

DISCUSSION PAPER SERIES

IZA DP No. 12628

**Does Class Size Matter in Postgraduate
Education?**

Alessio Gaggero
Getinet Haile

SEPTEMBER 2019

DISCUSSION PAPER SERIES

IZA DP No. 12628

Does Class Size Matter in Postgraduate Education?

Alessio Gaggero

University of Nottingham

Getinet Haile

University of Nottingham and IZA

SEPTEMBER 2019

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Does Class Size Matter in Postgraduate Education?*

The paper examines the impact of class size on postgraduate grades using administrative data from one of the largest Schools of a Russell Group University in the UK. As well as estimating Fixed Effects models on the population of postgraduate candidates in the School, we exploit a policy change aimed at reducing class size to implement a regression discontinuity design (RDD). We find that class size does impact grades adversely overall; and the policy aimed at reducing class size impacts grades favourably. Our findings are robust to alternative specifications as well as being supported by the validity tests we conducted.

JEL Classification: I21, I23

Keywords: higher education, class size, grades, RDD

Corresponding author:

Alessio Gaggero
University of Nottingham
South Building, Jubilee Campus
Wollaton Road
Nottingham, NG8 1BB
United Kingdom
E-mail: alessio.gaggero@nottingham.ac.uk

* The authors would like to extend their thanks to the data Department of the anonymous UK University for the administrative data used in this paper. The usual disclaimer applies.

1. Introduction and Motivation

The issue of class-size and its effect on student learning has been extensively studied in primary and secondary school settings. The weight of the evidence to date supports the view that smaller classes promote student learning (see Angrist and Lavy 1999 for extensive review). In the context of tertiary education, where independent learning is a major part of the education landscape, class-size may not be as important. However, this may not mean that the question of class-size is altogether unimportant in this context given the resource implications of larger class-sizes, including the ease with which students access their professors. Such concerns may be particularly valid given two important recent developments in the higher education sector in the UK and elsewhere in the OECD. First, there has been a significant rise in tertiary education, which is driven by supply side policies in these countries as Bandiera *et al.* (2010) noted. Secondly, there have been changes in the higher education funding environment, particularly following the great recession of 2008, with a view to sustainable financing of higher education. The Government abolished paying teaching grants to universities in September 2012, and the cap on tuition fees has been raised significantly to make up for funding shortfalls (Crawford *et al.* 2014). The changes mean that most universities now source a significant part of their funding from tuition fees.¹ This has made the postgraduate sector, particularly the more lucrative international postgraduate market, a lot more attractive to universities generally and, in particular, the Russell Group universities given their established international reputation.

There is extensive literature linking class-size to test scores in schools (Kruger, 1999; Kruger and Whitmore, 2001; Angrist and Lavy, 1999; Browning and Heinesen, 2007; Leuven *et al.*, 2008; Hanushek, 1979; Hoxby, 2000; Case and Deaton, 1999). However, there is a dearth of evidence relating to tertiary education. De Paola and Scoppa (2011), Monks and Schmidt (2011), Bandiera *et al.* (2010), Kokkelenberg *et al.* (2008) and De Paola *et al.* (2013) are some of the few recent studies, which found class-size having a negative effect on college scholastic outcomes reaffirming earlier findings in Gibbs *et al.* (1996) in the context of tertiary education, which found that students in larger classes perform less well. In a recent study Huxley *et al.* (2018) have reported significant variation in teaching intensity across higher education in the UK, which they attribute to variation in class-size. Examining the link

¹ Barr and Turner (2013) dwell on these in the context of the US, which they describe as growing conflict between expanding enrolments in postsecondary education and contracting public budget support. The independent Browne Review also dealt with the issue of sustainable HE funding in the UK with the recommendation that more of the burden of funding HE be placed on graduates.

between tertiary-level class-size and test score is therefore a worthy exercise particularly given the on-going debate on the future of higher education funding and the recent recommendation for reducing the cap on tuition fee.²

In this paper, we provide evidence of the link between class-size and postgraduate grades. Similarly to Bandiera *et al.* (2010), we first estimate this relationship by means of student fixed effects regressions. The outcome of this approach, however, may still suffer given the possibility that students self-select into modules for reasons that are unobservable to the researcher. In an ideal experiment, one would control for such unobservables by randomly allocating students into modules of different sizes. The empirical design used in this paper, mimics this ideal experiment by exploiting a recent policy change, which is aimed at maintaining a high standard of teaching, that envisaged double teaching (*double-up*, hereinafter) on the basis of the specific number of students enrolled to a certain module. Specifically, module convenors with enrolment size above a certain level were expected to split students into two groups and double teach the module content as a result of the policy. Our use of the discontinuous module enrolment function and the policy change to examine the link between class-size and postgraduate grades is likely to yield superior instrumental variables estimates.³

We use administrative data of a postgraduate programme of one of the largest Schools of a Russell Group public university in the UK to examine the impact of the class-size policy on postgraduate grades in a Regression Discontinuity Design (RDD). Our empirical results echo earlier research and suggest that indeed class size matter significantly for postgraduate education. Specifically, our RDD estimates suggest that students exposed to the double-up policy earned significantly higher grades vis-à-vis their counterparts who were not affected by the double teaching policy. We find these results to be strongest for British students. Moreover, our results suggest that the double-up policy significantly reduced the probability of students failing in their module. Importantly, our results are robust to a variety of specifications. The RDD estimates are insensitive to the inclusion (or exclusion) of various control variables and different functional form specifications of the running variable. More importantly, we demonstrate that our results are unlikely to be driven by discontinuities in pre-intervention individual characteristics, or endogenous sorting around the threshold.

² See the May 2019 Auger Review report.

³ Angrist and Lavy (1999) use a similar approach to study the effect of class size on grades in Israeli elementary schools.

The question of class-size is not yet well understood in the context of postgraduate education, but these developments are likely to lend some prominence to existing concerns about increasing student-to-staff ratio and worse student performance in large class settings.⁴ The same issues have become increasingly relevant in tertiary education, particularly given the current funding climate in higher education institutions. Examining the link between postgraduate class-size and postgraduate scholastic outcomes is therefore of significant policy interest both for the higher education sector and postgraduate students alike, which this paper aims to achieve.

The remainder paper is organised as follows. Section 2 describes the data used and institutional setting. Section 3 outlines the identification strategy used and discusses the results. Section 4 presents the sensitivity analysis before the final Section concludes the paper.

2. Data and Institutional Background

We use administrative data covering the population of all enrolees in the postgraduate (PG) programme of one of the largest Schools of a Russell Group public university in the UK for the academic year 2017/18. The PG programme has some 16 specialities in total and recruitment to the programme depends on applicants' prior academic achievement at an undergraduate level, which is typically a strong 2.1 or equivalent, language proficiency in the case of international applicants and candidates' character references. Candidates join one of 15 speciality areas once enrolled. The programme requires students to attend a set of core and elective modules. A candidate's performance in modular final examinations leads to the award of modular scores that range between 0 and 100.⁵ The scores obtained are typically averaged across programme modules to yield overall postgraduate grade, which gets translated into four distinct degree classifications: distinction (>70%), merit/credit (60-70%), pass (50-60%) and fail (<50%).

The school's success in attracting increasing number of postgraduate applicants in recent years and the desire to maintain a high standard of teaching prompted the school to pursue a policy of double teaching (*double-up* hereinafter) since 2008 depending on module enrolment size. The policy stipulates that module convenors with enrolment size above a certain level

⁴ The Dearing's Review highlights these issues. Also see Figure A. 1 on trends in PG students in the study institution.

⁵ We expect the modular grade to be largely, if not entirely, objective outcome measure particularly given the type of marking arrangements in the school, where exam scripts are marked by the professor concerned before they go through two additional layers of checks/moderations by internal and external moderators.

split students into two groups and double teach.⁶ The policy has evolved over the years with varying cut-off points for triggering the double-up. Since 2016 the recommended cut-off point has been 110 students per module, so that a module with enrolment size in excess of 110 students become candidate module to double teach by splitting students enrolled on the module into two (or more) smaller classes.

The study sample includes 987 full-time MSc students, who could attend up to 10 different modules during the academic year, yielding a total of 7,696 student-module observations. The data includes information on students' modular grades, the number of students enrolled in each module and a set of student attributes including age, gender, and nationality. Table 1 reports basic summary statistics, which indicates the typical postgraduate student being 23 years of age and 70% of them being females.⁷ Academically, students in the sample achieve an overall mark of 61%, on average, with 20.1%, 42.5% and 34.1% of the students achieving a distinction, merit and pass degree classifications, respectively, while 3.4% failed.

[Table 1 about here]

3. Econometric Methods and Results

3.1 Baseline Analysis

The paper first conducts a baseline analysis on the link between class-size and grades following Hanushek (1979) and using the following panel data model:

$$Y_{i,m} = \alpha_i + \beta S_m + X'_{i,m}\gamma + \varepsilon_{i,m}, \quad (1)$$

where $Y_{i,m}$ represents the *standardised* module (test) score of student i on module m ; α_i is a student-specific fixed effect; S_m is the number of students enrolled in module m , and β is the main parameter of interest, as it measure the effect of class-size on grades; $X'_{i,m}$ is a set of control variables that might affect students' outcomes, which include student characteristics, such as age, gender, and country of origin, as well as module characteristics, such as a whether the module was core or optional, or whether the module was taught in the spring term. Finally, $\varepsilon_{i,m}$ is a random error term.

⁶ The policy was not compulsory, however, and there are some exceptions to it – for example depending on classroom size, where some modules with large-size lecture halls may opt out of the double-up.

⁷ These are figures comparable to the national average postgraduate student characteristics in England in 2016/17 as compiled by the Higher Education Student Statistics (<https://www.hesa.ac.uk/news/11-01-2018/sfr247-higher-education-student-statistics/numbers>).

[Table 2 about here]

Table 2 reports the descriptive results based on Equation (1). Columns (1) and (2) report estimated coefficients from a pooled regression model, while columns (3) and (4) account for student-specific fixed effects. Overall, the results confirm previous findings in the literature and show a strong and negative effect of class-size on grades, and suggest the presence of nonlinear relationship of the concave type.⁸ This highlights that unobserved student characteristics play a role in determining test scores. Moreover, the results in the Table reveal that girls and British students tend to achieve better modular scores than their male and international counterparts, respectively. The results also suggest that, on average, students tend to perform better in spring term and in optional modules, which is intuitive in some sense given that students are likely to enrol in optional modules that they expect to perform better at.

In order for the β coefficient estimated from Equation (1) to be reliable, it requires the assumption $E[\varepsilon_{i,m} | S_m, X'_{i,m}, \alpha_i] = 0$ to be satisfied. This assumption entails that, conditional on characteristics, $X'_{i,m}$, and student-specific fixed effects, α_i , students select into modules of different sizes randomly. This may be a strong assumption if students were to sort into modules of different sizes according to their preferences and/or their idiosyncratic gains. For example, as noted by Bandiera *et al.* (2010), students may choose modules of smaller sizes to maximise their time spent with professors and minimise their effort into a specific subject. This type of sorting would lead to a downward bias in the estimated coefficients. In this study we tackle this potential issue by exploiting the double-up policy mentioned above in a regression discontinuity design framework.

3.2 Regression Discontinuity Design

We use a Regression Discontinuity Design (RDD) to estimate the effect of the double-up on modular grade. The RDD design was first introduced by Thistlethwaite and Campbell (1960) and then formalised by Hahn *et al.* (2001) who derived the necessary conditions for identification of causal effects. RDD are becoming increasingly popular in empirical studies given that the assumptions needed for identification of causal effects are quite weak. The

⁸ Notice, the difference in sample size is because in this exercise we exclude students who were exposed to the double-up teaching.

defining feature of this class of models is that the probability of receiving the treatment changes discontinuously as a function of an assignment variable being above or below a certain cut-off point. The underlying idea of a RDD is that, as in a randomised experiment, for individuals just above and below the pre-identified cut-off point, assignment to treatment is as good as random. Drawing from Angrist and Lavy (1999), in this case we exploit the fact that the probability of receiving *double-up* changes discontinuously depending on enrolment number in a specific module, which we denote by S_m , being above a certain cut-off point s_0 .

There are two types of RDD: the sharp and the fuzzy design (see Trochim, 1984). In the sharp design, treatment status depends deterministically on the running variable being above or below the cut-off point s_0 . In contrast, in the fuzzy design the probability of receiving the treatment is known to be discontinuous in the cut-off point s_0 , but it is not a deterministic function of S_m and selection on unobservables may therefore still be an issue. As noted earlier, the double-up policy might not have been implemented in some cases even when $S_m \geq 110$. This makes it essential that we implement the fuzzy RDD design.

Formally, let S_m be the (discrete) running variable of enrolment size on a module, m , and the cut-off point of interest is $s_0=110$, such that for students enrolled in modules with $S_m \geq s_0$, the probability of being exposed to the *double-up* jumps from zero to positive. Further, define an assignment to *double-up* rule $Z_m = 1(S_m \geq s_0)$, and let D_m be the *double-up* indicator, that identify students in modules where double-up took place. Then we can write:

$$P\{D_m = 1\} = \begin{cases} f_1(S_m), & S_m \geq s_0 \\ f_0(S_m), & S_m < s_0 \end{cases} \quad (2)$$

where, due to the discontinuity at the cut-off point, $f_1(S_m) \neq f_0(S_m)$.⁹ In the spirit of Hahn *et al.* (2001), a regression framework for a fuzzy RDD is offered by the instrumental variable approach so that:

$$Y_{i,m} = \alpha_i + g(S_m) + \beta D_m + X'_{i,m}\gamma + \varepsilon_{i,m} \quad (3)$$

where $g(\cdot)$ is a high-order polynomial and the *double-up* indicator, D_m , is instrumented by Z_m allowing the implementation of the fuzzy RDD parametrically using all observations in the sample in estimating equation (3).¹⁰

⁹ In this particular case $f_0(S_m)$ equals zero by definition.

¹⁰ Alternatively, one can implement the local linear estimator – a non-parametric estimator – which estimates equation (3) only for a sub-sample of the students in an arbitrarily small neighbourhood around the cut-off, s_0 ,

[Figure 1 about here]

Figure 1 presents a local polynomial smoothing (LPS) based plot of the probability of being exposed to *double-up* as a function of class-size. The figure shows a statistically significant discontinuity in the probability of receiving the *double-up* depending on class-size. It also reveals that to the left of s_0 the probability of treatment is a deterministic function of the running variable as in a sharp design. Battistin and Rettore (2008) define this type of design as a *partially* fuzzy RDD. Accordingly, the population can be divided into three groups: eligible participants (those on a module with $S_m \geq 110$ and received *double-up*), eligible non-participants (those on a module with $S_m \geq 110$ but without *doubled-up*) and non-eligible (those on a module with $S_m < 110$). Such a design is particularly advantageous for two reasons. *Firstly*, while in the case of a conventional fuzzy RDD we would need standard instrumental variable assumption, including exclusion restriction and monotonicity, in the particular case of the partially fuzzy RDD only the continuity assumption needs to be satisfied. *Secondly*, the partially fuzzy RDD allows estimating the average treatment effect on the treated (ATT), rather than the average treatment effect for compliers (LATE).¹¹ Notice, however, that the estimated parameters from this type of design are still defined as ‘local’ because they can only be informative of the effect of the double-up policy in the neighbourhood of the cut-off point.

[Figure 2 about here]

Figure 2 depicts results from a local polynomial smoothing (LPS) regression, which examines the effect of the double-up policy on the main outcome of interest (test score) graphically. It shows a significant jump in the outcome of interest at the pre-identified cut-off point of 110 students per class. Specifically, students who were enrolled in modules of sizes just above the threshold and, consequently, exposed to the double-up policy achieved significantly higher test scores than their counterparts without the policy. The findings in Figure 2 are thus consistent with the hypothesis that class-size does matter for academic achievements in postgraduate education.

and without the polynomial term, $g(\cdot)$. We chose the parametric approach to preserve sample size, but results are robust when performing the non-parametric approach (see next Section).

¹¹ I.e., as in experimental framework where all subjects who receive treatment are compliers (see Bloom, 1984). Here we do not have always takers (see extensive discussion on this in Angrist and Pischke (2008), pp. 161-166).

We next test the robustness of these findings in a regression framework, as specified in Equation (3), controlling for a number of other confounding factors. Table 3 reports results from this exercise, which represent first stage estimates of the effect of class size on the probability of being exposed to the double-up policy.

[Table 3 about here]

The results confirm the graphical intuition of Figure one and show that individuals in classes of size 110 and above, who were assigned to the double teaching, were significantly more likely to have actually been exposed to the policy than their counterparts who were not.

Table 4 reports RDD estimates of the effect of the double-up policy on grades, using the (partially fuzzy) RDD approach. Our preferred specification accounts for student fixed effects, thus controlling for predetermined unobservable characteristics such as academic ability and family background. Column (1) reports results from the basic specification with a second order polynomial term with no additional covariates. Column (2) includes the full set of covariates as explained above, while columns (3) and (4) include, respectively, third and fourth degree polynomial terms in addition to test the robustness of the results. Also, the RDD estimates in Table 4 and below are clustered by the running variable as suggested by Lee and Card (2008).¹²

The results show a significant and positive effect of the double-up policy on students' academic performances. Specifically, estimated coefficients in column (2) suggest that students exposed to the double-up policy achieved, on average, significantly higher modular grade than did their counterparts without double-up. These estimates are robust to the inclusion of various controls and different functional form choices.¹³

[Table 4 about here]

To check the robustness of our results, we estimated non-parametric RDD as specified in equation (3), but focusing only on the sub-sample of subjects within arbitrarily small

¹² Lee and Card (2008)'s clustering approach is the standard approach to date, which involves clustering using observations with similar/comparable values in the running variable around the cut-off as members of the same cluster. However, in their recent paper Kloeser and Rothe (2018) have recommended against this approach.

¹³ In Appendix Table A.2, we also report RDD estimates for a number of falsification tests, which show that no significant effect is found for different cut-off points

windows around the cut-off point. Results from this analysis are reported in Table 5, which broadly confirm that our results are robust.¹⁴

[Table 5 about here]

In table 6, we report heterogeneous effects of the double-up policy by disaggregating the sample into gender and broad nationality categories. The results reveal hardly any gender-specific difference in the effects of the double-up policy. On the other hand, the policy appears to have a differential effect linked to the broad nationality category. Specifically, the students who benefitted from the double-up policy are British students. This may suggest that international students with a relative inexperience of the higher education culture and their relative lack of fluency in the medium of instruction do not appear to gain from the double-up policy as their British counterparts do.

[Table 6 about here]

Finally, in Table 7 we investigate the potential heterogeneous effects of class-size, and report RDD estimates of the effects of the double-up policy separately by degree classifications. The results indicate that the policy significantly decreased the probability that students achieved the bottom degree classifications as can be gleaned from Column 4. Overall, the results presented provide compelling evidence that class size does matter, and that students who were exposed to the double-up policy achieved significantly higher grades. In the section that follows, we present a series of tests and checks to confirm the validity of these findings.

[Table 7 about here]

4. Validity

The main assumption, which needs to be satisfied for our identification strategy to produce unbiased estimates, is the continuity assumption. Borrowing from the jargon of the treatment effects literature, let $\{Y_i^1, Y_i^0\}$ be the potential outcome for individual i in case of treatment

¹⁴ Optimal bandwidth was selected using the procedure implemented by Calonico, Cattaneo and Titiunik (2014) and by Imbens and Kalyanaraman (2012).

and in the absence of treatment, respectively. Then, the continuity assumption that needs to be satisfied for the validity of a partially fuzzy RDD can be formally written as follows:

$$E[Y_i^0 | S_m = s_0^+] - E[Y_i^0 | S_m = s_0^-] = 0, \quad (4)$$

where s_0^+ and s_0^- represent, respectively, students just above and below the cut-off point, s_0 . In our case, this assumption entails that students enrolled in modules just above and just below the pre-identified cut-off level are identical in every respect, both in terms of observables and unobservables, but only differ in the probability of being exposed to the double-up policy. A direct way to assess the validity of this assumption is to examine if pre-intervention variables do not change discontinuously around the cut-off point. Figure 3 depicts the results from a set of local polynomial smoothing (LPS) regressions of variables such as age, gender, and country of origin that could not have been affected by the policy. Accordingly, we do not observe a statistically significant discontinuity at the cut-off point for any of these variables.

[Figure 3 about here]

As noted by McCrary (2008), however, the continuity assumption may be invalidated in cases where the treatment assignment rule is a public knowledge. Specifically, if people knew about the discontinuous nature of the assignment to treatment mechanism, those who expect to benefit from the double-up would manipulate the running variable, S_m , in order to receive the intervention, and selection bias would still be an issue. Given the nature of the double-up policy, it is quite unlikely that this would happen in our setting. However, in order to dispel any potential concerns about students sorting around the running variable, we also implemented the “donut hole” approach suggested by Barreca *et al.* (2016).¹⁵ The main idea behind this approach is that units closest to the cut-off are those most likely to have engaged in manipulation. Consequently, excluding such units from the analysis would eliminate any potential concern. In Table 8 we report RDD estimates of the double-up policy on test scores, excluding students who were enrolled in modules within 1 and 5 enrolment size points

¹⁵ Figure A.2 in the Appendix presents the outcome of the McCrary test, obtained through the STATA routine DCdensity.

around the cut-off. As the results reported in Table 8 confirm, our results remain robust to such validation tests.¹⁶

[Table 8 about here]

5. Conclusion

The paper examined the link between class-size and postgraduate grades using administrative data covering the population of candidates in one of the largest Schools of a Russell Group public university in the UK. As well as estimating fixed effects regressions we exploited a policy change aimed at reducing class-size to construct instrumental variables estimates of the impact of class-size on postgraduate grades using regression discontinuity design (RDD). We found that class-size impacts modular grades adversely confirming the well-established link between class-size and student performance. On the other hand, the policy designed to reduce class-size is found to have a significant positive impact on postgraduate modular grades. Importantly, we also found that the policy has reduced the probability that postgraduate students fail in their programme of study.

As noted earlier, supply side policies have led to a significant increase in tertiary level education overall. On the other hand, the changing funding environment in higher education institutions has made student fees a vital part of higher education funding, particularly at the PG level. This has renewed institutions' drive to recruit more students thus further reinforcing the effects of the supply side policies. In turn, this has revived some of the pre-existing concerns regarding student-to-staff ratio and the quality of tertiary education. Against this background, there has been a dearth of evidence on the effect of class-size on postgraduate grades, which this paper attempted to contribute to.

As noted earlier, it is apparent that independent learning is an integral part of the tertiary education landscape. However, the current funding climate within the higher education sector may make the question of class-size and its impact on student outcomes in that context all the more important. The recent Auger Review into post-18 education and funding has, for example, recommended reducing the cap on tuition fee. Faced with the prospect of reduction in the cap on tuition fee, higher education institutions may increase student intake to make up

¹⁶ Additionally, our findings are robust when implementing the treatment effect derivative (TED) of the estimated RDD, constructed by Dong and Lewbel (2015), which test for the stability of the RDD estimates (Cerulli *et al.*, 2017).

for funding shortfalls. If so, this is likely to have implications for class-size and student performance in tertiary education. The recent government undertakings (DFE 2018) to understand the role of contact hours and class-size highlights such concerns and underscores the importance of contact hours and class-size in determining student performance. Given these, the findings in this paper regarding double teaching are likely to be informative for policy makers, higher education institutions and students alike. As Huxley *et al.* (2017) noted, however, double teaching may not always be beneficial if doing so were to compromise teaching quality in some sense. Future work may usefully examine this issue in a wider context than has been done in this paper.

References

- Angrist, J. D., & Pischke, J.-S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.
- Angrist, J. D. & Lavy, V. (1999). Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement, *The Quarterly Journal of Economics*, 533-575.
- Bandiera, O., Larcinese, V., & Rasul, I. (2010). Heterogeneous Class Size Effects: New Evidence from a Panel of University Students. *Economic Journal*, 120(549), 1365–1398.
- Barr, A. and Turner, S. (2013). Expanding Enrollments and Contracting State Budgets: The Effect of the Great Recession on Higher Education, *The Annals of The American Academy*, 650, 168-193.
- Barreca, A. I., Lindo, J. M., & Waddell, G. R. (2016). Heaping-induced bias in regression-discontinuity designs. *Economic Inquiry*, 54(1), 268–293.
- Battistin, E., & Rettore, E. (2008). Ineligibles and eligible non-participants as a double comparison group in regression-discontinuity designs, *Journal of Econometrics*, 142(2), 715–730.
- Bloom, H. (1984). Accounting for No-Shows in Experimental Evaluation Designs. *Evaluation Review*, 8(2), 225–246.
- Browning, M., & Eskil Heinesen. (2007). Class Size, Teacher Hours and Educational Attainment. *The Scandinavian Journal of Economics*, 109(2), 415-438.
- Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust Nonparametric Confidence Intervals for Regression Discontinuity Designs. *Econometrica*, 82(6), 2295-2326.
- Thistlethwaite, D. L. and Campbell, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology*, 51(6), 309–316.

- Campbell, D. T. (1969). Reforms as Experiments, *American Psychologist*, XXIV, 409-429.
- Lee, D. & Card, D. (2008). Regression discontinuity inference with specification error. *Journal of Econometrics*, 142(2), 655–674.
- Case, A., & Deaton, A. (1999). School Inputs and Educational Outcomes in South Africa. *The Quarterly Journal of Economics*, 114(3), 1047-1084.
- Cerulli, G., Dong, Y., Lewbel, A., & Poulsen, A. (2017). Testing Stability of Regression Discontinuity Models. In *Advances in Econometrics, Volume 38* (pp. 317–339). Emerald Publishing Limited.
- Crawford, C., Crawford, R., & Jin, W. (2014). Estimating the public cost of student loans (No.R94). IFS Reports, Institute for Fiscal Studies.
- de Paola, M., Ponzio, M., & Scoppa, V. (2013). Class size effects on student achievement: heterogeneity across abilities and fields. *Education Economics*, 21(2), 135–153.
- de Paola, M., & Scoppa, V. (2011). The effects of class size on the achievement of college students. *Manchester School*, 79(6), 1061–1079.
- Department for Education (2018). Teaching Excellence and Student Outcomes Framework: Subject level (technical document to support the government consultation)
- Department for Education (2019). Independent panel report to the Review of Post-18 Education and Funding, May 2019.
- Dong, Y., & Lewbel, A. (2015). Identifying the Effect of Changing the Policy Threshold in Regression Discontinuity Models. *The Review of Economics and Statistics*, 97(5), 1081–1092.
- Gibbs, G., Lucas, L., & Simonite, V. (1996). Class size and student performance: 1984–94. *Studies in higher education*, 21(3), 261-273.
- Hahn, J., Todd, P. E., & van der Klaauw, W. (2001). Identification and Estimation of treatment effects using a regression-discontinuity design. *Econometrica*, 69(1), 201–209.
- Hanushek, E. A. (1979). Conceptual and empirical issues in the estimation of educational production functions. *Journal of Human Resources*, 14 (3), 351-388
- McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, 142(2), 698–714.
- Hoxby, C. (2000). The Effects of Class Size on Student Achievement: New Evidence from Population Variation. *The Quarterly Journal of Economics*, 115(4), 1239-128
- Huxley, G., Mayo, J., Peacey, M. W., & Richardson, M. (2018). Class size at university. *Fiscal Studies*, 39(2), 241-264.
- Imbens, G., and Kalyanaraman, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of economic studies*, 79(3), 933-959.
- Kokkelenberg, E., Dillon, M and Christy, S. (2008) The effect of class size on student grades at a public university, *Economics of Education Review*, 27, 221-233.

- Kolesar, M. & Rothe, C. (2018). Inference in Regression Discontinuity Designs with a Discrete Running Variable, *American Economic Review*, 108(8), 2277-2304.
- Krueger, A. (1999). Experimental Estimates of Education Production Functions. *The Quarterly Journal of Economics*, 114 (2), 497-532.
- Krueger, A., & Whitmore, D. (2001). The Effect of Attending a Small Class in the Early Grades on College-Test Taking and Middle School Test Results: Evidence from Project STAR. *The Economic Journal*, 111(468), 1-28.
- Leuven, E. , Oosterbeek, H. and Rønning, M. (2008), Quasi-experimental Estimates of the Effect of Class Size on Achievement in Norway. *Scandinavian Journal of Economics*, 110: 663-693.
- Monks, J. and Schmidt, R. (2011) The Impact of Class Size on Outcomes in Higher Education, *The B.E. Journal of Economic Analysis & Policy*, 11 (1), 1-17.
- Trochim, W. M. K. (1984). *Research Design for Program Evaluation: The Regression-Discontinuity Approach*. Beverly Hills: Sage Publications.

Table 1: SUMMARY STATISTICS

Variable	Mean	Standard Deviation	Min.	Max.	N
Personal Attributes:					
Female	0.697	0.46	0	1	987
Age	23.32	2.554	20	42	987
Country of Origin Attributes:					
British	0.083	0.276	0	1	987
Chinese	0.605	0.489	0	1	987
Other Countries	0.312	0.464	0	1	987
Academic Performances:					
Overall Mark	61.035	11.03	0	94	7696
Distinction	0.201	0.401	0	1	7696
Merit	0.425	0.494	0	1	7696
Pass	0.341	0.474	0	1	7696
Fail	0.034	0.18	0	1	7696
Class Size:	141.309	102.0192	1	396	7696

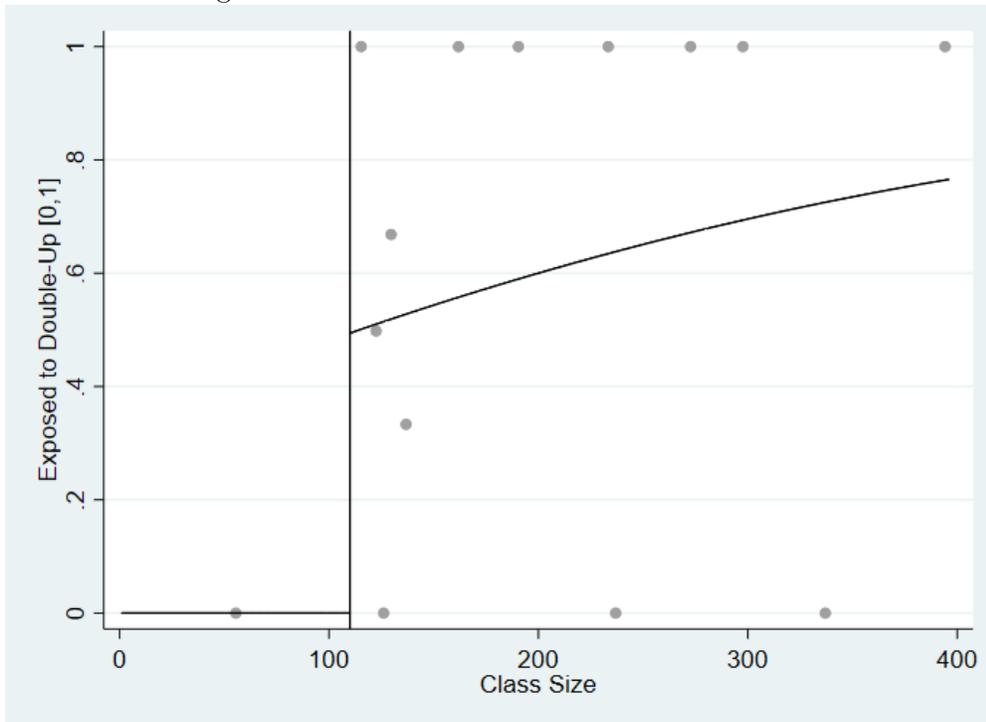
Note: The table reports summary statistics of the sample of interest.

Table 2: BASELINE ESTIMATES

	Pooled Model		Fixed Effects Model	
	(1) Overall Mark	(2) Overall Mark	(3) Overall Mark	(4) Overall Mark
Class Size	-0.0330*** (0.0071)	-0.0262*** (0.0079)	-0.0293*** (0.0082)	-0.0197** (0.0086)
Class Size ²	0.0010*** (0.0002)	0.0009*** (0.0002)	0.0010*** (0.0002)	0.0010*** (0.0002)
Covariates:				
Core Module [0,1]		-0.0461 (0.0325)		-0.1644*** (0.0323)
Spring Term Module [0,1]		0.0717** (0.0341)		0.0749*** (0.0277)
10 Credits		-0.6232*** (0.1367)		-0.6059*** (0.2187)
15 Credits		-0.5055*** (0.0984)		-0.4365*** (0.1060)
20 Credits		-0.4075** (0.1687)		-0.9239*** (0.1941)
Female [0,1]		0.1157*** (0.0360)		- (-)
Years of Age		0.0073 (0.0059)		- (-)
British [0,1]		0.1811*** (0.0632)		- (-)
Chinese [0,1]		-0.1117*** (0.0376)		- (-)
Constant	0.2014*** (0.0432)	0.4403** (0.1910)	0.1569*** (0.0518)	0.5876*** (0.1287)
Observations	5383	5383	5382	5382

Note: The table presents baseline estimates of the effect of class size on student overall mark. The difference in sample size is due to the fact that for this exercise we exclude students who were exposed to the double-up teaching. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure 1: GRAPHICAL ANALYSIS - FIRST STAGE



Note: The figure shows local polynomial estimates of the probability of being exposed to the double-up teaching as a function of module enrolment size.

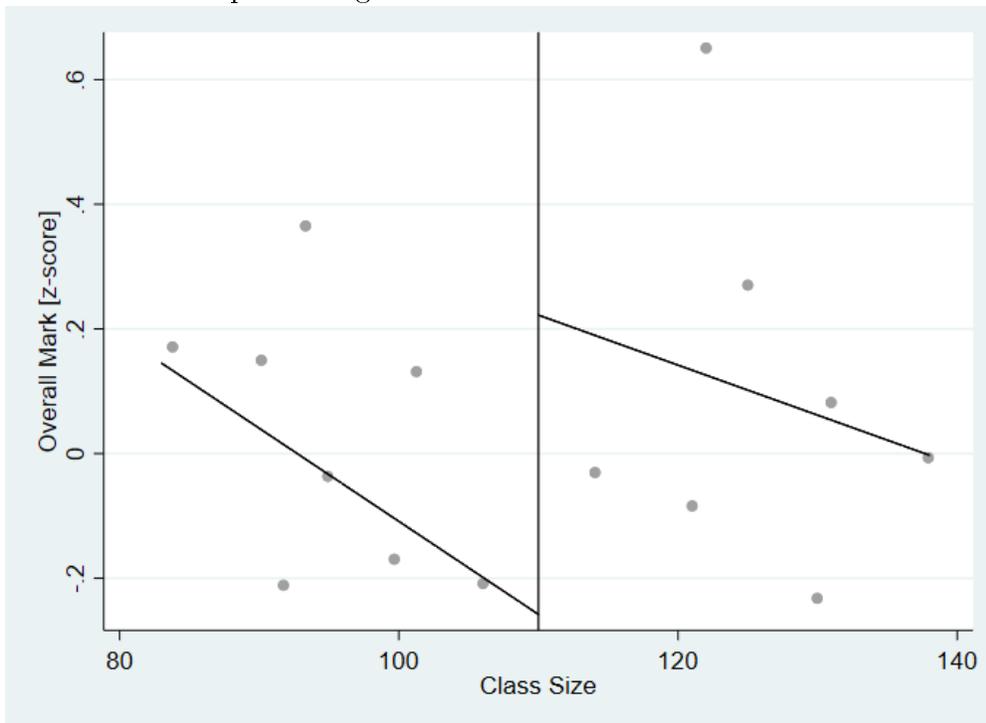


Figure 2: GRAPHICAL ANALYSIS - LOCAL POLYNOMIAL ESTIMATES

Note: The figure shows local polynomial estimates of the effect of the double-up on students' overall mark.

Table 3: FIRST STAGE ESTIMATES

	Pooled Estiamtes		Student Fixed Effects	
	(1) Actual Double-up	(2) Actual Double-up	(3) Actual Double-up	(4) Actual Double-up
Assigned Double-up [0,1]	0.4944*** (0.0150)	0.4944*** (0.0150)	0.7163*** (0.0182)	0.7110*** (0.0190)
Running Variable				
Class Size	0.0045*** (0.0017)	0.0045*** (0.0017)	-0.0111*** (0.0023)	-0.0106*** (0.0024)
Class Size ²	0.0001*** (0.0000)	0.0001*** (0.0000)	0.0004*** (0.0000)	0.0004*** (0.0000)
Covariates:				
Core Module [0,1]	-0.0148* (0.0088)	-0.0148* (0.0088)		-0.0201* (0.0105)
Spring Term Module [0,1]	0.0142* (0.0086)	0.0142* (0.0086)		0.0043 (0.0089)
10 Credits	0.0263* (0.0154)	0.0263* (0.0154)		-0.0749*** (0.0166)
15 Credits	0.0070 (0.0079)	0.0070 (0.0079)		-0.0309** (0.0123)
20 Credits	0.0246** (0.0116)	0.0246** (0.0116)		-0.0742*** (0.0187)
Female [0,1]	0.0206** (0.0090)	0.0206** (0.0090)		- (-)
Years of Age	0.0005 (0.0013)	0.0005 (0.0013)		- (-)
British [0,1]	0.0148 (0.0130)	0.0148 (0.0130)		- (-)
Chinese [0,1]	-0.0286*** (0.0090)	-0.0286*** (0.0090)		- (.)
Constant	-0.0534 (0.0374)	-0.0534 (0.0374)	-0.0040 (0.0140)	0.0352 (0.0257)
Observations	7696	7696	7696	7696

Note: The table presents first stage estimates of the effect of class size on the probability of being exposed to the double-up teaching. Robust standard errors in parentheses are clustered by unique values of the running variable. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: RDD ESTIMATES WITH STUDENT FIXED EFFECTS

	(1)	(2)	(3)	(4)
	Overall Mark	Overall Mark	Overall Mark	Overall Mark
Double-up [0,1]	0.3314** (0.1289)	0.3237** (0.1284)	0.3250** (0.1301)	0.9117*** (0.2519)
Running variable:				
Class Size	-0.0376*** (0.0095)	-0.0395*** (0.0101)	-0.0197 (0.0230)	0.3665*** (0.1120)
Class Size ²	0.0006*** (0.0002)	0.0007*** (0.0002)	-0.0004 (0.0014)	-0.0457*** (0.0140)
Class Size ³			0.0000 (0.0000)	0.0018*** (0.0006)
Class Size ⁴				-0.0000*** (0.0000)
Covariates:	X	✓	✓	✓
Observations	7696	7696	7696	7696

Note: The table presents RDD estimates of the impact of the double-up policy on student overall mark. Results are conditional on a vector of covariates as shown in table 3. Robust standard errors in parentheses are clustered by unique values of the running variable. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: RDD ESTIMATES, AT AN ARBITRARILY SMALL WINDOW AROUND THE CUT-OFF

	Non-Parametric RDD		
	(1)	(2)	(3)
	Overall Mark Window [80, 140]	Overall Mark Window [50, 170]	Overall Mark Window [20, 200]
Double-up [0,1]	0.4344* (0.2276)	0.2733* (0.1504)	0.3947*** (0.1399)
Running variable:			
Class Size	-0.0066 (0.0405)	-0.0076 (0.0226)	-0.0210** (0.0084)
Covariates:	✓	✓	✓
Observations	3242	4712	5721

Note: The table presents RDD estimates of the impact of the double-up policy on student overall mark. Results are conditional on a vector of covariates as shown in table 3. Robust standard errors in parentheses are clustered by unique values of the running variable. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: RDD ESTIMATES, DISAGGREGATED BY GENDER AND COUNTRY OF ORIGIN

	(1) Overall Mark Boys	(2) Overall Mark Girls	(3) Overall Mark British	(4) Overall Mark Chinese
Double-up [0,1]	0.3155** (0.1487)	0.3215** (0.1535)	0.4737** (0.1890)	0.2295 (0.1534)
Running variable:				
Class Size	-0.0414*** (0.0148)	-0.0380*** (0.0119)	-0.0166 (0.0261)	-0.0284** (0.0121)
Class Size ²	0.0005* (0.0003)	0.0008*** (0.0002)	-0.0005 (0.0007)	0.0006*** (0.0002)
Covariates:	✓	✓	✓	✓
Observations	2292	5404	618	4743

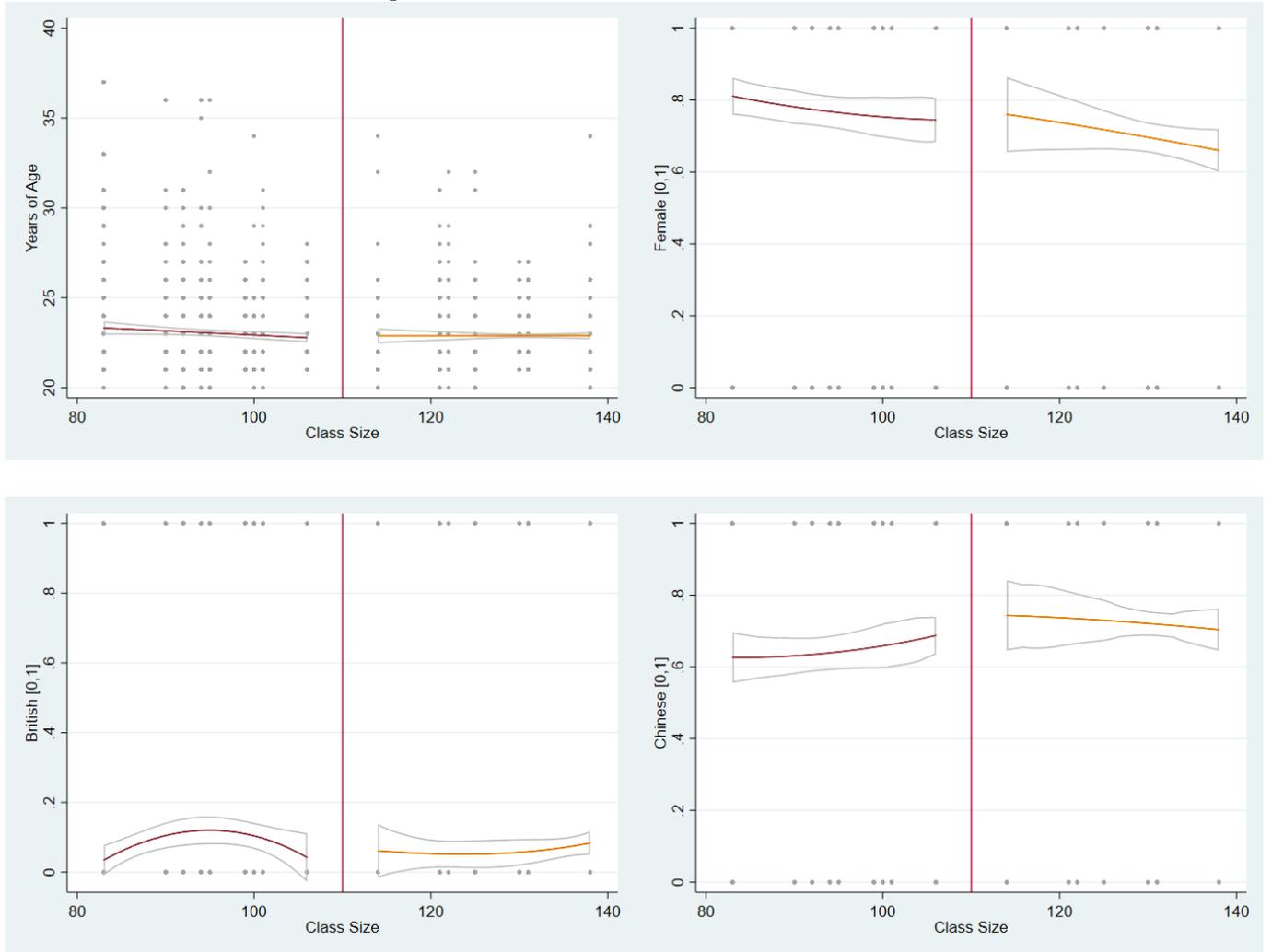
Note: The table presents RDD estimates of the impact of the double-up policy on student overall mark. Results are conditional on a vector of covariates as shown in table 3. Robust standard errors in parentheses are clustered by unique values of the running variable. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: RDD ESTIMATES, DISAGGREGATED BY DEGREE CLASSIFICATION

	(1) Distinction	(2) Merit	(3) Pass	(4) Fail
Double-up [0,1]	0.0146 (0.0516)	0.0504 (0.0497)	0.0168 (0.0444)	-0.0818*** (0.0298)
Running variable:				
Class Size	0.0004 (0.0046)	-0.0212*** (0.0048)	0.0114** (0.0048)	0.0095*** (0.0036)
Class Size ²	0.0000 (0.0001)	0.0004*** (0.0001)	-0.0003*** (0.0001)	-0.0002** (0.0001)
Covariates:	✓	✓	✓	✓
Observations	7696	7696	7696	7696

Note: The table presents RDD estimates of the impact of the double-up policy on student overall mark. Results are conditional on a vector of covariates as shown in table 3. Robust standard errors in parentheses are clustered by unique values of the running variable. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure 3: RDD CONTINUITY TEST.



Note: The figure shows local polynomial estimates of pre-treatment variables as a function of enrolment size.

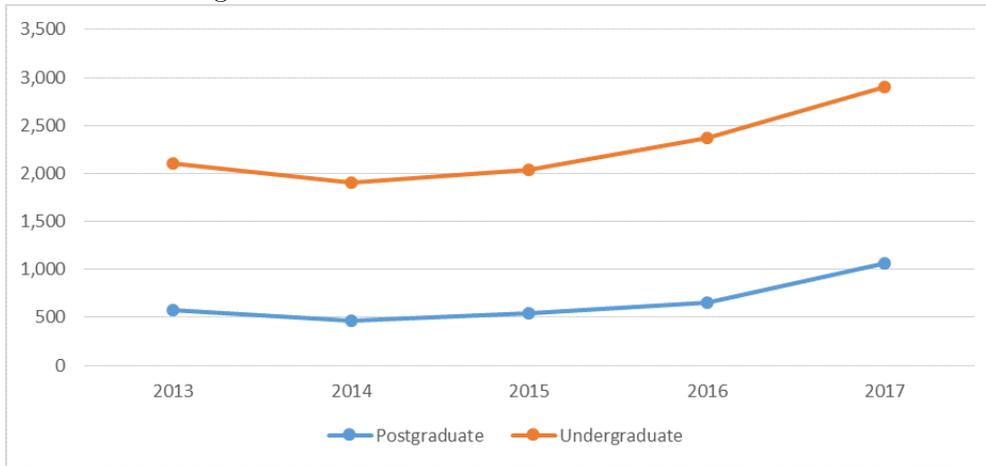
Table 8: RDD ESTIMATES, DONUT-HOLE APPROACH

	Outcome: Overall Mark	
	(1) Excluded [+1, -1] Around the Cut-off	(2) Excluded [+5, -5] Around the Cut-off
Double-up [0,1]	0.2851** (0.1340)	0.3237** (0.1284)
Running variable:		
Class Size	-0.0367*** (0.0105)	-0.0395*** (0.0101)
Class Size ²	0.0006*** (0.0002)	0.0007*** (0.0002)
Covariates:	✓	✓
Observations	7270	7696

Note: The table presents RDD estimates of the impact of the double-up policy on student overall mark. Results are conditional on a vector of covariates as shown in table 3. Robust standard errors in parentheses are clustered by unique values of the running variable. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

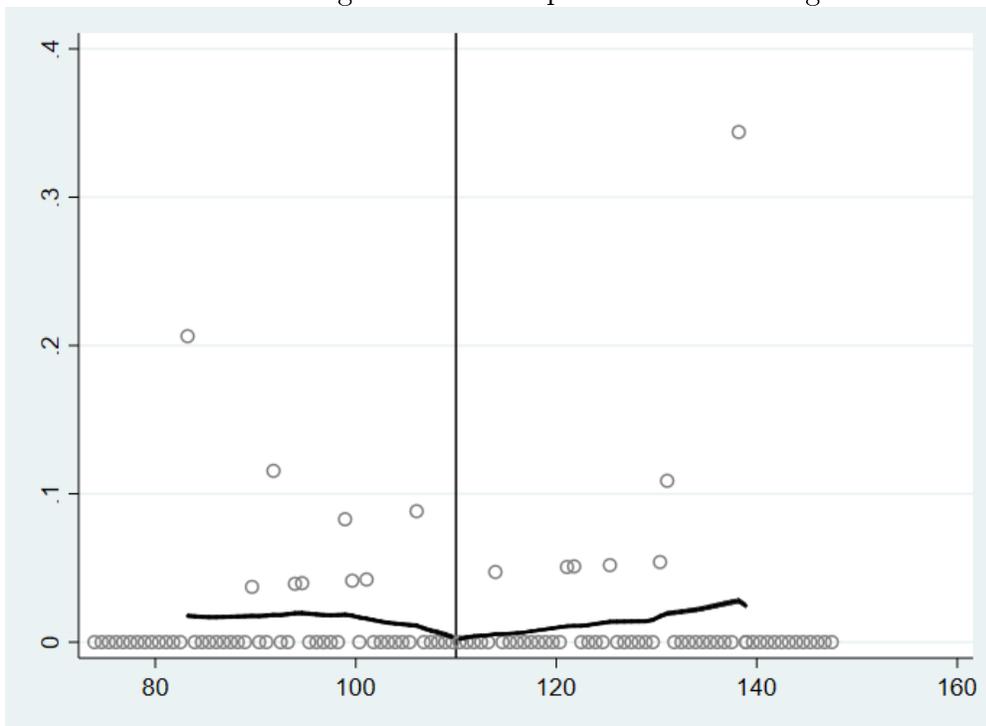
Appendix

Figure A.1: STUDENT REGISTRATION OVER THE LAST FIVE YEARS.



Note: The figure shows number of student registered at the University over the last five years.

Figure A.2: Manipulation of Running Variable



Note: The figure reports evidence of no manipulation around the cut-off.

Table A.1: RDD ESTIMATES, FALSIFICATION TEST

	(1) Overall Mark Cutoff=20	(2) Overall Mark Cutoff=135	(3) Overall Mark Cutoff=160
Double-up [0,1]	-0.0291 (0.3583)	-1.5868 (2.7737)	-0.9451 (1.1601)
Running variable:			
Class Size	-0.0172 (0.0226)	0.0811 (0.1673)	0.0406 (0.0668)
Class Size ²	0.0004 (0.0003)	-0.0009 (0.0023)	-0.0004 (0.0009)
Covariates:	✓	✓	✓
Observations	7696	7696	7696

Note: The table presents RDD estimates of the impact of the double-up policy on student overall mark. Results are conditional on a vector of covariates as shown in table 3. Robust standard errors in parentheses are clustered by unique values of the running variable. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.