based on survey data.

The school characteristic variables include educational facilities, adequate stationery, the availability of qualified teachers, teachers’ commitment, and the school educational climate, and were obtained from the school principal survey. This instrument adopted the same items as those used in the TIMSS.

I also control for some contextual data related to school locality.⁶ The locality controls include the locality type (rural, urban, or refugee camp), the poverty rate based on a 2009 consumption survey,⁷ the population per locality in 2013, the proportion of a locality under area C,⁸ and whether the

⁶ “Locality” is the smallest administrative unit defined by the PCBS. There are 528 localities in the West Bank.

⁷ This is the most recent available survey, conducted by the World Bank and PCBS and issued in 2013.

⁸ Area C: Areas in the West Bank that are still under full Israeli military and civil control based on the Oslo Accords of 1993, while the PA has civil and security control in area A. The PA has civil autonomy but no security control in area B (Vishwanath et al. 2014). Communities in area C are living in difficult life circumstances due to the lack of major services. For more details, see www.btselem.org/topic/Area_c.

locality is affected by the separation wall.⁹ The primary source of this information is the Palestine Central Bureau of Statistics (PCBS).

Finally, the dataset includes detailed information on students' physical and mental health, and teacher/peer relations in school. These data are drawn from the HBSC survey and student background questionnaires answered by households. I interpret this information as a likely channel through which class size may affect children's behavior. Section 7 provides more details about these questions and the way in which these questions are used to indicate behavioral problems.

To reduce bias due to omitted variables and to avoid reducing the sample size due to missing values of control variables, as well as increasing the precision of the estimates, some control variables with missing values are substituted by mean values, and an indicator for such missing values is created. In the final analysis, and after imposing some restrictions according to the suggested identification strategy, the sample ends up with approximately 4,700 students in grades 5–9 distributed across 416 different classes and enrolled in 91 schools located in 71 distinct localities throughout the West Bank. Table 1 presents a summary of the statistics of the dependent and control variables. Due to the sampling requirement, female students are overrepresented in the sample (67 percent).

Next, to check the potential endogeneity problems, I perform regressions to account for the class size variation between and within schools. Table A.1 in the appendix presents the results of the regressions of average class size in grades 5–9 on the individual, locality, and school level of controls. The results with the school fixed effect are presented in the last column. Locality attributions and class size are highly correlated, as indicated in all the columns. The presence of the separation wall is associated with an increase in class size by 1.8 students. The result could be attributed to the fact that the separation barrier reduces the mobility abilities for those people who live in densely populated localities plagued by the wall. Both the poverty rate and the locality's proportion in area C have a significant adverse impact on class size. Usually, schools in poor

⁹ The Israeli West Bank barrier or wall is a separation barrier built by the Israeli government in the West Bank along the 1949 Armistice Line known as the "Green Line" (B'Tselem 2012). The barrier divides Palestinian communities, encircles some, and isolates others from their surroundings, while separating East Jerusalem from the rest of the West Bank (UNSCO 2014).

communities and marginalized areas are less attractive to parents due to their lack of educational resources. Additionally, communities in area C are characterized as having inadequate primary services and suffer from unfortunate life conditions (UNSCO 2014).

Table 1: Descriptive Statistics

Variable	Mean	StdDev.	Min.	Max.
Panel A: Individual-level observations: 4,664				
Cognitive tests score (child)	60.65	16.66	0.55	94.48
Female child	0.66	0.47	0	1
Refugee status	0.42	0.49	0	1
High school diploma (father)	0.28	0.45	0	1
College degree (father)	0.19	0.39	0	1
High school diploma (mother)	0.30	0.46	0	1
College degree (mother)	0.15	0.35	0	1
Living standard scale (10 points)	3.53	2.27	0	9
Net monthly income (NIS) (5-point scale) ¹⁰	1.66	1.38	1	5
Books at home (5-point scale) ¹¹	1.52	1.23	1	5
No. of student's siblings	4.38	2.98	0	18
Number of rooms (above median)	0.54	0.50	0	1
School average 2010/11	71.89	16.07	23.33	99.42
School average 2011/12	69.69	16.00	18.50	99.83
Panel B: Class-level observations: 416				
Average class size	30.59	7.02	6	44.5
Cognitive test score (class)	60.62	11.28	27.98	85.31
Panel C: School-level observations: 91				
UNRWA school	0.45	0.50	0	1
School lack of resources (18-point scale)	6.02	4.47	0	17
School educational atmosphere total (40-point scale)	26.95	3.54	17	38
Problems among teachers (6-point scale)	1.73	1.64	0	6
Panel D: Locality-level observations: 71				
Presence of the separation wall	0.376	0.484	0	1
Percentage of the locality in area C	0.303	0.316	0	1
Poverty headcount rate	0.194	0.131	0	0.5035
Population per locality	24,735	48,168	237	195,733
Urban locality	0.416	0.493	0	1

¹⁰ The scale is: (1) <NIS 1,500; (2) NIS 1,500–NIS 2,499; (3) NIS 2,500–NIS 3,999; (4) NIS 4,000–NIS 5,000; and (5) >NIS 5,000.

¹¹ The scale is: (1) 0–10 books; (2) 11–25 books; (3) 26–100 books; (4) 101–200 books; (5) >200 books.

Panel E: Descriptive statistics for the secondary outcomes and mechanism

Variable	Observations	Mean	Std Dev.	Min.	Max.
Bully–victim	4,313	0.24	0.42	0	1
Bully–act	4,297	0.2	0.4	0	1
Fight in school and injury during past 12 months	4,272	0.93	1.26	0	4
SDQ impact: overall difficulties	3,619	1.53	0.70	1	4
School teacher support	4,253	3.94	1.76	1	10
Peer relations	4,316	5.50	2.35	1	15

In column (5) in table A.1 in the appendix, I add the cohort fixed effect to the specification; both school- and locality-level controls do not change and remain significant. The class size is likely to be smaller in grade nine, which represents the final level in the preparatory stage. School principals tend to distribute grade nine students into small classes, since those students tend to be more sociable and interactive, as they see themselves as superior to younger students. Parents may also decide to move their children to larger schools that contain a secondary level. The results in table A.1 suggest that students from disadvantaged localities attend smaller classes. Moreover, students from more affluent families and students in grade nine tend to attend smaller classes. If the selective placement in small and large classes were disregarded, the precise estimate of class size on achievement would be biased. In the next section, I will address the strategy to deal with this problem.

5. IDENTIFICATION STRATEGY

5.1 Econometric Model

The empirical specification for the model is presented below. The model assumes that the outcome variable (cognitive score, bullying, and violent behavior) of student i , enrolled in school k in town l (Y_{ikl}) is generated by the following equation:

$$Y_{ikl} = \alpha + \delta CS_{kl} + \beta X_{ikl} + \vartheta S_{kl} + \omega T_l + \varepsilon_{ikl} \quad (1)$$

where X_{ikl} is a vector of observable attributes of a student's individual-level controls, including sex, a dummy for the month of birth,¹² refugee status, a dummy for the student's grade, and household characteristics, such as parental level of education measured in three categories (less than, with, or more than a high school diploma), household income, the household living standard index (10 points),¹³ the number of books and rooms at home, and the number of siblings. S_{kl} is a vector of observable school-level controls and includes the school type (UNRWA versus governmental school), a school resource index, teacher characteristics, and a school climate index, as reported by school principals. T_l stands for town-level variables, like population and poverty rate, locality type (urban versus rural or camp), the presence of the separation wall, and the proportion of the locality under area C. CS is the average class size that student i attended during the 2012/13 academic year, and ε_{ikl} denotes all the other determinants of achievement, such as the unobserved attributes of the student, parents, school, and locality. The coefficient of interest is δ , the class size effect. Y is the outcome variable representing the student cognitive scores, bullying, and violent behavior. All the standard errors are clustered at the school level.

As indicated before, the endogeneity problem may produce biased ordinary least squares (OLS) estimates of the class size effect, and the exogenous source of variation in class size is needed for reliable identification. The available dataset provides information only on the average class size within the grade rather than the actual class size. In these data, there are two thresholds for the maximum class size number: 40 and 45 students/class in public and UNRWA schools, respectively. Following the literature, I exploit the maximum class size rule in the RD design induced by Angrist and Lavy (1999) to predict the average class size to be a nonlinear and discontinuous function of enrollment. The number of classes in grade j at school k in year t can be described by (f_{jkt}):

¹² The month of birth determines children's school entry age. The School Entry law could affect child outcomes other than through educational attainment. For example, Bedard and Dhuey (2006) indicated that older children in the same classroom may be treated differently (e.g., be placed in more advanced programs) or experience fewer social problems. Cascio and Schanzenbach (2007) pointed out that entrance start dates are correlated with school performance, grade repetition, and diagnoses of learning disabilities.

¹³ The household living standard index represents whether a household owns a TV, mobile phone, DVD player, air conditioner, car, computer, central heating, washing machine, or refrigerator.

$$\text{In governmental schools : } (f_{jkt}) = \text{int} \left[\frac{N_{jkt} - 1}{40} \right] + 1 \quad (2)$$

$$\text{In UNRWA schools : } (f_{jkt}) = \text{int} \left[\frac{N_{jkt} - 1}{45} \right] + 1 \quad (3)$$

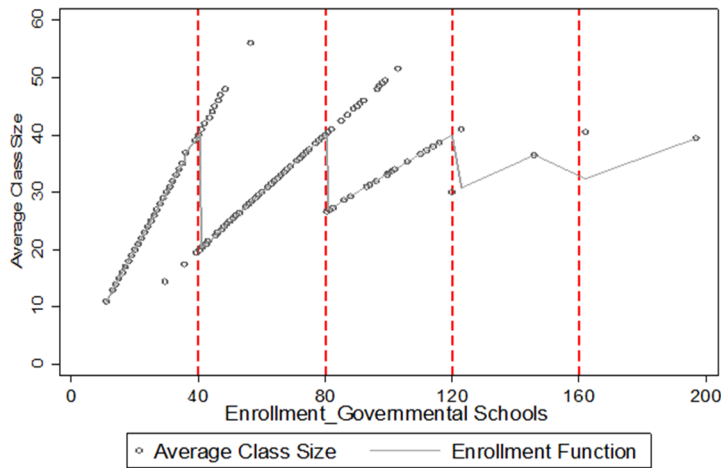
where N_{jkt} is the total enrollment in grade j in school k in year t , and $\text{int}(x)$ is the largest integer smaller than or equal to x .

Figures 2.a and 2.b plot the theoretical class size and the actual average class size against the cohort enrollment. The actual average class size is very close to the predicted one in UNRWA schools and the majority of governmental schools, except in a few cases, as shown in figure 2.a. This indicates that these schools do not have the resources to follow the maximum class size rule. However, the graph shows that there is no good variation to identify class size effects, as no variation is yielded in the first stage when the actual average class size is instrumented by the theoretical class size induced by equations (2) and (3). Focusing on those schools that comply with the maximum size rule only produces no differences in the class size coefficient between OLS and 2SLS estimates. Therefore, I restrict the sample to those classes that comply with the maximum class size rule. Then, I estimate the effect of class size with the same specification as in equation (1) by using the reduced form, since in this case the theoretical and average class size should be the same.¹⁴ As a result, 473 observations distributed over 20 governmental schools in 43 classes are excluded from the main analysis.¹⁵ One concern may still be that the class size effect would be overestimated as a result of excluding noncompliant schools from the estimating sample. Thus, those schools remaining in the analysis could have a positive selection in terms of resources and differ systematically left and right of thresholds. Table A.8 in the appendix reports the results for all observations without excluding noncompliant schools from the sample and shows no significant differences from the main results reported in the table 3.

¹⁴ The restricted sample contains only schools that comply with the maximum class size rule, since instrumenting the actual average class size with the theoretical average class size yields no variation in the first stage. I apply an OLS regression rather than 2SLS since the two provide the same results. One of the limitations of this approach is increasing the margin of error due to the decreasing sample size. In addition, the adopted identification strategy is valid only for a restricted subsample; therefore, the results may be less conclusive and might affect the external validity. However, this depends on the size of the affected sample, which is very small in the current case.

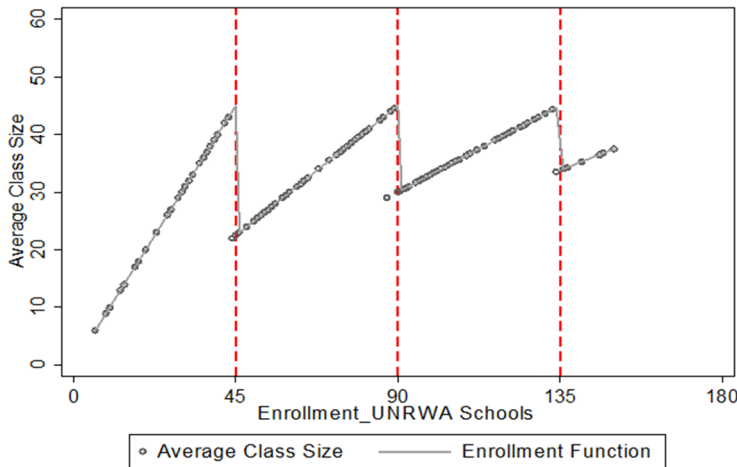
¹⁵ Table A.4 represents the differences in observed variables (particularly at the school level) between schools that comply and schools that do not comply with the maximum class size rule.

Figure 2.a: Average and Predicted Class Size (governmental schools)



Source: Own calculations based on survey data and school enrollment.

Figure 2.b: Average and Predicted Class Size (UNRWA schools)



Source: Own calculations based on survey data and school enrollment.

5.2 Threats to Validity

This identification strategy risks being influenced by unobserved school choice (specifically omitted variables) and the possibility of parents sorting across regions in search of better services, including schools. To deal with these potential problems, I control for a variety of school covariates using the principals' survey to construct school educational indicators. These indicators contain three indexes that capture many school characteristics and reflect the education climate: (1) an educational atmosphere index, such as teacher job satisfaction, curriculum, parental support, positive involvement in the educational process, and so on; (2) problematic behaviors of teachers, such as

late school arrivals and unjustified absenteeism; and (3) a school resource index, for example adequate stationery, the availability of qualified teachers, educational space, and heating/cooling and lighting systems.

Regarding the threat of crossregional movements, the West Bank is characterized as being geographically fragmented (Vishwanath et al. 2014), diminishing the capacity for internal and external mobility. The presence of the separation wall—existing in many area C localities within West Bank territories—restricts movement, causing political and economic instability. This in turn ruptures West Bank communities. The most recent available mobility report from the PCBS (2010) confirms that the majority of West Bank residents (more than 90 percent) have the same residency as their birthplace.¹⁶

Another threat to validity is sorting near cutoffs. The regression discontinuity design (RDD) induced by the maximum class size rule assumes that the predicted class size is unrelated to the student or school characteristics. This assumption is violated in private schools (Urquiola and Verhoogen 2009), as well as in public schools (Danny, Gradstein, and Reuven 2009) when higher socioeconomic status (SES) households sort themselves into higher quality schools as a response to a school administration’s attempts to manipulate enrollment to avoid adding additional classrooms. Therefore, RD estimates overestimate the effect of class size on student outcomes, since the students after the cutoffs come from higher SES backgrounds. Such sorting seems to be unlikely in the West Bank, where students are required to attend the closest school to their residence and mobility across school districts is rare. Moreover, to deal with that issue, I estimate the RD model based on Urquiola and Verhoogen’s (2009) specification, which can be written as follows:

$$Y_{is} = \gamma E(CS_s | N_s) + \alpha(N_s) + \mu_{is} \quad (4)$$

$$E(CS_s | N_s) = \theta 1(N_s \geq C_u) + b(N_s) \quad (5)$$

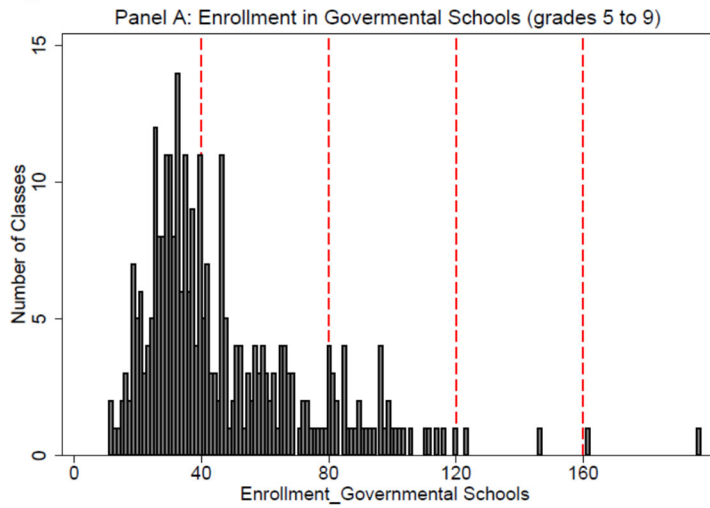
¹⁶ Other researchers have indicated the low mobility in the West Bank, for example Di Maio and Nandi (2013) and Mansour and Rees (2012).

where i and s indicate individual and school, respectively. Y_{is} is the outcome of an individual student enrolled in school s . $E(CSs)$ is the average class size in school s , N_s is the grade enrollment, C_u is the value of the class size cutoff (i.e., 40 in governmental schools and 45 in UNRWA schools), and $a(\cdot)$ and $b(\cdot)$ are flexible functions of enrollment. Figure (3) presents a histogram of enrollment in grades 5–9 in both governmental and UNRWA schools. The histogram does not show any pattern in which the numbers of class sections are stacked exactly at the cutoff point, except in a few cases in the public schools at the first and the second cutoff point. The analysis will focus on the two first cutoff points since there are not enough class sections after the third cutoff. In general, the enrollment is not high in both types of schools, which gives the estimation more precision, since Maimonides' rule better explains the class size effect at low levels of enrollment (Angrist, Battistin, and Vuri 2017). Furthermore, the not-high enrollment level reduces the possibility of manipulation of the forcing variable, enrollment, in this case.

Like Urquiola and Verhoogen (2009), table 2 presents regressions on household characteristics with the indicator variables for the class size cutoffs and enrollment. (I use dummies indicating whether enrollments are above each of the cutoff points and piecewise linear splines for enrollments are used.) The results provide evidence that there is no significant jump at the first or second enrollment cutoff for the selected covariates (mother's education, household income, and household standard of living index). Moreover, table A.5 in the appendix provides additional evidence using RD estimates induced by Calonico, Cattaneo, and Titiunik (2014) to verify that there is no violation of RDD assumptions with regard to the sorting of household characteristics at the cutoff points. These results are consistent with the visual evidence obtained from figures 4.1.a, 4.1.b, 4.2.a, and 4.2.b, which confirm that mothers' schooling and the household standard of living index are not significantly different before or after the cutoff points in both types of school.

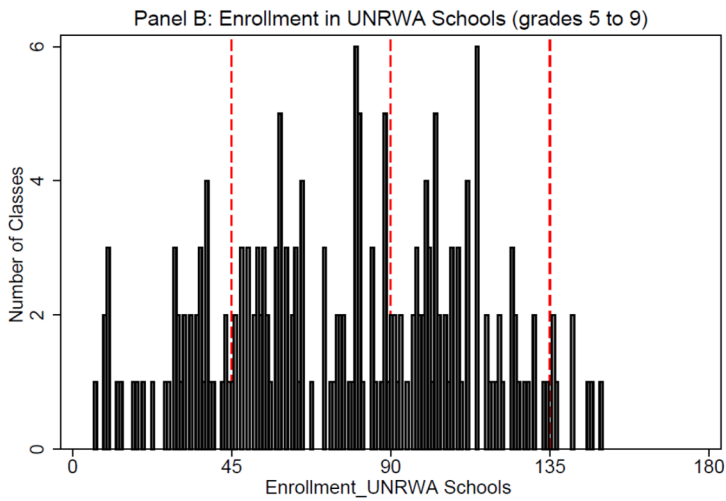
The main estimation assumes that controlling several levels of covariates, such as including school characteristics and locality level and excluding schools that severely deviate from the maximum class size rule, is sufficient to correct the biases of selective placement into schools with different class sizes.

Figure 3.a: Enrollment in Governmental Schools



Source: Own calculations based on school enrollment.

Figure 3.b: Enrollment in UNRWA Schools



Source: Own calculations based on school enrollment.

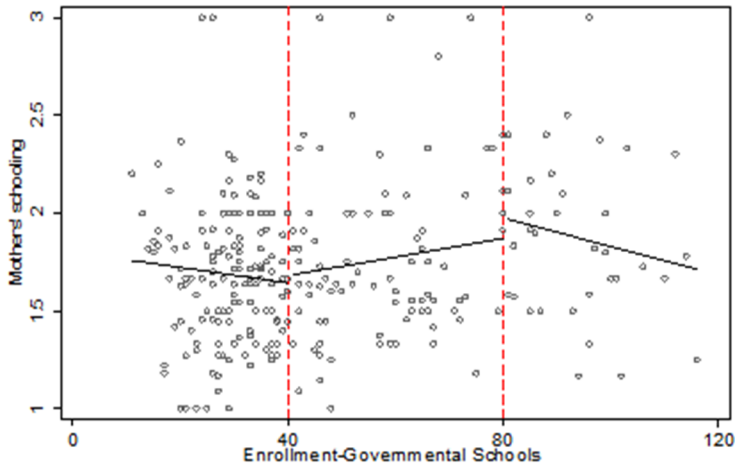
Table 2: Behavior of Selected Variables around Enrollment Cutoffs

Dep. var.:	(1)	(2)	(3)
	Moth. edu.	Income	Living standard
Panel A: Governmental schools			
$1\{N \geq 41\}$	0.024 (0.129)	0.356 (0.283)	0.017 (0.364)
$1\{N \geq 81\}$	-0.230 (0.200)	0.060 (0.222)	-0.299 (0.295)
N	-0.004 (0.006)	-0.016 (0.012)	-0.002 (0.009)
$(N-41) \times 1\{N \geq 41\}$	0.007 (0.009)	0.005 (0.015)	0.004 (0.017)
$(N-81) \times 1\{N \geq 81\}$	0.001 (0.006)	0.014 (0.011)	0.009 (0.016)
Constant	1.797*** (0.184)	2.682*** (0.382)	4.317*** (0.301)
Observations	1,961	1,927	2,089
R-squared	0.004	0.008	0.002
Panel B: UNRWA schools			
$1\{N \geq 46\}$	-0.039 (0.168)	0.148 (0.209)	0.003 (0.319)
$1\{N \geq 91\}$	0.085 (0.101)	0.047 (0.162)	-0.066 (0.245)
N	0.007 (0.008)	-0.005 (0.012)	0.013 (0.012)
$(N-46) \times 1\{N \geq 46\}$	-0.008 (0.009)	0.004 (0.013)	-0.015 (0.014)
$(N-91) \times 1\{N \geq 91\}$	0.003 (0.005)	-0.002 (0.008)	-0.005 (0.010)
Constant	1.427*** (0.102)	1.467*** (0.267)	2.071*** (0.413)
Observations	1,582	1,525	1,630
R-squared	0.011	0.002	0.006

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, and * p<0.1.

N denotes the enrollment function.

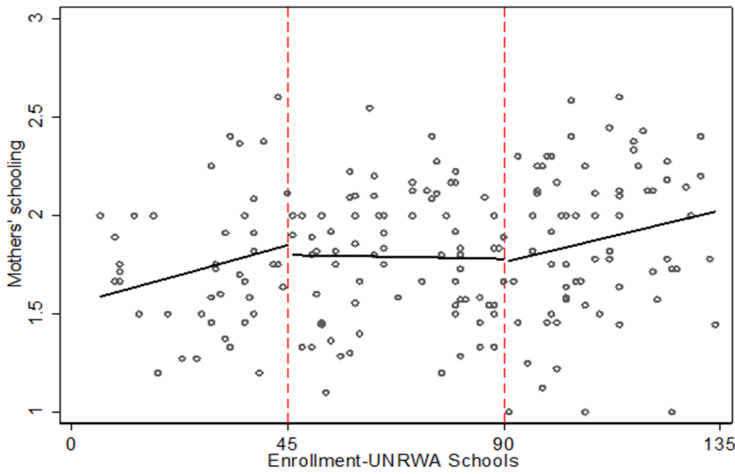
Figure 4.1.a: Enrollment and Mothers' Schooling (government schools)



Source: Own calculations based on survey data and school enrollment.

Note: Mothers' schooling is measured on a three-level scale: (1) less than high school diploma (12 years), (2) high school diploma, and (3) college degree.

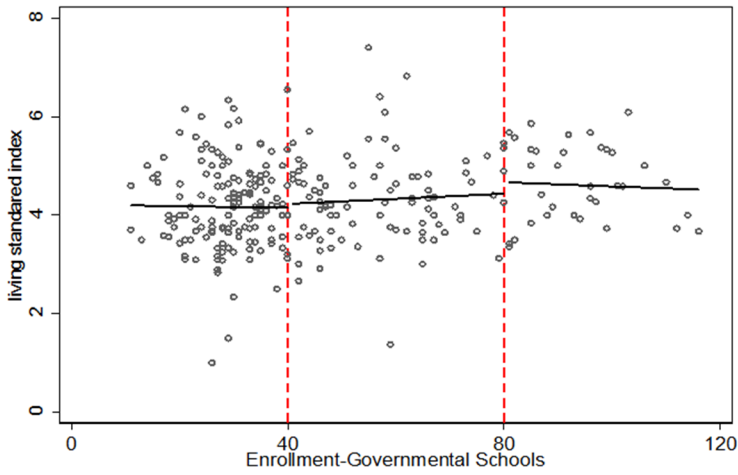
Figure 4.1.b: Enrollment and Mothers' Schooling (UNRWA schools)



Source: Own calculations based on survey data and school enrollment.

Note: Mothers' schooling was measured on a three-level scale: (1) less than high school diploma (12 years), (2) high school diploma and (3) college degree.

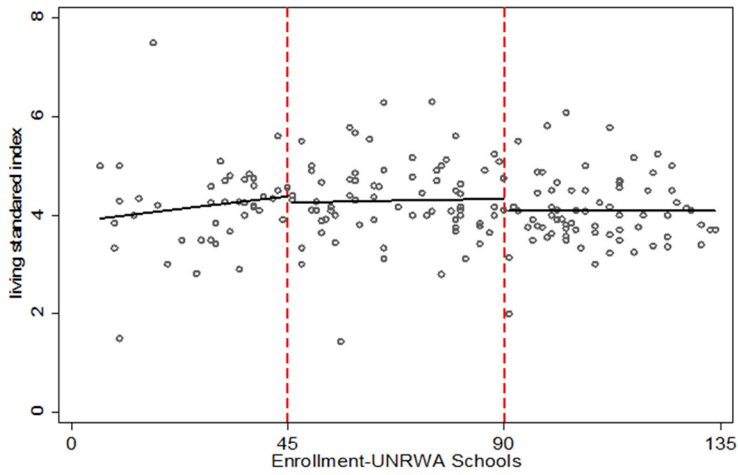
Figure 4.2.a: Enrollment and Student’s Living Standard Index (government schools)



Source: Own calculations based on survey data and school enrollment.

Note: Living standard index is an indicator of whether the student’s household owns a TV, mobile phone, DVD player, air conditioner, cars, etc.

Figure 4.2.b: Enrollment and Student’s Living Standard Index (UNRWA schools)



Source: Own calculations based on survey data and school enrollment.

Note: Living standard index is an indicator of whether student’s household owns a TV, mobile phone, DVD player, air conditioner, cars, etc.

6 RESULTS

6.1 Cognitive Tests Results

The estimates of class size effects based on the identification strategy illustrated in the previous section are presented in table 3. Each reported coefficient comes from the various specifications. The results in all the columns indicate the negative correlation between class size and cognitive abilities. Column (1) represents the effect of class size without any control except the enrollment function, which suggests an insignificant negative relation. Columns (2) and (3) are obtained from a specification with individual and locality controls; the estimates are negative but still insignificant. The result in column (4) is obtained after adding school-level controls. The estimates increase but remain statistically insignificant. Column (5) reports the results from specifications that also include performance in the previous two academic years. Including controls produces small negative class size estimates that are statistically insignificant. Table 3 suggests that reducing the average class size by one pupil has no significant impact on a student's cognitive test score. Figure 5.a plots the average class cognitive test scores against enrollment. The absence of a jump in scores at cutoffs in data from governmental schools suggests that class size reduction leaves the scores unchanged. However, figure 5.b shows a significant reduction in the average class score after the first and second cutoffs in UNRWA schools.

Table A.2 in the appendix also reports the effect of other control variables on students' cognitive abilities. Overall, females, refugees, and UNRWA school students outperform males, nonrefugees, and public school students. The results are dramatically negatively affected by the locality's poverty index. As expected, cognitive ability is strongly related to a student's grade and school performance.

There are strands in the literature indicating that the class size effect varies across subjects and grades in terms of the sign and size of the estimates. Table 4 demonstrates that class size in general has a negative effect on students' performance in all the cognitive subtests in all the grades, but these estimates are not statistically significant. The magnitude of these estimates, however, is substantially larger and statistically significant at the 1 percent level in grade six. The coefficient in column (3) suggests that reducing the class size by one student is significantly associated with higher cognitive scores by 0.4 percentage points compared with approximately 0.075 percentage

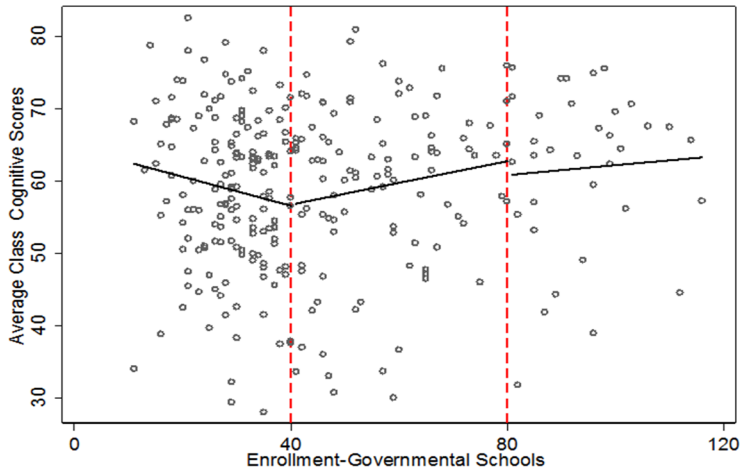
points in other grades. Table 4 also shows that the estimate for grade nine is greater than those obtained in the antecedent two grades. In fact, the one common characteristic in grades six and nine is that these two grades represent the final stage in their level. Traditionally, schooling in Palestine is divided into three stages: elementary (grades 1–6), preparatory (grades 7–9), and secondary (grades 10–12). Seniority affects conduct, especially for those who are moving to a higher stage. Students in these grades are wont to feel superior to students in lower grades and tend to be more sociable in class and less controllable. Furthermore, moving to the next stage may cause huge levels of uncertainty and anxiety (Hofstede 1983) and may lead to aggressive behavior, such as bullying.

Table 3: Class Size and Students' Cognitive Test Scores

Dep. var.: cognitive tests	(1)	(2)	(3)	(4)	(5)
Average class size	-0.007 (0.100)	0.053 (0.084)	-0.047 (0.072)	-0.096 (0.069)	-0.075 (0.066)
Enrollment	0.007 (0.026)	-0.027 (0.020)	-0.012 (0.022)	-0.015 (0.023)	-0.027 (0.022)
Constant	60.484*** (2.889)	38.931*** (4.027)	49.100*** (4.572)	40.288*** (7.640)	9.091 (7.792)
Observations	4,664	4,664	4,664	4,664	4,642
R-squared	0.000	0.257	0.288	0.305	0.467
Mean of dep. var.	60.65	60.65	60.65	60.65	60.65
Std err. of dep. var.	16.66	16.66	16.66	16.66	16.66
Individual-level controls	NO	YES	YES	YES	YES
Locality-level controls	NO	NO	YES	YES	YES
School-level controls	NO	NO	NO	YES	YES
School GPA in 2011/2012	NO	NO	NO	NO	YES

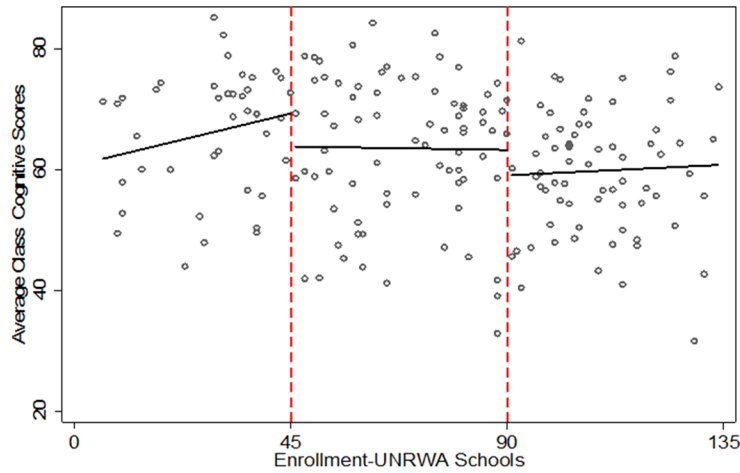
Notes: The robust standard errors clustered at the school level are reported in parentheses. The individual-level controls include sex, dummies for the month of birth and refugee status, household income, father's and mother's education, the standard of living scale (with nine points), the number of books at home, the number of rooms above the median per locality, the number of siblings, and the student's grade. The locality-level controls contain the poverty rate, the proportion of the locality under area C, the presence of the separation wall, the population in 2013, and the locality type (dummy for urban). The school-level controls are the school resource scale, school learning atmosphere and problems among teachers, and school type (UNRWA or governmental). The school GPA includes the school GPA for the 2010/11 and 2011/12 academic years. *** p<0.01, ** p<0.05, and * p<0.1.

Figure 5.a: Enrollment and Average Class Cognitive Test Scores (government schools)



Source: Own calculations based on survey data and school enrollment.

Figure 5.b: Enrollment and Average Class Cognitive Test Scores (UNRWA schools)



Source: Own calculations based on survey data and school enrollment.

Table 4: OLS Estimates: The Effect of Class Size on Cognitive Tests per Grade

	(1)	(2)	(3)	(4)	(5)	(6)
Grade	All	5	6	7	8	9
All cognitive test parts	-0.075 (0.066)	0.051 (0.131)	-0.392*** (0.145)	-0.067 (0.119)	0.054 (0.139)	-0.198 (0.137)
Numerical tests	-0.032 (0.097)	0.070 (0.186)	-0.378* (0.190)	0.168 (0.161)	0.101 (0.213)	-0.231 (0.212)
Figural tests	-0.102 (0.082)	0.064 (0.171)	-0.464** (0.177)	-0.121 (0.142)	-0.004 (0.168)	-0.192 (0.153)
Verbal tests	-0.069 (0.062)	0.037 (0.124)	-0.348*** (0.128)	-0.095 (0.118)	0.080 (0.128)	-0.192 (0.130)
No. of observations	4,642	918	916	927	924	957
Enrollment	YES	YES	YES	YES	YES	YES
Individual-level controls	YES	YES	YES	YES	YES	YES
School-level controls	YES	YES	YES	YES	YES	YES
Locality-level controls	YES	YES	YES	YES	YES	YES
Student GPA in 2011 and 2012	YES	YES	YES	YES	YES	YES

Notes: The robust standard errors clustered at the school level are reported in parentheses. The individual-level controls include sex, dummies for the month of birth and refugee status, household income, father's and mother's education, the standard of living scale (with nine points), the number of books at home, the number of rooms above the median per locality, the number of siblings, and the student's grade. The locality-level controls contain the poverty rate, the proportion of the locality under area C, the presence of the separation wall, the population in 2013, and the locality type (dummy for urban). The school-level controls are the school resource scale, school learning atmosphere and problems among teachers, and school type (UNRWA or governmental). The school GPA includes the school GPA for the 2010/11 and 2011/12 academic years. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

6.2 Heterogeneous Class Size Effects on Cognitive Test Scores

I will now investigate whether this insignificant effect of class size reduction reported in the main results will have a different effect on the cognitive results for specific subgroups. There is evidence that the gains from small classes are largest for students from disadvantaged backgrounds, as indicated by Angrist and Lavy (1999) and Krueger (1999). Other researchers have emphasized that the outcome of the class depends on the teacher characteristics (Woessmann and West 2006) and student achievement records (Konstantopoulos 2008).

Table 5 presents the results for specific groups using the same specification as in equation (1). The first subgroup is classified according to gender; the results differ substantially from the results for girls and boys together but are still statistically insignificant. Column (3), panel (A) suggests that

boys benefit more from class size reduction than girls. As for the second subgroup, I divide the students into two groups according to their school achievement in the previous academic year (2011/12). Column (4) indicates that class size has a larger negative influence on students with a lower-than-average GPA (>70 percent) but no significant effect on the opposite group. Subgroups 4, 5, and 6 in panel (B) are classified according to the locality type, teacher problem index, and school type, respectively. All of the reported estimates are statistically insignificant. Notably, the class size reduction has approximately zero effect on UNRWA students, which is consistent with the UNRWA claim that there is no evidence to support a class size reduction policy. To conclude, table 5 does not provide evidence in favor of heterogeneous class size effects. The only exception is for students with low academic achievement profiles and those from disadvantaged backgrounds, who benefit the most from a reduction in class size.

Table 5: Class Size Effects for Subgroups

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A						
Dep. Var.: cognitive tests	Child gender		School GPA in previous year		Mother education	
	Girls	Boys	<70	>70	<12 years	> 12 years
class Size	-0.045 (0.044)	-0.132 (0.153)	-0.164* (0.091)	0.008 (0.063)	-0.097 (0.069)	0.044 (0.083)
Constant	15.763** (6.080)	7.740 (14.245)	1.933 (8.092)	13.202 (8.612)	9.515 (7.743)	-1.227 (11.335)
Observations	3,095	1,547	2,397	2,240	3,960	682
R-squared	0.515	0.419	0.383	0.405	0.456	0.511
Mean of Dep. Var.	59.04	63.18	60.32	54.73	67.13	59.77
Std.Err. of Dep. Var.	16.82	14.81	16.84	16.70	13.95	16.60
Panel B						
	Teacher problem index		Locality type		School type	
	Equal to 0	0 > and < 6	Urban	Rural/Camp	UNRWA	GOV
class Size	-0.049 (0.064)	-0.112 (0.089)	0.043 (0.107)	-0.115 (0.091)	-0.001 (0.081)	-0.071 (0.093)
Constant	27.260*** (9.389)	6.489 (8.654)	-8.672 (9.775)	20.581* (10.573)	10.121 (7.751)	10.387 (11.657)
Observations	1,057	3,585	1,937	2,705	2,065	2,577
R-squared	0.564	0.469	0.504	0.464	0.459	0.489
Mean of Dep. Var.	63.58	62.04	55.79	59.53	66.19	62.80
Std.Err. of Dep. Var.	15.94	15.83	18.83	16.93	15.71	16.15
Individual-level controls	YES	YES	YES	YES	YES	YES
Locality-level controls	YES	YES	YES	YES	YES	YES
School-level controls	YES	YES	YES	YES	YES	YES
Cohort FE	YES	YES	YES	YES	YES	YES
School GPA in 2011/2012	YES	YES	YES	YES	YES	YES

Notes: The robust standard errors clustered at the school level are reported in parentheses. The individual-level controls include sex, dummies for the month of birth and refugee status, household income, father's and mother's education, the standard of living scale (with nine points), the number of books at home, the number of rooms above the median per locality, the number of siblings, and the student's grade. The locality-level controls contain the poverty rate, the proportion of the locality under area C, the presence of the separation wall, the population in 2013, and the locality type (dummy for urban). The school-level controls are the school resource scale, school learning atmosphere and problems among teachers, and school type (UNRWA or governmental). The school GPA includes the school GPA for the 2010/11 and 2011/12 academic years. *** p<0.01, ** p<0.05, and * p<0.1.

6.3 Effect of Class Size on Bullying and Violent Behavior

Table A.6 in the appendix shows the results of the marginal effect based on the model as outlined in equation (1) estimated using the probit model. Class size does not seem to be associated with any statistically significant effect on bullying indicators for perpetrator or perpetrator/victim characteristics, while the results reveal a significant relation to the students' self-reported indication of being bullied by others. The result in table 6, column (3) suggests that increasing the class size by one student increases the predicted probability of the student having been bullied during the past two months by 0.009 percentage points. After adding the school fixed effect to control for between-school endogeneity, the probability increases to 0.0135 with higher precision power. Next, I investigate the effect of class size on students' self-reported probability of being bullied using a different specification, as illustrated in table 6, panel (A). Columns (1) through (5) include a linear function of the forcing variable (enrollment), while column (6) presents the result of the quadratic function of enrollment and adds the school fixed effect to the model. The estimate is a very precisely estimated zero and reveals a significant effect of class size. The size of the estimation is remarkably similar when controlling for a student self-reporting as a bully. Column (6) demonstrates that a ten-student class size increase will increase the probability of a student being bullied by 13 percent, which corresponds to 0.55 percent of the mean.

In the second part of table 6, I present the results for the effect of class size on violence among peers. All the estimations (columns [1] through [6]) confirm that larger classes increase violence among students. However, the precision of these estimations varies according to each specification. Column (4) shows the results under a full set of controls, including the school fixed effect. The coefficient suggests that a ten-student class size increase is associated with increasing average student engagement in fighting during the past 12 months and needing medical treatment by 0.10 times. After including the quadratic function of enrollment in the estimation, shown in column (6), the coefficient does not change; however, here the precision power is lower (with 10 percent significance).

Table 6: Effect of Class Size on Behavioral Problems

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A, dep. var.: bullying victim. Probit estimation						
Class size	0.00902* (0.00514)	0.00818* (0.00451)	0.00944** (0.00427)	0.0135** (0.00636)	0.0131** (0.00623)	0.0129** (0.00616)
Bullying act			1.374*** (0.0613)		1.381*** (0.0657)	1.381*** (0.0658)
Observations	4,313	4,292	4,224	4,290	4,222	4,222
Mean of dep. var.	0.236	0.236	0.236	0.232	0.232	0.232
Std err. of dep. var.	0.425	0.425	0.425	0.422	0.422	0.422
Panel B, dep. var.: violent behavior. OLS estimation						
Class size	0.0127*** (0.00453)	0.0101** (0.00390)	0.00923** (0.00387)	0.0109** (0.00542)	0.00947* (0.00551)	0.00975* (0.00536)
Being bullied			0.464*** (0.0610)		0.468*** (0.0601)	0.468*** (0.0601)
Observations	4,223	4,202	4,202	4,202	4,202	4,202
R-squared	0.004	0.055	0.078	0.087	0.109	0.110
Mean of dep. var.	0.918	0.918	0.918	0.918	0.918	0.922
Std err. of dep. var.	1.255	1.255	1.255	1.255	1.255	1.256
Enrollment	YES	YES	YES	YES	YES	YES
Enrollment squared	NO	NO	NO	NO	NO	YES
Individual/locality controls	NO	YES	YES	YES	YES	YES
School-level controls	NO	YES	YES	NO	NO	NO
School F.E.	NO	NO	NO	YES	YES	YES
Bullying act/victim	NO	NO	YES	NO	YES	YES

Notes: The robust standard errors clustered at the school level are reported in parentheses. The controls include: individual-level controls, sex, dummies for the month of birth and refugee status, household income, father's and mother's education, the standard of living scale (with nine points), the number of books at home, the number of rooms above the median per locality, the number of siblings, and the student's grade. The locality-level controls contain the poverty rate, the proportion of the locality under area C, the presence of the separation wall, the population in 2013, the locality type (dummy for urban), and the school GPA for the 2010/11 and 2011/12 academic years. The school-level controls are the school resource scale, school learning atmosphere and problems among teachers, and school type (UNRWA or governmental). *** p<0.01, ** p<0.05, and * p<0.

Focusing more narrowly on the discontinuity cutoff, as in Urquiola and Verhoogen (2009) and Van der Klaauw (2002), table 7 presents the OLS¹⁷ results for five student enrollment bands around the first and second cutoffs. I add an indicator for whether schools' enrollment is above or below the respective cutoff to investigate the effect of class size around each discontinuity. Panels (A) and (B) present the results for the violent behavior outcome for the two schooling systems, while the victim of bullying outcome is presented in panels (C) and (D). The first two columns include all the

¹⁷ The probit model and OLS yield similar outcomes for the bullying victim outcome.

controls in equation (1), while the school fixed effect is added in columns (3) and (4). Columns 3 and 4 reveal that the effect of class size on self-reported conduct problems is positive for both types of schools and significant in some cases even after adding the school fixed effect. The only exception is the result for self-reported victims of bullying for UNRWA schools, which yields a negative relation.¹⁸

Based on the RDD validity, as table 2 demonstrates, there is reason to interpret the consistent estimates produced in tables 6 and 7 as a causal effect of class size on the self-reporting of violent behavior and self-reported indication of being a victim of violence. The next section attempts to propose a mechanism that might serve as a mediator channel related to the context of this study.

¹⁸ The same analysis is performed for the main outcome (cognitive test scores) and the results are presented in table A.3 in the appendix. The estimates of the effect of class size on students' cognitive test scores are uniformly negative, although they are not statistically significant.

Table 7: Within Five Student Enrollment Band Regressions: Class Size, Violent Behavior, and Victims of Bullying

Dep. var. violent behavior cutoff	(1) 1st	(2) 2nd	(3) 1st	(4) 2nd
Panel A: Governmental schools				
Dep. var.: Violent behavior	(40 students)	(80 students)	(40 students)	(80 students)
Enrollment (N)	36≥N≥45	76≥N≥85	36≥N≥45	76≥N≥85
Class size	0.016 (0.010)	0.895 (0.676)	-0.030*** (0.010)	0.895 (0.676)
Observations	426	68	426	68
R-squared	0.170	0.507	0.206	0.507
Panel B: UNRWA schools				
Dep. var.: Violent behavior	(45 students)	(90 students)	(45 students)	(90 students)
Enrollment (N)	41≥N≥50	86≥N≥95	41≥N≥50	86≥N≥95
Class size	0.043*** (0.002)	0.053*** (0.014)	0.043*** (0.002)	0.069 (0.047)
Observations	155	197	155	197
R-squared	0.248	0.329	0.248	0.352
Panel C: Governmental schools				
Dep. var.: Bullying victim	(40 students)	(80 students)	(40 students)	(80 students)
Enrollment (N)	36≥N≥45	76≥N≥85	36≥N≥45	76≥N≥85
Class size	0.002 (0.003)	0.638 (0.388)	0.007** (0.003)	0.638 (0.388)
Observations	427	67	427	67
R-squared	0.186	0.594	0.222	0.594
Panel D: UNRWA schools				
Dep. var.: Bullying victim	(45 students)	(90 students)	(45 students)	(90 students)
Enrollment (N)	41≥N≥50	86≥N≥95	41≥N≥50	86≥N≥95
Class size	-0.001 (0.002)	-0.007** (0.002)	-0.001 (0.002)	-0.012 (0.009)
Observations	156	195	156	195
R-squared	0.359	0.257	0.359	0.271
Controls	YES	YES	YES	YES
School FE	NO	NO	YES	YES

Notes: The robust standard errors clustered at the school level are reported in parentheses. The controls include the same covariates in table 6. *** p<0.01, ** p<0.05, and * p<0.1.

7. POSSIBLE MECHANISMS

The characteristics of school children who are victims of bullying are associated with symptoms in the domains of emotion, behavior, hyperactivity, and peer relationships (Moura et al. 2011). A strand of literature has also suggested that personality traits might contribute to typical bullying and aggressive behavior (Tani et al. 2003). Eriksen, Nielsen, and Simonsen (2012) suggested that the teachers' perception is a mechanism channel through which bullying could affect victims' school achievements. Teachers perceive bullied children to be particularly weak in terms of academic skills. This perception affects teachers' confidence in and willingness to invest in victims and bullies. They also view bullied children as having worse moods and worse social competencies.

In this section, I will propose two transmission channels that may arise through which class size can affect the self-reporting of being bullied and engagement in violent behavior. The first group contains mental health attributes assessed by SDQ¹⁹ questions as part of the parental background questionnaire. The second group discusses student/peer/teacher relations. Students' peer relations were measured using an index established by three survey choices: "The students in my class enjoy being together"; "Most of the students in my class are kind and helpful"; and "Other students accept me as I am." Meanwhile, the teacher support index was obtained through two self-reported survey questions: "When I need extra help from the teachers I can get it"; and "My teachers are interested in me as a person." The answers to each question were rated on a five-point scale (strongly agree, agree, neither agree nor disagree, disagree, and strongly disagree). Then, the total points were considered as a final measurement for each index (on a fifteen-point scale for peer relations and a ten-point scale for teacher support), where more points indicate worse peer relations and teacher support.

¹⁹ The Strengths and Difficulties Questionnaire (SDQ) is a brief behavioral screening questionnaire about 3–16-year olds. There are currently three self-reported versions of the SDQ in the following categories: children, parents, and teachers. Each version includes between one and three of the following components: 25 items on psychological attributes, an impact supplement, and follow-up questions (Goodman 1997; Goodman, Lamping, and Ploubidis 2010). In this study, I employ the parent-reported version and the first two components of the questionnaire: 25 items on psychological attributes and the impact supplement. These extended versions of the SDQ ask whether the respondent thinks the young person has a problem and, if so, enquire further about chronicity, distress, social impairment, and burden to others (Goodman 1999). For a detailed description of the SDQ, see <http://www.sdqinfo.com/>.

Table 8 shows the results of the probit and OLS regressions of self-reported victims of bullying and engagement in violent behavior from the SDQ total scores, scores for internalizing and externalizing problems, and the impact supplement component of the SDQ questionnaire,²⁰ as well as peer relations and students' self-reported teacher support. The results in table 8 obtained from the same specification are used in equation (1). I also control for school fixed effects and employ the quadratic function of the forcing variable "enrollment." Columns (1) and (3) present 12 separate regressions, in which each proposed mediating variable is included one at a time. Even so, both mental health indicators on the one hand and students' relations and teacher support on the other hand are strongly related to the two outcome conduct problems. Notably, the SDQ impact supplement coefficient is larger (more than two times) than the other mediator coefficients. Columns (2) and (4) in table 8 show the results when the two mediator channels are jointly included. The SDQ's internalizing and externalizing components are still strong predictors of the two outcomes. The sizable influence of the SDQ impact supplement starts to vanish when victimization of bullying is the outcome, while it becomes smaller but still has a significant effect on self-reported violent behavior. In addition, teacher support still has a remarkable impact on self-reported victimization of bullying and violent behavior.

²⁰ Parents were asked: "In general, do you think that your child has difficulties in one or more of the following areas: emotion, concentration, behavior, harmony, and dealings with others?" The answers to this question were provided on a four-point scale indicating zero difficulties, a few difficulties, clear difficulties, and severe difficulties.

Table 8: Potential Mechanism

Dep. var.	(1)	(2)	(3)	(4)
	Probit estimation		OLS estimation	
	Victim of bullying		Violent behavior	
SDQ total	0.0124*** (0.00465)		0.0102*** (0.00386)	
N	3,687		3,686	
SDQ internalizing	0.0116 (0.00889)	0.00392 (0.0108)	0.0267*** (0.00722)	0.0251*** (0.00839)
N	3,687		3,686	
SDQ externalizing	0.0220*** (0.00681)	0.0189** (0.00865)	0.00464 (0.00545)	-0.00527 (0.00745)
N	3,687		3,686	
SDQ impact supplement difficulties	0.0911** (0.0385)	0.0309 (0.0441)	0.0897*** (0.0292)	0.0590* (0.0341)
N	3,519		3,515	
Peer relation	0.0252** (0.0106)	0.0171 (0.0120)	0.00220 (0.0104)	-0.0109 (0.0108)
N	4,474		4,434	
School teacher support	0.0470*** (0.0143)	0.0354* (0.0183)	0.0334*** (0.0115)	0.0291** (0.0136)
N	4,347		4,366	
Enrollment	YES	YES	YES	YES
Enrollment squared	YES	YES	YES	YES
Controls	YES	YES	YES	YES
School FE	YES	YES	YES	YES
Observations		3,354		3,356
R-squared				0.104
Mean of dep. var.	0.239	0.230	0.924	0.907
Std err. of dep. var.	0.427	0.421	1.258	1.245

Notes: The robust standard errors clustered at the school level are reported in parentheses. The controls are the same as in the other specifications, as indicated in equation (1). *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

Next, to demonstrate whether these channels are on the causal path from class size to students' conduct problems proxied by self-reported victims of bullying and violent behavior, I estimate the same specification as in equation (1) with student–peer relations, teacher support, and SDQ components as outcomes. The results from table 9 suggest that peer relations and student mental health, obtained by the SDQ impact supplement questions, are the possible transmission mechanism

explaining the positive correlation between class size and conduct problems.

Table9: Class Size Effect on Teacher Support, Peer Relations, and SDQ

Dep. var.:	(1) Peer relation	(2) School teacher support	(3) SDQ total	(4) SDQ internalizing	(5) SDQ externalizing	(6) SDQ impact supplement difficulties
Class size	0.0192* (0.0113)	-0.00970 (0.00740)	-0.000812 (0.0204)	0.00273 (0.0121)	0.00160 (0.0146)	0.00559** (0.00264)
Observations	4,434	4,366	3,686	3,708	3,689	3,605
R-squared	0.054	0.093	0.144	0.103	0.149	0.128
Enrollment	YES	YES	YES	YES	YES	YES
Enrollment squared	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES
School FE	YES	YES	YES	YES	YES	YES
Mean of dep. var.	5.530	3.972	12.55	6.114	6.540	1.527
Std err. of dep. var.	2.328	1.786	5.587	3.098	3.537	0.690

Notes: The robust standard errors clustered at the school level are reported in parentheses. The individual-level controls include sex, dummies for the month of birth and refugee status, household income, father’s and mother’s education, the standard of living scale (with nine points), the number of books at home, the number of rooms above the median per locality, the number of siblings, and the student’s grade. The locality-level controls contain the poverty rate, the proportion of the locality under area C, the presence of the separation wall, the population in 2013, the locality type (dummy for urban), and the school GPA for the 2010/11 and 2011/12 academic years. The school-level controls are the school resource scale, school learning atmosphere and problems among teachers, and school type (UNRWA or governmental). *** p<0.01, ** p<0.05, and * p<0.

Table 10 presents a further explanation of the SDQ impact supplement question. If the child or parent answered “yes” to the question, “Overall, I think that the child has difficulties in one or more of the following areas: emotions, concentration, behavior, or being able to get on with other people,” then five further questions were posed about the severity of these difficulties. Taken together, the five questions concerning overall distress and social impairment can be used to generate an impact score that ranges from 0–10.²¹ The impact supplement provides an important estimate of the burden of the problems, which is an essential part of the diagnostic criteria in the

²¹ The impact supplement contains five questions: difficulties “upset or distress me”; “interfere with home life”; “interfere with friendships”; “interfere with classroom learning”; and “interfere with leisure activities.” Each item has four options and ranges from 0–2, 0 indicating “not at all” or “only a little, with 1 indicating “quite a lot” and 2 indicating “a great deal.” The impact scores can be used as continuous variables or could be classified as normal, borderline, or abnormal. A total impact score of 2 or more is abnormal, a score of 1 is borderline, and a score of 0 is normal. For more details, see: <http://www.sdqinfo.com/>.

current diagnostic classification system. Table 10 presents a different estimation for the SDQ impact supplement and shows that an increase in class size is significantly associated with an increase in these difficulties. Adding the school fixed effect to the estimate results in a loss of the significant power of the coefficient, but the sign is still in the right direction.

8. CONCLUSION

Of the various contributions to the literature on class size effects, some studies have found a significant relation between class size and achievement, and others have not. While the literature remains inconclusive, less is known about the effect of class size on outcomes depending on noncognitive skills, such as well-being and conduct problems. This study investigates the impact of class size on students' cognitive skills, bullying, and violent behavior in West Bank schools. I employ discontinuous regression changes in class size in governmental and UNRWA schools. I also provide evidence that the RDD assumptions are not violated due to discontinuities in the relationship between enrollment and household characteristics at cutoff points induced by a maximum class size rule. With regard to cognitive skills, the results show a small and insignificant effect of class size on pupils' achievements, except in grade six, in which the class size has a notable adverse impact on the students' results.

Table 10: SDQ Impact Supplement and Class Size

Dep. var.: SDQ impact supplement			
	(1)	(2)	(3)
Outcome classification	Continuous variable	Borderline	Abnormal
Panel A: Without school fixed effect			
Class size	0.0141** (0.0056)	0.00372*** (0.0012)	0.00269** (0.0011)
Mean of dep. var.	0.93	0.221	0.153
Std err. of dep. var.	1.833	0.415	0.36
Observations	3,295	3,296	3,297
Enrollment	YES	YES	YES
Enrollment squared	YES	YES	YES
Controls	YES	YES	YES
School FE	NO	NO	NO
Panel A: With school fixed effect			
Class size	0.010 (0.0078)	0.00288 (0.0018)	0.00205 (0.0014)
Mean of dep. var.	0.930	0.221	0.154
Std err. of dep. var.	1.833	0.415	0.361
Observations	3,488	3,488	3,488
Enrollment	YES	YES	YES
Enrollment squared	YES	YES	YES
Controls	YES	YES	YES
School FE	YES	YES	YES

Notes: The robust standard errors clustered at the school level are reported in parentheses. The controls include individual-level controls, sex, dummies for the month of birth and refugee status, household income, father's and mother's education, the standard of living scale (with nine points), the number of books at home, the number of rooms above the median per locality, the number of siblings, and the student's grade. The locality-level controls contain the poverty rate, the proportion of the locality under area C, the presence of the separation wall, the population in 2013, the locality type (dummy for urban), and the school GPA for the 2010/11 and 2011/12 academic years. The school-level controls are the school resource scale, school learning atmosphere and problems among teachers, and school type (UNRWA or governmental). *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

Additionally, the findings show no evidence in favor of heterogeneous class size effects on subgroups. Only students with low performance records benefit from a reduction in class size. The second part of this study concludes that class size reduction contributes to mitigating conduct problems among students measured by self-reported indication of being a victim of bullying and engagement in violent behavior. The main estimations suggest that a ten-student class size reduction will decrease the probability of a student being bullied by 13 percent and the number of instances of fighting during the past 12 months by 0.10 on average. Finally, the dataset provides evidence that

peer relations and data from the SDQ impact supplement questions are a potential mechanism underlying the relationship between students' conduct problems and class size.

Schools are often reluctant to implement a class size reduction policy, arguing that the increase in the overall educational costs cannot justify the small or nonexistent impact on student achievement. Costs notwithstanding, recall that a lack of noncognitive skills, such as conduct, can hurt school performance and the academic process as a whole. The findings presented above demonstrate that wider contextual characteristics, namely being bullied and fighting, are alleviated by a smaller class size. Therefore, the long-run return to mitigating behavioral problems with respect to student achievement would offset the incremental costs of decreasing the class size. The costs and benefits of a class size reduction policy should be weighed carefully, especially if the school resources are limited, the school suffers from ill-behaved pupils, or pupils come from a disadvantaged background. Much of the gain from a class size reduction policy depends on student and teacher characteristics in the contextual setting.

REFERENCES

- Abdeen, Z., A. Brunstein-Klomek, O. Nakash, N. Shibli, M. Nagar, H. Agha, S. Hallaq, Y. Kanat-Maymon, H. Juerges, I. Levav, and R. Qasrawi. 2018. "The association between political violence and the connection between bullying and suicidality among Palestinian youth." *Suicide and Life Threatening Behavior* 48(1): 95–104.
- Akerhielm, K. 1995. "Does class size matter?" *Economics of Education Review* 14(3): 229–41.
- Altinok, N., and G. Kingdon. 2012. "New evidence on class size effects: A pupil fixed effects approach." *Oxford Bulletin of Economics and Statistics* 74(2): 203–34.
- Angrist, J. D., and V. Lavy. 1999. "Using Maimonides' rule to estimate the effect of class size on scholastic achievement." *The Quarterly Journal of Economics* 114(2): 533–75.
- Angrist, J. D., E. Battistin and D. Vuri. 2017. "In a small moment: Class size and moral hazard in the Italian Mezzogiorno." *American Economic Journal: Applied Economics* 9(4):216–49.
- Asadullah, M. N. 2005. "The effect of class size on student achievement: Evidence from Bangladesh." *Applied Economics Letters* 12(4): 217–21.
- B'Tselem. 2012. "The long term impact of Israel's separation barrier in the West Bank." Technical report. Jerusalem: B'Tselem: The Israeli Information Center for Human Rights in the Occupied Territories. Available at: https://www.btselem.org/publications/summaries/201210_arrested_development
- Bedard, K., and E. Dhuey. 2006. "The persistence of early childhood maturity: International evidence of long-run age effects." *The Quarterly Journal of Economics* 121(4): 1437–72.
- Bingley, P., V. Myrup Jensen, and I. Walker. 2005. "The effects of school class size on length of post-compulsory education: some cost-benefit analysis." IZA Discussion Paper No. 1605. Bonn, Germany: Institute for the Study of Labor (IZA).
- Bonesrønning, H. 2003. "Class size effects on student achievement in Norway: Patterns and explanations." *Southern Economic Journal* 69(4): 952–65.
- Browning, M., and E. Heinesen. 2007. "Class size, teacher hours and educational attainment." *Scandinavian Journal of Economics* 109(2): 415–38.
- Calonico, S., M. D. Cattaneo, and R. Titiunik. 2014. "Robust nonparametric confidence intervals for regression-discontinuity designs." *Econometrica* 82(6): 2295–326.
- Carlsson, M., G. B. Dahl, B. Öckert, and D. O. Rooth. 2015. "The effect of schooling on cognitive skills." *Review of Economics and Statistics* 97(3): 533–47.

- Cascio, E., and D. W. Schanzenbach. 2007. "First in the class? Age and the education production function." NBER Working Paper No. w13663. Cambridge, MA: National Bureau of Economic Research (NBER).
- Cattell, R.B., and A. K. S. Cattell. 1960. *Measuring intelligence with the culture fair tests*. Champaign, IL: Institute for Personality and Ability Testing.
- Danny, C. Z., M. Gradstein, and E. Reuven. 2009. "Class Size and the Regression Discontinuity Design: The Case of Public Schools." IZA Working Paper No. 4679. Bonn, Germany: Institute for the Study of Labor.
- Dee, T. S., and M. R. West. 2011. "The non-cognitive returns to class size." *Educational Evaluation and Policy Analysis* 33(1): 23–46.
- Denny, K., and V. Oppedisano. 2013. "The surprising effect of larger class sizes: Evidence using two identification strategies." *Labour Economics* 23: 57–65.
- Di Maio, M., and T. K. Nandi. 2013. "The effect of the Israeli–Palestinian conflict on child labor and school attendance in the West Bank." *Journal of Development Economics* 100(1):107–16.
- Eriksen, T. L., H. S. Nielsen, and M. Simonsen. 2012. "The effects of bullying in elementary school." IZA Discussion Paper No. 6718. Bonn, Germany: Institute for the Study of Labor (IZA).
- Fredriksson, P., B. Öckert, and H. Oosterbeek. 2012. "Long-term effects of class size." *The Quarterly Journal of Economics* 128(1): 249–85.
- Gary-Bobo, R.J., and M. B. Mahjoub. 2013. "Estimation of class-Size effects, using 'Maimonides' Rule' and other instruments: The case of French junior high schools." *Annals of Economics and Statistics/ANNALES D'ÉCONOMIE ET DE STATISTIQUE* 111-112: 193–25.
- Goodman, A., D. L. Lamping, and G. B. Ploubidis. 2010. "When to use broader internalising and externalising subscales instead of the hypothesised five subscales on the Strengths and Difficulties Questionnaire (SDQ): data from British parents, teachers and children." *Journal of Abnormal Child Psychology* 38(8): 1179–91.
- Goodman, R. 1997. "The Strengths and Difficulties Questionnaire: a research note." *Journal of Child Psychology and Psychiatry* 38(5): 581–86.
- . 1999. "The extended version of the Strengths and Difficulties Questionnaire as a guide to child psychiatric caseness and consequent burden." *The Journal of Child Psychology and Psychiatry and Allied Disciplines* 40(5): 791–99.
- Hofstede, G. 1983. "National cultures in four dimensions: A research-based theory of cultural differences among nations." *International Studies of Management and Organization* 13(1-2): 46–74.

- Hoxby, C. M. 2000. "The effects of class size on student achievement: New evidence from population variation." *The Quarterly Journal of Economics* 115(4): 1239–85.
- Jabr, D., and S. Cahan. 2014. "Schooling effects on cognitive development in a difficult environment: the case of refugee camps in the West Bank." *International Studies in Sociology of Education* 24(2): 165–88.
- Jakobsson, N., M. Persson, and M. Svensson. 2013. "Class-size effects on adolescents' mental health and well-being in Swedish schools." *Education Economics* 21(3): 248–63.
- Konstantopoulos, S. 2008. "Do small classes reduce the achievement gap between low and high achievers? Evidence from Project STAR." *The Elementary School Journal* 108(4): 275–91.
- Krueger, A. B. 1999. "Experimental estimates of education production functions." *The Quarterly Journal of Economics* 114(2): 497–532.
- Lazear, E. P. 2001. "Educational production." *The Quarterly Journal of Economics* 116(3): 777–803.
- Leuven, E., H. Oosterbeek, and M. Rønning. 2008. "Quasi-experimental estimates of the effect of class size on achievement in Norway." *Scandinavian Journal of Economics* 110(4): 663–93.
- Mansour, H., and D. I. Rees. 2012. "Armed conflict and birth weight: Evidence from the al-Aqsa Intifada." *Journal of Development Economics* 99(1): 190–99.
- MoEHE (Ministry of Education and Higher Education). 2016. "About the Ministry." Ministry of Education and Higher Education—Palestine website. Accessed April 25, 2018. Available at: <http://www.moehe.gov.ps/en/About-the-Ministry/Education-System>
- MoFP (Ministry of Finance and Planning). 2016. "Citizen budget." Technical report. Ramallah, Palestine: Ministry of Finance and Planning, State of Palestine. Available at: <http://www.pmf.ps/documents/10192/654283/2016CitizensBudget.pdf>
- Moshoeshoe, R. 2015. "Average and Heterogeneous Effects of Class Size on Educational Achievement in Lesotho." Economic Research Southern Africa Working Paper No. 496. Cape Town: Economic Research Southern Africa.
- Moura, D. R. D., A. C. N. Cruz, and L. D. Á. Quevedo. 2011. "Prevalence and characteristics of school age bullying victims." *Jornal de Pediatria* 87(1): 19–23.
- Nisbett, R. E. 2009. *Intelligence and how to get it: Why schools and cultures count*. New York: WW Norton and Company.
- Olweus, D. 1997. "Bully/victim problems in school: Facts and intervention." *European Journal of Psychology of Education* 12(4):495.

- Ortar, G., and A. Shachor. 1980. *MILTA: A battery of tests for ages 9 through 18*. Jerusalem: Ministry of Education and Culture (in Hebrew).
- PCBS (Palestinian Central Bureau of Statistics). 2010. "Migration survey in the Palestinian territory." Technical report. Ramallah, Palestine: Palestinian Central Bureau of Statistics. Available at: http://www.pcbs.gov.ps/Portals/_pcbs/PressRelease/Migration_e.pdf
- . 2013. "Palestinian children, cases and statistics." Technical report. Ramallah, Palestine: Palestine Central Bureau of Statistics (in Arabic). Available at: http://www.pcbs.gov.ps/Portals/_PCBS/Downloads/book1969.pdf
- . 2017. "Status of the rights of Palestinian children. 2016." Technical report. Ramallah, Palestine: Palestinian Central of Bureau and Statistics. Available at: <http://www.pcbs.gov.ps/Downloads/book2275.pdf>
- Raven, J.C. 1983. *The standard progressive matrices, 1938–83*. New York: Psychological Corporation.
- Sarzosa, M., and S. Urzúa. 2015. "Bullying among adolescents: the role of cognitive and non-cognitive skills." NBER Working Paper No. w21631. Cambridge, MA: National Bureau of Economic Research.
- Smith, P.K., R. Catalano, P. Slee, Y. Morita, J. Junger-Tas, and D. Olweus (eds.). 1999. *The nature of school bullying: A cross-national perspective*. Hove, UK: Psychology Press.
- Tani, F., P. S. Greenman, B. H. Schneider, and M. Fregoso. 2003. "Bullying and the Big Five: A study of childhood personality and participant roles in bullying incidents." *School Psychology International* 24(2): 131–46.
- Thorndike, R. L., E. P. Hagen, and I. Lorge. 1971. *Cognitive abilities test: Kindergarten-grades 12/13. Multi-level ed., levels AH (grades 3–13). Primary batteries 1 u. 2 (grades K–3)*. Boston: Houghton Mifflin.
- UNESCO (United Nations Educational, Scientific and Cultural Organization). 2011. "World data on education, 7th edition 2010/11." Technical report. Paris: UNESCO. Available at: http://www.ibe.unesco.org/fileadmin/user_upload/Publications/WDE/2010/pdf-versions/Palestine.pdf
- UNRWA (United Nations Relief and Works Agency for Palestine Refugees in the Near East). 2014. "Educational research briefs publication plan." Technical report. Amman, Jordan: UNRWA. Available at: https://www.unrwa.org/sites/default/files/research_and_development_unit_research_briefs_-_2014_english.pdf

- UNSCO (Office of the United Nations Special Coordinator for the Middle East Peace Process). 2014. "Report to the Ad Hoc Liaison Committee New York." Technical report, September 22. New York: UNSCO. Available at: https://unsco.unmissions.org/sites/default/files/un_ahlc_report_sept_2014.pdf
- Urquiola, M. 2006. "Identifying class size effects in developing countries: Evidence from rural Bolivia." *Review of Economics and Statistics* 88(1): 171–77.
- Urquiola, M., and E. Verhoogen. 2009. "Class-size caps, sorting, and the regression-discontinuity design." *American Economic Review* 99(1): 179–215.
- Van der Klaauw, W. 2002. "Estimating the effect of financial aid offers on college enrollment: A regression–discontinuity approach." *International Economic Review* 43(4): 1249–87.
- Vishwanath, T., B. Blankespoor, F. Calandra, N. Krishnan, M. Mahadevan, and M. Yoshida. 2014. "Seeing believes: Poverty in the Palestinian territories." Technical report. Washington, DC: World Bank Group.
- Woessmann, L., and M. West. 2006. "Class-size effects in school systems around the world: Evidence from between-grade variation in TIMSS." *European Economic Review* 50(3):695–736.

APPENDIX

Table A.1: OLS Regression with Class Size as the Dependent Variable

Dep. var: Average class size	(1)	(2)	(3)	(4)	(5)	(6)
Enrollment	0.125*** (0.00250)	0.134*** (0.016)	0.117*** (0.015)	0.128*** (0.015)	0.127*** (0.015)	0.041 (0.028)
Female child		0.029 (0.878)	0.496 (0.810)	-0.113 (0.853)	-0.103 (0.855)	-1.119*** (0.164)
Refugee dummy		-1.102 (0.813)	-0.851 (0.673)	-0.228 (0.824)	-0.286 (0.824)	-0.433 (0.284)
Fathers' schooling						
Less than 12 years		-0.192 (0.661)	-0.293 (0.607)	-0.265 (0.590)	-0.226 (0.605)	-0.448 (0.436)
12 years		0.120 (0.660)	-0.007 (0.610)	-0.117 (0.598)	-0.009 (0.618)	-0.188 (0.442)
College degree		-0.192 (0.652)	-0.271 (0.602)	-0.462 (0.583)	-0.399 (0.592)	-0.484 (0.421)
Mothers' schooling						
Less than 12 years		-0.453 (0.725)	-0.707 (0.687)	-0.275 (0.672)	-0.341 (0.689)	-0.331 (0.497)
12 years		-0.221 (0.814)	-0.461 (0.758)	0.021 (0.697)	-0.047 (0.714)	-0.122 (0.509)
College degree		-0.343 (0.871)	-0.495 (0.809)	-0.176 (0.760)	-0.331 (0.769)	-0.491 (0.540)
Living standard scale (10 points)		0.093 (0.086)	0.105 (0.075)	0.074 (0.070)	0.106 (0.068)	0.024 (0.049)
Net monthly income (NIS)		-0.264 (0.179)	-0.209 (0.136)	-0.174 (0.113)	-0.175 (0.112)	-0.061 (0.058)
Number of books at home		-0.041 (0.130)	-0.114 (0.124)	-0.134 (0.117)	-0.103 (0.118)	-0.119 (0.079)
No. of student's siblings		-0.044 (0.067)	-0.002 (0.050)	-0.013 (0.047)	0.004 (0.047)	0.005 (0.028)
Number of rooms (above median)		1.110** (0.437)	0.736* (0.375)	0.655* (0.355)	0.655* (0.360)	0.127 (0.216)
Presence of the separation wall			1.725 (1.074)	1.758* (0.982)	1.776* (0.982)	8.434*** (2.004)
Percentage of the locality in area C			-4.467* (2.564)	-4.296 (2.624)	-4.403* (2.622)	-2.775** (1.323)
Poverty headcount rate			-11.888*** (3.238)	-12.870*** (3.086)	-12.862*** (3.098)	24.597*** (3.615)
Population per locality			0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)
Locality type (urban)			-1.227 (1.066)	-1.932* (1.100)	-1.943* (1.101)	3.887*** (0.649)

Continued ...

Table A.1: Continued

UNRWA school				-2.203 (1.351)	-2.127 (1.353)	
School lack of resources (18-point scale)				-0.170** (0.085)	-0.169* (0.085)	
School educational atmosphere total (40-point scale)				0.112 (0.138)	0.112 (0.138)	
Problems among teachers (6-point scale)				-0.627** (0.252)	-0.623** (0.251)	
Student grade						
Student in grade 6					-0.350 (0.759)	-0.125 (0.752)
Student in grade 7					-0.508 (0.757)	-0.131 (0.759)
Student in grade 8					-0.921 (0.756)	-0.671 (0.755)
Student in grade 9					-2.040** (0.873)	-2.205** (0.869)
School GPA 2011						-0.006 (0.017)
School GPA 2012						0.007 (0.016)
Constant	24.28*** (0.172)	24.815*** (1.258)	28.787*** (1.587)	28.325*** (3.809)	28.809*** (3.755)	15.899*** (1.576)
Observations	5,106	5,106	5,106	5,106	5,106	5,080
R-squared	0.315	0.326	0.380	0.414	0.422	0.669
Individual-level controls	NO	YES	YES	YES	YES	YES
Locality-level controls	NO	NO	YES	YES	YES	YES
School-level controls	NO	NO	NO	YES	YES	NO
Cohort FE	NO	NO	NO	NO	YES	YES
School GPA in 2011/12	NO	NO	NO	NO	NO	YES
School FE	NO	NO	NO	NO	NO	YES

Note: The robust standard errors clustered at the school level are reported in parentheses. *** p<0.01, ** p<0.05, and * p<0.

Table A.2: Class Size and Students' Cognitive Test Scores

Dep. var.: Cognitive tests	(1)	(2)	(3)	(4)	(5)
Average class size	-0.007 (0.100)	0.053 (0.084)	-0.047 (0.072)	-0.096 (0.069)	-0.075 (0.066)
Enrollment	0.007 (0.026)	-0.027 (0.020)	-0.012 (0.022)	-0.015 (0.023)	-0.027 (0.022)
Female child		7.691*** (1.812)	8.167*** (1.709)	7.299*** (1.496)	6.193*** (1.387)
Refugee status		3.884*** (1.363)	4.545*** (1.487)	1.710 (1.075)	1.950* (1.010)
Student grade					
Student in grade 6		6.709*** (0.840)	6.626*** (0.820)	6.701*** (0.811)	9.270*** (0.732)
Student in grade 7		10.917*** (0.812)	10.984*** (0.791)	11.037*** (0.790)	13.970*** (0.713)
Student in grade 8		14.028*** (0.984)	13.994*** (0.973)	13.969*** (0.974)	17.975*** (0.970)
Student in grade 9		16.413*** (0.905)	16.323*** (0.894)	16.254*** (0.912)	20.463*** (0.951)
Fathers' schooling					
Less than 12 years		0.175 (1.910)	0.053 (1.882)	-0.012 (1.855)	-0.507 (1.749)
12 years		2.408 (1.941)	1.985 (1.884)	1.894 (1.894)	-0.057 (1.752)
College degree		4.031** (1.946)	3.599* (1.907)	3.346* (1.917)	-0.036 (1.811)
Mothers' schooling					
Less than 12 years		-1.346 (1.748)	-1.395 (1.697)	-1.715 (1.632)	-0.117 (1.502)
12 years		1.017 (1.828)	0.688 (1.748)	0.277 (1.707)	0.481 (1.594)
College degree		2.661 (2.003)	2.671 (1.910)	2.053 (1.887)	0.196 (1.734)
Living standard index (10 points)		0.492** (0.218)	0.465** (0.191)	0.434** (0.182)	0.163 (0.169)
Net monthly income (NIS)		0.738*** (0.249)	0.631*** (0.231)	0.736*** (0.227)	0.446** (0.199)
Number of books at home		0.913*** (0.272)	0.882*** (0.263)	0.842*** (0.265)	0.291 (0.237)
No. of student's siblings		-0.740*** (0.167)	-0.595*** (0.157)	-0.535*** (0.150)	-0.203 (0.133)
Number of rooms (above the median)		0.947 (0.769)	0.553 (0.701)	0.374 (0.718)	-0.299 (0.595)
Presence of the separation wall			-0.129 (1.420)	0.049 (1.379)	-0.354 (1.278)
Percentage of the locality in area C			-3.978 (3.224)	-3.833 (2.947)	-4.367 (2.906)

Continued ...

Table A.2: Continued

Poverty headcount rate			-30.594***	-28.463***	-23.534***
			(6.064)	(6.236)	(6.245)
Population per locality (2013)			-0.000	0.000	-0.000
			(0.000)	(0.000)	(0.000)
Locality type (urban)			-0.031	-0.259	0.245
			(1.667)	(1.649)	(1.462)
UNRWA school				4.435**	3.785**
				(1.703)	(1.560)
School lack of resources (18-point scale)				0.184	0.178
				(0.148)	(0.138)
School educational atmosphere				0.358*	0.281
				(0.185)	(0.189)
Problems among teachers				-0.528	-0.240
				(0.371)	(0.381)
School GPA 2011					0.195***
					(0.043)
School GPA 2012					0.284***
					(0.040)
Constant	60.484***	38.931***	49.100***	40.288***	9.091
	(2.889)	(4.027)	(4.572)	(7.640)	(7.792)
Observations	4,664	4,664	4,664	4,664	4,642
R-squared	0.000	0.257	0.288	0.305	0.467
Individual-level controls	NO	YES	YES	YES	YES
Locality-level controls	NO	NO	YES	YES	YES
School-level controls	NO	NO	NO	YES	YES
School GPA in 2011/12	NO	NO	NO	NO	YES
Mean of dep. var.	60.65	60.65	60.65	60.65	60.65
Std err. of dep. var.	16.66	16.66	16.66	16.66	16.66

Note: The robust standard errors clustered at the school level are reported in parentheses. *** p<0.01, ** p<0.05, and * p<0.

Table A.3: Within Enrollment Band OLS Estimation, Five-student Interval

Dep. var.: cognitive tests	Cutoff			
	First	Second	First	Second
Panel A: Governmental schools				
	(1)	(2)	(3)	(4)
	(40 students)	(80 students)	(40 students)	(80 students)
Enrollment (N)	$36 \geq N \geq 45$	$76 \geq N \geq 85$	$36 \geq N \geq 45$	$76 \geq N \geq 85$
Class size	-0.121 (0.244)	-0.560 (0.690)	0.271* (0.143)	-41.133* (14.931)
Observations	492	82	429	68
R-squared	0.002	0.051	0.574	0.825
Controls	NO	NO	YES	YES
Panel B: UNRWA schools				
	(45 students)	(90 students)	(45 students)	(90 students)
Enrollment (N)	$41 \geq N \geq 50$	$86 \geq N \geq 95$	$41 \geq N \geq 50$	$86 \geq N \geq 95$
Class size	0.353 (0.303)	-0.248 (0.496)	0.308*** (0.026)	0.320 (0.386)
Observations	166	216	164	202
R-squared	0.036	0.008	0.662	0.652
Controls	NO	NO	YES	YES

Notes: The robust standard errors clustered at the school level are reported in parentheses. The controls include sex, dummies for the month of birth and refugee status, household income, father's and mother's education, the standard of living scale (with nine points), the number of books at home, the number of rooms above the median per locality, the number of siblings, and the student's grade. The locality-level controls contain the poverty rate, the proportion of the locality under area C, the presence of the separation wall, the population in 2013, and the locality type (dummy for urban). The school-level controls are the school resource scale, school learning atmosphere, and problems among teachers. The school GPA includes the school GPA for the 2010/11 and 2011/12 academic years. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.4: Schools That Comply/Do Not Comply with the Maximum Class Rule Characteristics

	Obs.	Mean	S.D.	Min.	Max
Panel A: Schools that comply with the maximum class size rule					
Cognitive tests score (child)	4664	60.62	11.28	27.98	85.31
Cognitive tests score (class)	4664	60.65	16.66	0.55	94.48
Class size	4664	30.59	7.02	6	44.5
UNRWA school	4664	0.45	0.50	0	1
School lack of resources (18-point scale)	4664	6.02	4.47	0	17
School educational atmosphere total (40-point scale)	4664	26.95	3.54	17	38
Problems among teachers (6-point scale)	4664	1.73	1.64	0	6
Panel B: Schools that do not comply with the maximum class size rule					
Cognitive tests score (child)	473	59.69	12.01	31.77	75.97
Cognitive tests score (class)	473	59.69	17.01	2.76	90.06
Class size	473	43.41	3.69	40	56
UNRWA school	473	0.00	0.00	0	0
School lack of resources (18-point scale)	473	5.82	4.76	0	15
School educational atmosphere total (40-point scale)	473	26.40	2.51	22	31
Problems among teachers (6-point scale)	473	1.74	1.68	0	6

Table A.5: RD Regressions for Pretreatment Covariates

Dep. var.	First	Second	First	Second	First	Second	First	Second
	Mother's edu.	Mother's edu.	Living standard	Living standard	Mother's edu.	Mother's edu.	Living standard	Living standard
	Panel A: Governmental schools				Panel B: UNRWA Schools			
Cutoff value	40	80	40	80	45	90	45	90
RD_ estimate	0.284	0.0320	0.437	0.647	0.0617	0.145	-0.106	0.208
Effective observations	681	504	724	723	348	703	357	835
Left of c	498	381	529	558	165	340	167	366
Right of c	183	123	195	165	183	363	190	469
Total observations	2253	2253	2404	2404	1739	1739	1797	1797
Robust 95% CI	[-.07; .78]	[-.5; .35]	[-.52; 1.61]	[-.32; 1.74]	[-.38; .61]	[-.14; .47]	[-1.1; 1.06]	[-.33; .85]
Kernel type	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular
BW type	CCT	CCT	CCT	CCT	CCT	CCT	CCT	CCT
Conventional stdererror	0.176	0.179	0.432	0.455	0.215	0.131	0.455	0.257
Conventional p-value	0.107	0.858	0.311	0.155	0.774	0.270	0.815	0.420
Robust p-value	0.100	0.728	0.314	0.176	0.638	0.301	0.968	0.391
Order loc. poly. (p)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Order bias (q)	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000
BW loc. poly. (h)	8.456	23.443	8.638	33.691	10.850	19.960	10.892	24.124
BW bias (b)	13.109	33.322	12.191	52.140	17.472	29.931	15.938	38.509

Table A.6: Effect of Class Size on Bullying

Dep. var:	(1)	(2)	(3)	(4)	(5)	(6)
	Victim		Bully		Bully-victim	
Class size	0.00818*	0.0135**	0.00330	0.0100	-0.000579	0.00290
	(0.00451)	(0.00636)	(0.00600)	(0.00826)	(0.00562)	(0.00722)
Enrollment	0.000276	-0.000434	0.00122	0.000538	0.00100	0.000475
	(0.00115)	(0.00244)	(0.00146)	(0.00400)	(0.00145)	(0.00301)
Constant	0.752*	0.259	0.494	-0.423	0.918**	0.315
	(0.392)	(0.278)	(0.424)	(0.327)	(0.402)	(0.301)
Observations	4,292	4,290	4,224	4,201	4,224	4,201
Individual-level controls	YES	YES	YES	YES	YES	YES
Locality-level controls	YES	YES	YES	YES	YES	YES
School-level controls	YES	NO	YES	NO	YES	NO
School GPA	YES	YES	YES	YES	YES	YES
School FE	NO	YES	NO	YES	NO	YES
Mean of dep. var.	0.232	0.232	0.123	0.122	0.195	0.195
Std err. of dep. var.	0.422	0.422	0.328	0.328	0.396	0.396

Notes: The robust standard errors clustered at the school level are reported in parentheses. The controls include sex, dummies for the month of birth and refugee status, household income, father's and mother's education, the standard of living scale (with nine points), the number of books at home, the number of rooms above the median per locality, the number of siblings, and the student's grade. The locality-level controls contain the poverty rate, the proportion of the locality under area C, the presence of the separation wall, the population in 2013, and the locality type (dummy for urban). The school-level controls are the school resource scale, school learning atmosphere, and problems among teachers. The school GPA includes the school GPA for the 2010/11 and 2011/12 academic years. *** p<0.01, ** p<0.05, and * p<0.1.

Table A.7: OLS Estimation (quartile distribution)

Dep. var.: Cognitive tests	(1)	(2)	(3)
	Distribution		
	Top 25%	Top 50%	Top 75%
Average class size	-0.058 (0.038)	-0.019 (0.034)	-0.077** (0.039)
Enrollment	-0.031*** (0.012)	-0.028** (0.012)	-0.024** (0.009)
Constant	-7.835** (3.555)	7.004** (3.345)	25.330*** (3.552)
Observations	4,642	4,642	4,642
Mean of dep. var.	60.65	60.65	60.65
Std err. of dep. var.	16.66	16.66	16.66
Individual-level controls	YES	YES	YES
Locality-level controls	YES	YES	YES
School-level controls	YES	YES	YES
School GPA in 2011/2012	YES	YES	YES

Notes: The robust standard errors clustered at the school level are reported in parentheses. The controls include sex, dummies for the month of birth and refugee status, household income, father's and mother's education, the standard of living scale (with nine points), the number of books at home, the number of rooms above the median per locality, the number of siblings, and the student's grade. The locality-level controls contain the poverty rate, the proportion of the locality under area C, the presence of the separation wall, the population in 2013, and the locality type (dummy for urban). The school-level controls are the school resource scale, school learning atmosphere, and problems among teachers. The school GPA includes the school GPA for the 2010/11 and 2011/12 academic years. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

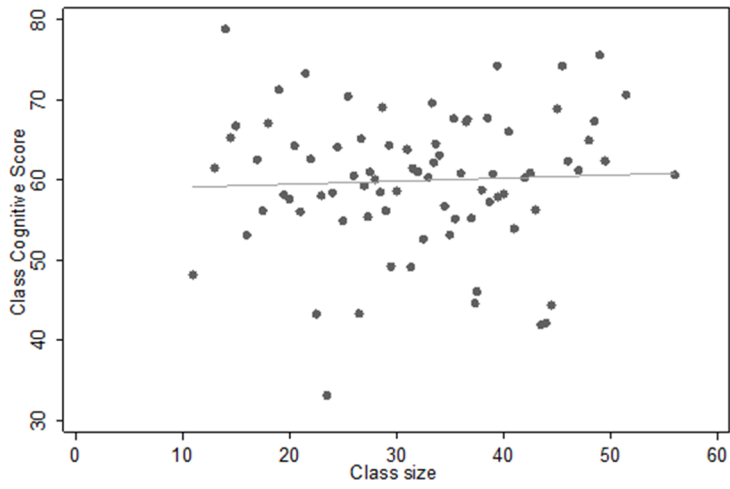
Table A.8: Class Size and Students' Cognitive Test Scores (all observations)

Dep. var.: Cognitive tests	(1)	(2)	(3)	(4)	(5)
Average class size	-0.042 (0.094)	0.035 (0.080)	-0.055 (0.076)	-0.062 (0.075)	-0.062 (0.071)
Enrollment	0.017 (0.025)	-0.020 (0.021)	-0.006 (0.022)	-0.013 (0.022)	-0.023 (0.021)
Constant	60.824*** (2.889)	39.430*** (4.056)	49.138*** (4.476)	38.411*** (7.474)	7.815 (7.695)
Observations	5,106	5,106	5,106	5,106	5,080
R-squared	0.001	0.257	0.289	0.304	0.460
Mean of dep. var.	60.65	60.65	60.65	60.65	60.65
Std err. of dep. var.	16.66	16.66	16.66	16.66	16.66
Individual-level controls	NO	YES	YES	YES	YES
Locality-level controls	NO	NO	YES	YES	YES
School-level controls	NO	NO	NO	YES	YES
School GPA in 2011/2012	NO	NO	NO	NO	YES

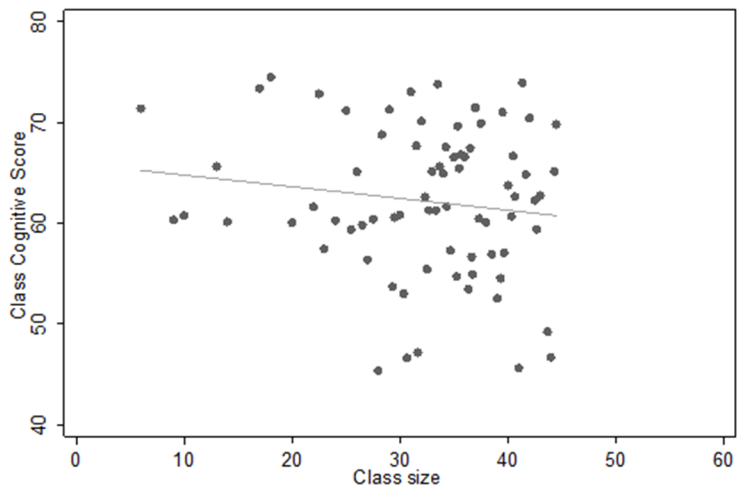
Notes: The robust standard errors clustered at the school level are reported in parentheses. The individual-level controls include sex, dummies for the month of birth and refugee status, household income, father's and mother's education, the standard of living scale (with 9 points), the number of books at home, the number of rooms above the median per locality, the number of siblings, and the student's grade. The locality-level controls contain the poverty rate, the proportion of the locality under area C, the presence of the separation wall, the population in 2013, and the locality type (dummy for urban). The school-level controls are the school resource scale, school learning atmosphere and problems among teachers, and school type (UNRWA or governmental). The school GPA includes the school GPA for the 2010/11 and 2011/12 academic years. *** p<0.01, ** p<0.05, and * p<0.1.

Figure A.1.a: Class Size and Average Class Cognitive Tests Scores Relationship

Panel A: Governmental Schools

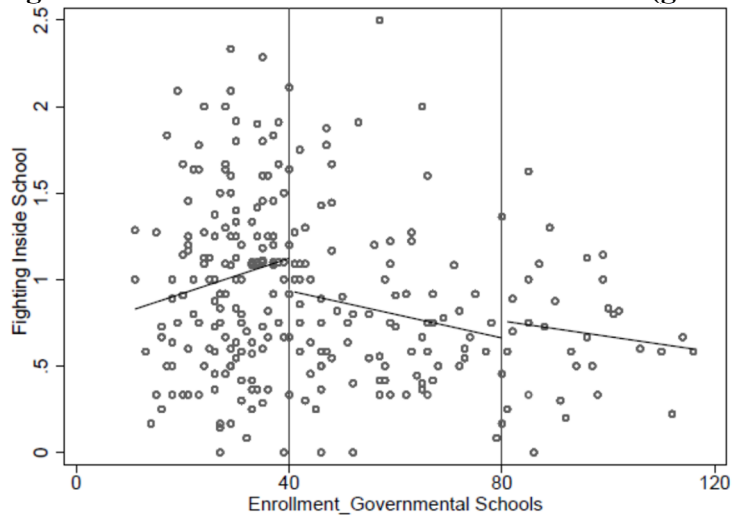


Panel A.1.b: UNRWA Schools



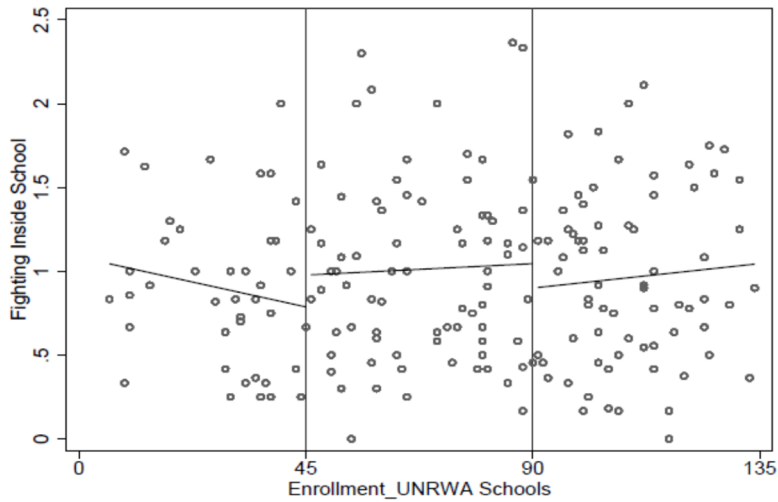
Source: Own calculations based on survey data.

Figure A.2.a: Enrollment and Violent Behavior (governmental schools)



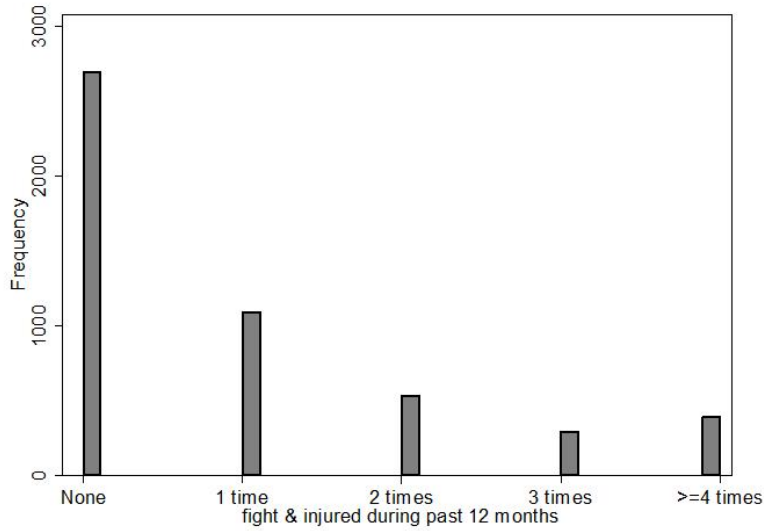
Source: Own calculations based on HBSC survey and school enrollment

Figure A.2.b: Enrollment and Violent Behavior (UNRWA schools)



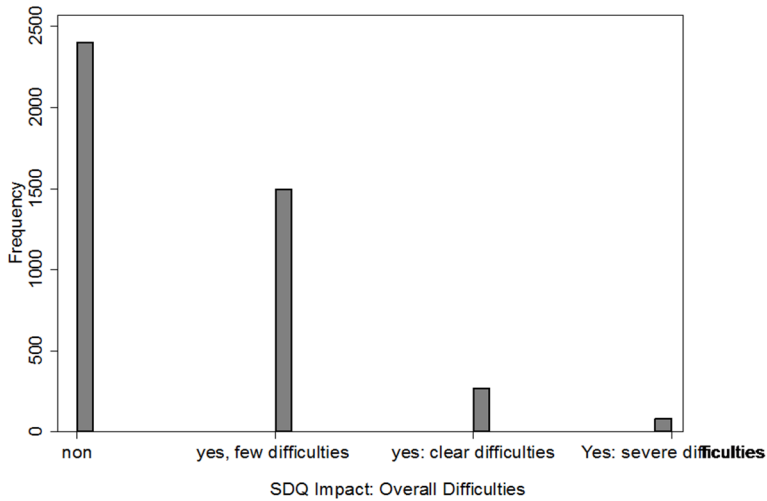
Source: Own calculations based on HBSC survey and school enrollment.

Figure A.3: Violent Behavior Distribution



Source: Own calculations based on the HBSC survey.

Figure A.4: SDQ Follow-up Question



Source: Own calculations based on the parental questionnaire.

Note: Mothers/parents were asked: “In general, do you think that your child has difficulties in one or more of the following areas: emotion, concentration, behavior, harmony, or dealings with others?”