University of New South Wales

School of Economics

Honours Thesis

The Impact of Selective High Schools on Student Achievement

Evidence from New South Wales, Australia

Author:
Kai Zen
Student ID: 3414707

Supervisor:
Associate Professor Gigi Foster

Bachelor of Economics (Honours in Economics)
and
Bachelor of Economics (Economics, Econometrics)
Bachelor of Commerce (Finance)

15th November, 2016

Declaration

I hereby declare that the content of this thesis is my own work and that, to the best of my knowledge, it contains no material that has been published or written by another person or persons, except where due acknowledgement has been made. This thesis has not been submitted for award of any other degree or diploma at the University of New South Wales or any other educational institution.

Kai Zen
15th November, 2016

Disclaimer: The findings and views expressed here are those of the author and should not be attributed to the NSW Department of Education, or units thereof. All remaining errors are my own.

Acknowledgements

I would first like to express my deep gratitude to my supervisor, Gigi Foster, for all the guidance and support she has given me over the year. I would like to give particular mention to Gigi's pragmatic and insightful advice concerning my research aims and ideas, as well as all her tireless work in helping me procure the necessary data. Her patience, intellect and meticulous attention to detail have helped to bring much-needed clarity to this thesis.

I would also like to thank the Economics faculty at UNSW for their central role in my education not only this year, but throughout my time at university. I thank Chris Gibbs for encouraging me to undertake Honours in the first place, and for taking me on as a Research Assistant last year, which helped further develop the research and programming skills that have since proved invaluable. I thank Geni Dechter for all her behind-the-scenes work as Honours co-ordinator, as well as for acting as my discussant at the 2016 Student Research Fair and providing valuable suggestions for my thesis. I also thank Tess Stafford for guiding our thesis seminars and enduring our presentations, as well as everyone who contributed comments at the practice, faculty, and Student Research Fair presentations. In addition, I wish to take this opportunity to thank the benefactors of the Michael E. Grace Memorial Honours Scholarship for their financial support this year. I am grateful to the NSW Department of Education for allowing me access to the data set analysed in this thesis.

This year would have been all the more difficult without my fellow classmates in the 2016 UNSW Economics Honours cohort and the memorable times spent together in the Honours room, absurd and unproductive as they often were. Anecdotally, I can attest to strong peer effects on academic achievement — though at times it was difficult to identify the sign of such effects. I would like to give particular mention to the fellow members of the Macro Dream Team, Calvin He and Sébastien Printant, Tran Nguyen, who contributed helpful comments on my draft, as well as Cecilia Chang, Kelvin Guo, Damoon Sadeghian, and Nathan Walsh.

Finally, I would like to thank my friends and family, who have always been there for me and who have always believed in me. In particular, I want to express my endless thanks to my Mum and Dad for all their unbounded love, care, and reminders to eat fruit, as well as to my Grandma for her wisdom and for always supporting me in my pursuits.

Contents

List of Figures

List of Tables

Abstract

Selective high schools in the Australian state of New South Wales (NSW) provide an opportunity for students to attend a public school with significantly higher-achieving peers — the average successful applicant scores more than two standard deviations higher on baseline numeracy tests than the state average. Competition for entrance into these schools is fierce, with general public opinion attributing the superlative academic success of selective school students at least in part to the selective school environment. Much recent attention has been paid to credible evaluations of similar selective programs in other jurisdictions. Studies by Abdulkadiroğlu, Angrist, and Pathak (2014) and Dobbie and Fryer (2014) in Boston, MA and New York City, NY find little-to-no significant effect of attending selective high schools on student achievement. In this paper, I employ fuzzy regression discontinuity designs on 18 NSW selective schools with varying gradations of selectivity to estimate causal effects of selective school attendance on performance in high-stakes university entrance assessments and participation rates in advanced coursework. This is the first such study of selective schools in NSW, which is home to the oldest and most extensive selective school system in Australia, using a newly matched dataset encompassing the school careers of three state-wide cohorts of selective school applicants. I find that receiving an offer to attend a selective school has only scattered and mostly insignificant impacts on overall student achievement and participation in advanced coursework. I do find suggestive evidence that selective schools benefit low socioeconomic status students, but that such students are typically underrepresented in selective schools, which has implications for Gifted and Talented education policy.

Chapter 1
Introduction

Students at academically selective schools tend to perform superlatively on academic measures — 8 of the 10 highest-performing[1] high schools in New South Wales (NSW), Australia are selective high schools according to an unofficial ranking by the *Sydney Morning Herald* (Ting and Bagshaw, 2015). The success of selective school students is often attributed to the selective school environment, including heightened teacher expectations, more rigorous instruction, and high-performing peers, although selective schools have also been criticised in the media for being high-pressured and academically strict (Broinowski, 2015). As a result of their reputation for academic success, competition for NSW selective schools tends to be fierce: in 2016, there were 13,118 total applicants and 4,215 available places (NSW Department of Education, 2016). Over the years 2007-2009, 5,882 students were in contention for a place at the selective school with the most first preferences and 621 were ultimately successful, a success rate of 10.6%.

Student 'tracking' by academic ability, such as through academically selective schools or classes, is a common educational policy in many jurisdictions, but its effects on students have only recently begun to be formally investigated. Straightforward comparisons of academic success are unable to differentiate the true treatment effect of attending a selective school from the selection bias caused by underlying differences between those who attend selective schools and those who do not, for example, in motivation or work ethic. I employ regression discontinuity designs to credibly evaluate the causal effects of offers to attend selective schools on academic achievement using a newly-matched dataset on the academic careers of all students applying to selective schools in NSW over the years 2007-2009. The intuition behind this strategy is that the selective school admission process in NSW relies on implicit cut-offs in a single metric of academic achievement (a 'selective score') to determine if a prospective student gains entrance to a given school or not. These discontinuities in treatment assignment (hence the use of regression discontinuity designs) form multiple quasi-experiments: students who score just above a selective school's cut-off receive the opportunity to attend that selective school (the 'treatment' group), while students just below do not (the 'control' group). However, because test scores are noisy, a student's precise position in the area close to the cut-off is essentially random, and so students on either side of the cut-off are otherwise comparable — such as in underlying ability, baseline motivation, and so on. By focusing on students close to the cut-off, I am then able to recover valid treatment and control groups with which I can isolate the causal effect of offers to attend a selective school on academic outcomes.

Policy evaluations of ability tracking in education from around the world find mixed results. Results appear to depend on the outcome measure considered, and comparisons across studies are made more difficult by policy and cultural heterogeneity. In addition, there are competing theoretical mechanisms behind the effects of tracking students, and some mechanisms may dominate others in certain contexts. For instance, selective school peer groups theoretically produce beneficial information spillovers, reduced negative externalities from a potentially less poorly-behaved peer group, and social network formation, which ultimately enhance human and

[1] As determined by the share of Higher School Certificate exams taken which resulted in a score above 90 out of 100. The other 2 schools are non-government schools.

social capital development (Dobbie and Fryer, 2011). Negative effects of selective schools on students might also exist, such as those caused by potentially stressful or competitive environments stemming from a high-achieving peer body. This could be particularly true for the selective school student who just meets the admission criteria, who might be negatively impacted by teaching targeted at a relatively high level, or a fall in their academic standing relative to their high-achieving peers.

As part of the public school system, selective schools also expand school choice for high-achieving students from low socioeconomic status backgrounds, and therefore potentially play a role in improved intergenerational mobility and equality of opportunity. However, some critics allege high levels of social exclusivity at selective schools (Smith, 2014), perhaps as a result of potential SES biases in the admissions process or inequality in the primary school system, which might actually perpetuate inequality.

The large number of selective schools in NSW allows me to examine selective school effects across a wide spectrum of relative selectivity, and I extend my analysis by estimating selective school effects for student subpopulations, as well as estimating the effects of average peer achievement more specifically. A study of the NSW selective schools program is particularly informative, as it is the most extensive (involving 21 fully selective and 24 partially selective schools) and oldest program of its kind in Australia. By comparison, there are four selective schools in Victoria, the second-most populous state in Australia; one in Western Australia; and three recently established, senior-level 'Queensland Academies' in Queensland. In addition, whereas the majority of NSW selective schools were established in the late 1980s, many of these other Australian selective schools were only established in the late 2000s.

Overall, I find limited and largely insignificant evidence for overall selective school effects on overall and individual subject-area achievement, as well as on participation in advanced coursework. These null results hold for both comparisons between students who receive offers to attend a selective school and those who do not and attend a non-selective government high school, as well as comparisons between students who receive offers to more- or less-selective schools. There is additionally little evidence of selective school effects that differ when estimated separately on low- and high-achieving students, suggesting that selective schools affect students across the ability range in a similar manner. I find suggestive evidence that male students generally benefit from selective schools, whereas female students only benefit at the exceptionally selective end. I do find some evidence that socioeconomically disadvantaged students benefit from selective schools, indicating that the primary benefits to selective schools accrue to those whose school choice is likely to be most limited. However, I find strong links between the average socioeconomic status of a student's primary school and both application rates as well as selective scores achieved. The proximal policy recommendations of this paper are then to implement measures to improve application rates amongst low socioeconomic status students, and to improve the identification and support of gifted and talented students in disadvantaged primary schools.

Chapter 2
Background literature

There is a sizable literature on the broad theme of student tracking, including selective classes or schools, and in the realm of Gifted and Talented education. Tracking policy is often justified by the purported benefits to grouping similar students in itself, and so much of this literature also coincides with the literature on peer effects in education. Evaluations of tracking programs and peer effects have been conducted in a diverse range of settings and jurisdictions, a summary of which is displayed in Table 2.1. However, as can be seen, a consensus on the effects of tracking has yet to be reached, perhaps as a result of the heterogeneity of institutional environments. This study contributes to this literature by examining the most extensive selective school system in Australia, where ability tracking policies, and educational institutions more broadly, likely most resembles those found in the U.K. or the U.S. External validity likely decreases for other jurisdictions as contextual differences become more pronounced, such as developing or non-English speaking countries.

Table 2.1: Summary of selective and similar ability-tracking program evaluations

Authors	Jurisdiction	Significant effects found for:		
		Test scores	Subject participation	University variables
Abdulkadiroğlu et al. (2014)	Boston, MA	No		No
Abdulkadiroğlu et al. (2014)	New York City, NY	No		
Atkinson, Gregg, and McConnell (2006)	UK	No		
Clark (2010)	UK	No	Yes	Yes
Clark and Del Bono (2016)	Aberdeen, Scotland			Yes
Cohodes (2015)	Boston, MA	No	Yes	Yes
Ding and Lehrer (2007)	Jiangsu, China	Yes		
Dobbie and Fryer (2014)	New York City, NY			No
Dobbie and Fryer (2011)	New York City, NY	No	Yes	
Duflo, Dupas, and Kremer (2011)	Kenya	Yes		
Dustan, de Janvry, and Sadoulet (2015)	Mexico City, Mexico	Yes		
Houng (forthcoming)	Unspecified Australian state	Yes		
Jackson (2010)	Trinidad & Tobago	Yes		
Pop-Eleches and Urquiola (2013)	Romania	Yes		
Zhang (2008)	Wuhan, China	No		

Studies in bold type are evaluations of selective secondary schools that bear the most resemblance to those found in NSW.

2.1 Peer effects: theory and evidence

The influential Coleman et al. (1966) report brought peer composition to the fore as a major input to student achievement by highlighting how peer groups experienced by disadvantaged students lead to poor levels of achievement. The justifications made by education policymakers for implementing tracking programs typically rest on the purported benefits from peer group homogeneity, which are thought to benefit students *per se*, as well as allowing for matching learning environments to student characteristics, for example through advanced curricula and smaller class sizes. Peer ability is explicitly stated as a significant channel through which NSW selective schools benefit their students: it is claimed that '[selective] schools can provide intellectual stimulation by grouping together gifted and talented students who may otherwise be isolated from a suitable peer group' (NSW Department of Education, 2015a). The central role of peer groups in tracking contexts such as selective school programs merits a review of peer effects, or more formally, the impact of the behaviour and characteristics of an individual's peers (potentially encompassing classmates, neighbours, friends, colleagues, etc.) on that individual.

The theoretical literature[1] provides competing, and at times mutually exclusive models of peer effects in education. Dobbie and Fryer (2014) outline some potential theoretical benefits to a high-achieving peer group: information spillovers, networking, and reduced negative externalities from a potentially less poorly-behaved peer group (the 'bad apple' model of peer effects, see also Lazear (2001)). Long-term benefits from less-disruptive peers are potentially quite large: Carrell, Hoekstra, and Kuka (2016) estimate that removing one disruptive peer from a primary-school classroom for one year would raise the present discounted value of the future earnings of that class by $100,000. Hoxby and Weingarth (2005) propose and test for a large variety of educational peer effect models, illustrating the potential ambiguity of theory in this area. The 'focus' model of peer effects suggests that peer homogeneity might be beneficial for all students (even those dissimilar to the peer group) because, say, education outcomes actually suffer when teachers attempt to target students with varying needs. The 'boutique' model suggests that students only benefit when they are surrounded by students with similar characteristics, because this allows teaching methods and materials to be more appropriately tailored. High-achieving students specifically may also act as role models, and inspire their peers to achieve (the 'shining light' model). On the other hand, high-achieving peer effects may be negative: marginal students may suffer from 'invidious comparison' to their higher-achieving peers, which may reduce self-esteem and confidence. It is also possible that peer diversity (across achievement or other dimensions) is beneficial to students (the 'rainbow' model), for example because a variety of viewpoints and backgrounds might foster better academic understanding. Hoxby and Weingarth (2005) exploit exogeneous student reassignment in a North Carolina county to test these various models, and find evidence supporting the boutique and focus models, implying that students benefit from student body homogeneity of the type found in selective school environments.

More recently, Tincani (2015) develops an alternate model of peer effects which depend on the degree of concern students have for their relative academic rank, and applies the model to an empirical setting with exogenous peer shocks. In this model, students with peers that are relatively close in ability have more incentive to compete, as there are large potential gains in rank; when students face relatively higher-achieving peers, the cost of exerting effort for the same potential gains in rank is higher. This model thus predicts that the likely higher salience of rank concerns in tracked settings such as selective schools leads to stronger competition and positive peer effects for middle-ability students, but negative peer effects for the low-ability students.

[1] For a survey of educational peer effects, including theorised mechanisms thereof, see Epple and Romano (2011).

The results found in the empirical literature are similarly ambiguous, with wide-ranging effects that appear to be sensitive to the methodology and institutional environment. Angrist (2014) provides a critical survey of the empirical peer effects literature that outlines common methodological issues, and observes that credible studies in which peer characteristics are determined exogenously to individual characteristics tend not to find significant causal peer effects. Sacerdote (2014) provides a separate survey of experimental and quasi-experimental analyses of peer effects, concluding that peer effects are highly context-specific, and seem to manifest more readily in outcomes such as crime, drinking behaviour, and career choices rather than test scores. Approximately half of the analyses examined find no significant peer effects. Importantly, robust educational peer effects may become evident when the peer effect is allowed to vary by the *individual's* achievement levels as well as the achievement level of that individual's peer groups. On the other hand, Hanushek, Kain, Markman, and Rivkin (2003) find highly significant and positive primary-school peer effects on learning consistent for students across the whole test score distribution. Moreover, no significant effects arising from the *variance* of peer achievement are found, which may contradict the theoretical model proposed by Tincani (2015), although rank concerns may be relatively muted in primary schools as compared to other contexts.

While many studies of peer effects in education focus on the academic quality of peers, peer effects can also potentially arise from other student characteristics, such as race and socioeconomic status. Channels by which race or socioeconomic peer effects operate might come in the form of teacher expectations or biases, or conformity to social group norms (Sacerdote, 2011). For instance, Austen-Smith and Fryer (2005) construct a model of 'acting white', referring to when black students perform below their potential in order to conform to peer group expectations. Hoxby (2000) find empirical results of substantial negative impacts of peer groups with higher shares of black or Hispanic students . More recent studies, however, find only small or insignificant racial peer effects. Abdulkadiroğlu et al. (2014) find no evidence of a substantial impact from large drops in the minority and low-income proportion of peers on student achievement at academically selective schools in Boston. Angrist and Lang (2004) find generally no impact on academic achievement from a desegregation program in Boston that resulted in a positive exogenous shock in the fraction of disadvantaged and minority students in peer groups. Similarly, Hoxby and Weingarth (2005) find evidence that peers' race, ethnicity and income have only slight effects on individual achievement after accounting for peer achievement.

When examining different school environments more generally, it can be convenient to think of peer effects encapsulating other school characteristics seemingly unrelated to peers, such as differences in funding levels, course menu and curriculum, teaching quality, and so on, if these differences are themselves driven by differences in peers (Abdulkadiroğlu et al., 2014; Sacerdote, 2011). For example, higher-achieving students are likely to demand more advanced mathematics courses, so if part of the effect of attending that school can be attributed to more intensive mathematics instruction, we could reframe this in terms of peer effects. Similarly, teachers are likely to match instruction and academic standards to their students, implying a higher academic standard or rigour in selective high schools. In this way, some of the effect of teacher behaviour can be at least partially reframed as a peer effect.

Most of the studies in the education literature mentioned thus far estimate or model peer effects on academic achievement (often measured in terms of standardised test scores), but many other important aspects of student welfare that are potentially affected by peer effects remain unexplored because of data availability, measurement, and identification concerns. For example, the NSW selective schools policy wording above states that ability tracking can provide 'intellectual stimulation', which may come in any number of forms that are not necessarily linked to individual assessment achievement. Student satisfaction, openness to experience, or mental health, to name some examples, may be improved for students at selective schools. Similarly, few studies exist which examine peer effects on long-term outcomes, such as higher-education, employment, or income outcomes. Examples include Carrell et al. (2016), which tracks long-term peer effects on later incomes as mentioned earlier, and a novel study by Shue (2013) which identifies strong evidence for peer effects in randomly assigned Harvard Business School peer groups with respect to later executive compensation and corporate strategic behaviour.

2.2 Common methods of studying the effects of academic tracking policies

The often ethically or administratively prohibitive nature of randomised controlled trials poses an identification issue for the evaluation of student tracking effects. The randomised controlled trial conducted by Duflo et al. (2011) is one exception, in which within-school tracking institutions were randomly assigned to Kenyan primary schools. The authors find benefits for the schools which employ tracking as a result specifically of the impact of peers on influencing teacher behaviour, that is, by inducing effort and teaching that was better-targeted to the average student level. More traditional studies of tracking involved largely straightforward comparisons of the achievement of various student groups, with control variables such as initial ability and student demographics (see Betts and Shkolnik (2000) for a survey of such studies). Yet such studies suffer from selection bias by omitting unobserved variables likely to impact both group placement and achievement measures, such as motivation and effort, which likely upwardly biases treatment effect estimates. Similarly, studies that attempt to estimate peer effects can be prone to pitfalls arising from a lack of credible exogenous peer group determination (Angrist, 2014).

More modern studies, such as Abdulkadiroğlu et al. (2014) and Cohodes (2015), have relied largely on quasi-experimental econometric methods, most commonly regression discontinuity designs (first introduced by Thistlethwaite and Campbell (1960)) which 'naturally' create balanced treatment and control groups. This is possible because assignment to tracking regimes often relies on a student being above or below some threshold in a noisy measure of ability, such as test scores (the 'forcing' or 'running' variable parameter in regression discontinuity terminology). The precise score attained by students in some close bandwidth around the threshold can be thought of as a random draw, and so for these students treatment assignment is as good as randomly assigned. In cases where regression discontinuity is unsuitable, for example due to unavailability of data on the forcing variable, some studies rely on matching estimators (see, for example, Atkinson et al. (2006) and Houng (forthcoming)). Using these methods, studies find generally mixed-to-null evidence for benefits to tracking on academic achievement.

2.3 Evaluations of comparable selective school systems around the world

Studies examining tracking or other education policies similar to NSW selective schools include those which report significant and positive (Ding and Lehrer, 2007; Jackson, 2010; Pop-Eleches and Urquiola, 2013), mixed (Clark, 2010) as well as little-to-no effects (Abdulkadiroğlu et al., 2014; Dobbie and Fryer, 2014). The presence or absence of selective school effects may depend on the outcome variable of interest, students' socioeconomic status and race, gender, as well as specific institutions which are likely to vary widely across jurisdictions.

Houng (forthcoming) provides an exploratory analysis of selective schools in an unspecified Australian state, finding significant gains for both English-speaking and non-English speaking background girls, but not for boys. One key difference between this and other policy evaluations (including this paper) is the availability of data on non-government students, thus removing a key source of bias. However, because of a lack of data on the variable determining treatment (the 'forcing variable'), propensity score matching is employed instead of more commonly used regression discontinuity designs.

In the U.S., Abdulkadiroğlu et al. (2014) evaluate the causal effects of peer characteristics on student achievement in the three 'exam (selective) high schools in Boston and three in New York City. They find little evidence of gains from exam high school attendance or peer effects on student achievement, although some gains are found for minority students in English test scores. In contemporaneous, independent work, Dobbie and Fryer (2014) conduct an evaluation of the same three exam high schools in New York City. They find a null or even negative impact on college enrollment, college graduation, and the quality of the college attended. A similar, earlier paper by Dobbie and Fryer (2011) show similarly null effects for student achievement as Abdulkadiroğlu et al. (2014), although benefits are found in terms of participation in advanced coursework.

Clark (2010) examines U.K. grammar (selective) schools using a historical dataset, finding small effects on standardised test scores but large impacts on course-taking and university enrolment. In a separate study examining contemporary U.K. grammar schools, Atkinson et al. (2006) employs propensity score matching and find little overall impact on academic achievement, but substantive benefits to low socioeconomic status students. The authors argue that inequality of opportunity, which leads to high average socieconomic status of grammar schools, prevents more of these low socioeconomic status students from benefiting from these schools. In a novel study examining the long-term impacts of selective school attendance in Scotland in the 1960s, Clark and Del Bono (2016) find that there are large gains to years of completed education, but generally insignificant effects on labour market outcomes such as income, employment and wages. The authors, however, do document large, negative and statistically significant impacts on female fertility rates: a decrease in the number of children per woman of about 0.4 (20% of the mean).

The secondary school system in Jiangsu, China as studied in Ding and Lehrer (2007) exhibits tracking on a large scale, with entrance examination scores determining admission to a clearly ranked menu of schools. Their dataset is further enriched by standardised measures of teacher quality which allows them to separately examine the effects of peers and teacher quality, the latter of which tends to be correlated with high school ranking. Strong and robust evidence is presented for positive, non-linear (with larger benefits for higher-ability students) peer effects, and increases in peer quality variation are detrimental to all students. They find that teacher quality accounts for a significant portion of the variation in the school fixed effect. In a separate China study, Zhang (2008) finds no effect for selective middle schools.

In Mexico, Dustan et al. (2015) find adverse effects for students admitted to Mexico City 'elite' schools, who experience a large rise in high school dropout probability, especially those with weaker baseline academic achievement and with long commutes. The authors also find positive effects of elite schools in Mexico on student exit exam achievement, even adjusting for differential attrition. In an earlier related paper, Dustan (2010) finds that these effects disproportionately accrue to higher-SES students due to biases in the selection process (both in their entrance exam scores and school preferences), a similar finding to Atkinson et al. (2006) in the U.K. discussed above.

Jackson (2013) examines the secondary school system in Trinidad & Tobago, which exhibits substantial school choice. The author finds benefits to schools that are more selective and which have higher-performing peers, with non-linear effects such that attending a school on the lower end of increased peer achievement has little effect compared to on the medium-and-higher end. In addition, the marginal benefits of selective schooling are found to be twice as large for females as for males.

2.4 Evaluations of other tracking programs

The selective secondary school evaluations above are part of a broader education literature on tracking policies, such as primary-school selective programs, as well as within-school tracking policies. Cohodes (2015) examines the long-term impacts of a tracking program in Boston for children in years 4-6 on university entrance exam scores, on-time graduation, and college enrollment.[2]. This program is found to have very little effect on university entrance exam scores, but has a positive effect on Advanced Placement participation, especially in mathematics. Significant on-time graduation and college enrollment benefits for minority students are also found. The contributory channels identified are teacher quality and mathematics acceleration, but not peer effects.

Within-school tracking policies are examined in Figlio and Page (2002), who compare outcomes for lower-ability students in high schools that internally track based on ability and those that do not, in a manner similar to partially selective high schools in NSW.[3] The authors find no evidence that tracking harms lower-ability students, and that it may in fact be beneficial for them. In the Southwestern U.S., Bui, Craig, and Imberman (2014) find generally small-to-no achievement gains for within-school Gifted and Talented programs across a range of schools using both regression discontinuity and randomised lottery designs. In a separate study, Card and Giuliano (2014) examine the impact of separate Gifted and Talented classrooms for 4th grade students in a large, unspecified U.S. school district, with three admittance regimes: non-disadvantaged with IQ ≥ 130, disadvantaged/ESL students with IQ ≥ 116, and students who perform superlatively but do not meet the IQ thresholds. While effects are generally absent as in Bui et al. (2014), significant gains are found only for students selected on past achievement rather than IQ thresholds, and especially so for disadvantaged and minority students.

[2] This program resembles the Opportunity Class (OC) program in NSW, not examined in this paper.
[3] Internal tracking is also a common practice in many schools, both selective and non-selective.

Chapter 3
Institutional background and admissions process

3.1 Secondary education in New South Wales

The secondary education system in NSW comprises school years 7 through 12, with students typically transitioning from a primary school (K-6) to a high school (7-12). High schools in NSW include both government (commonly referred to as 'public') and non-government (commonly referred to as 'private') schools, and the overwhelming majority of schools use the state curriculum developed by the Board of Studies Teaching & Educational Standards NSW.[1] Government schools include non-selective, fully selective, and partially selective schools. The majority of non-selective schools are what are called 'comprehensive high schools', with others having some particular focus such as technology or performing arts. Non-government schools include religious schools such as Catholic and Islamic schools, and non-religious schools such as Independent schools. As an indication of relative costs, average yearly fees[2] were estimated at A$988 to attend a government high school in 2016; A$10,174 to attend a 'systemic' (religious) school; and A$23,524 to attend a private school (Australian Scholarships Group, 2016).

Selective schools are high schools for which the basis of admission is academic achievement and giftedness. The stated policy goals of NSW selective schools are to 'cater for gifted and talented students who have superior to very superior academic ability which is matched by exceptionally high classroom performance' NSW Department of Education (2015a), and 'to cater for [...] students who may otherwise be isolated from a suitable academic peer group [...] teaching them in specialised ways and providing educational materials at the appropriate level' (High Performing Students Unit, 2016b). Unlike comprehensive high schools, where attendance eligibility is mainly determined by residence in a catchment area, geography is not considered in the selective schools admissions process. For fully selective schools, selective admission applies to the whole school; for partially selective schools, only some classes are selective, with otherwise no admission criteria to attend the non-selective classes beyond the usual catchment area residence requirements. The curriculum taught at selective schools remains nominally the same as the state-set curriculum, although students may be expected to learn at a more rigorous or faster level.

In 2014, there were 464 government secondary schools, and 381 non-government secondary schools in NSW (Centre for Eduation Statistics and Evaluation, 2015). Of those government secondary schools, 21 were fully selective, and 24 were partially selective.

3.2 Selective schools admission process

The admissions process described in this section is that which applied to the families applying to NSW selective schools in the years 2007-2009, which differs slightly to the current process. Applications are open to all students who are citizens of Australia or New Zealand, or permanent residents of Australia. Parents apply to selective schools on behalf of their child through an online

[1] For example, a minority of schools offer the International Baccalaureate.

[2] While there are no tuition fees for government high schools, this figure also includes payments such as voluntary school contributions. See Australian Scholarships Group (2016) for details.

application, ranking up to four selective high schools in order of preference (since 2016, this has been reduced to three preferences). This typically occurs around October-November 2 years prior to the planned commencement of Year 7, i.e., when the student is in Year 5. The student undertakes the statewide standardised selective school test around March of the following year (during Year 6).

Students are assigned a total calculated profile (referred to as the 'selective score' henceforth) out of 300 that combines both school achievement in mathematics and English in Year 5, and the selective school test results, with equal weighting. The selective score is designed to incorporate both high achievement, in the form of the overall Year 5 performance, as well as giftedness, in the form of the selective test score. To this end, the selective tests resemble IQ tests, and as such are designed to be unrelated to the syllabus. The selective school test itself is comprised of multiple choice questions in English (reading), mathematics, and General Ability, and a piece of writing in response to a stimulus for English (writing).

A student's school achievement in mathematics and English is moderated using the mean and standard deviation of the respective subject-area selective test scores gained by all applicants from that school, in order to account for school-level differences. In this way, a student who, say, ranks 20th at a school with many high-achieving students will not be disadvantaged because of the relative strength of their peers, as compared to a student who ranks 20th at a school with few high-achieving students.

Students may also apply for entry into a specific selective school in later years, wherein entrance is usually determined by some combination of a school-specific test, high school results, interviews and/or other factors. As later-year applications to selective schools are decentralised and school-specific processes, such applicants are not considered in this analysis.

Student selection is formally carried out by school-specific selection committees that make decisions independently of each other. Selection committees resolve illness/misadventure claims and areas of special consideration, similar to selection regimes in other jurisdictions (e.g. Clark (2010)), while students who are clearly above or below a school's admission standards are offered or not offered a place respectively. Each committee includes a 'cluster' or 'school' director[3], the selective high school's principal, a parent or community representative, and can include other school staff members.

Students are matched with selective schools via a process that bears resemblance to a student-proposing deferred acceptance (DA) algorithm first described by Gale and Shapley (1962) (see Abdulkadiroğlu, Pathak, Roth, and Sönmez (2005) for a discussion of DA algorithms in the Boston Public School system). That is, schools consider the entire pool of students who have expressed a preference for that school, rejecting the students with the lowest selective scores in excess of its capacity (after allowing for a number of 'reserve' or waitlist placements). However, a student only receives an offer for the most preferred school out of the schools that a student has applied for and whose admission criteria the student meets. The NSW matching mechanism should thus be incentive compatible, eliciting truth-telling in students' preferences. A more detailed explanation of the process as it applied for the cohorts in this paper is as follows:

1. Applicants[4] simultaneously propose to up to 4 schools which have been ordered by preference.

2. For each selective school, a computer ranks proposing students in order of the selective score.

3. Each school's selection committee gives consideration to various additional admission factors that might affect a student's performance and adjusts scores (and therefore ranks) if

[3]Similar to a superintendent in U.S. jurisdictions; no longer a member of selection committees since 2014.

[4]Legally speaking, parents lodge applications and make decisions on behalf of their child.

necessary. These factors include: Aboriginality, applicants that are in Year 5 or 7, illness/misadventure, disability, students with fewer than 48 months of English-language education, and late applications.

4. For each selective school, a computer preliminarily assigns students an offer, rejection, or placement on a 'reserve' (or waiting) list, according to their selective score rank, number of vacancies at the school and the expected number of acceptances (schools over-offer to account for non-acceptances, including due to offers from a higher-preferenced school).

5. If a student is offered a place at one or more of their chosen selective schools, the student receives one offer for the highest preferred of these schools only.

6. Applicants choose to accept or decline their offer (if any).

7. Selection committees may make further offers to reserve-placed students in order of their selective score, depending on the number of acceptances, declines, and withdrawals. Note that since non-acceptances may occur quite late in the year, and that by this time most reserve-placed students will have confirmed enrolment at some other high school, these late-offered students may possess selective scores that are relative outliers compared to the initially accepted student body.

3.3 NSW Educational assessments in depth

3.3.1 National Assessment Program Literacy and Numeracy

NAPLAN (National Assessment Program Literacy and Numeracy) is a suite of basic skills tests, first introduced in 2008, that is administered to all enrolled Australian students in Years 3, 5, 7, and 9 in the 'domains' of reading, writing, language conventions (spelling, grammar, and punctuation), and numeracy.[5] As an assessment of basic literacy and numeracy, NAPLAN test scores are likely to feature significant truncation at the top end of performance. For example, the highest performance band in Year 7 corresponds to the second highest performance band in Year 9, implying that the academic ability of students performing beyond a superlative level at two year levels ahead is likely to be be underestimated and/or not finely differentiated from one another. In addition, NAPLAN tests are relatively low-to-zero stakes from the point of view of the student[6] which possibly affects student behaviour, and especially so for high-achieving students for whom the tests are often relatively easy.

3.3.2 Higher School Certificate

The Higher School Certificate (HSC) is the highest educational award in NSW schools, comprising in-school assessments throughout Year 12, and final state-standardised exams in a minimum of 10 units of standardised courses (BOSTES (2014)).[7] Performance in the HSC ultimately determines a student's Australian Tertiary Admission Rank (ATAR), which is a percentile rank of scaled and moderated overall HSC performance relative to that student's state-wide cohort as at Year 7 (UAC 2015). The ATAR is the primary and usually sole criterion[8] for undertaking tertiary

[5] Prior to 2008, basic skills testing was conducted by individual states.

[6] However, students might believe that NAPLAN tests are high-stakes if they or their parents hold other beliefs regarding NAPLAN. In addition, school averages are presented on the government's `www.MySchool.edu.au` informational website, which might cause schools to stress its importance.

[7] Additionally, the student must complete a minimum of 12 units of courses in the 'preliminary pattern of study' in Year 11.

[8] One exception to this are undergraduate Medicine degrees, for which interviews and performance on an additional Undergraduate Medicine and Health Science Admission Test (UMAT) are also considered.

study, determining which degree and at which university one can study. The moderation process is similar to that involved in the construction of the selective score described above: students' in-school performance throughout Year 12 are adjusted by the performance of their fellow peers in the final state-standardised exams in order to account for school-level differences in assessment standards (BOSTES, 2011). Scaling refers to adjustments to moderated HSC subject scores for the purpose of calculating the ATAR in order to account for the differences in relative difficulty or student competition between courses (Yager, 2016). Typically, courses are worth 2 units each. All students are required to take 2 units of English, but are otherwise relatively free to select courses from the menu of courses they face at their school. Students may elect to take more than 10 units of courses if they wish, but the best 2 units of English and the best other 8 units only are counted for the purposes of determining the ATAR.

Th menu of courses is determined by each school, likely based on supply of appropriate teachers/materials and demand by the school cohort. The following subject areas are divided into courses of varying levels of specialisation: English, mathematics, history, and languages other than English. In particular, English and mathematics are split into two base-level ('2 unit') courses each: the lower level courses being English Standard and Mathematics General 2; the respective higher level courses being English Advanced and Mathematics (2 unit). Students may choose either the lower or higher course of each subject but not both simultaneously. In addition, students may undertake separate, increasingly more advanced '1 unit' courses, namely English Extension 1, English Extension 2, Mathematics Extension 1, and Mathematics Extension 2. These advanced courses are cumulative — to take English Extension 1, the student must be studying English Advanced; to take English Extension 2, the student must take both English Advanced and English Extension 1. One exception to the separability of extension courses is Mathematics Extension 2, whose study replaces Mathematics (2 unit). That is, unlike the English Advanced/Standard English dichotomy, the alternative course to Mathematics (2 unit) is not only the lower-level 2 unit course (Mathematics General 2), but also the more-advanced Mathematics Extension 2 course.

The significant course choice in the HSC program allows me to examine selective school effects for participation in specific courses as outcome variables of secondary interest. While HSC performance is the major determinant of the ATAR and therefore of university entrance, assumed knowledge of various subjects are widespread for specific university degrees.[9] For example, the Bachelor of Science at the University of New South Wales (UNSW) in 2015 (ATAR cut-off: 84.00) assumes knowledge of Mathematics (2 unit) and Chemistry, plus one more of Biology, Earth and Environmental Science, Physics or HSC Mathematics Extension 1 (depending on the student's chosen area of study) (UNSW Australia, 2014).

The ATAR can be thought of as a 'sufficient statistic' for high school academic achievement, given that the ATAR is the figure ultimately supplied to students and used by employers and universities. However, for the purposes of my paper, I use the Tertiary Entrance Score (TES) as my primary outcome variable, an alternate measure of HSC achievement that proxies for ATAR due to legislated confidentiality restrictions on ATAR data (Yager, 2016). The TES is the post-scaling aggregate of the HSC subject scores achieved by a student (using the same subject rules described above) calculated by the DOE prior to transformation into the ATAR by the independent Universities Admissions Centre (UAC). The ATAR is thus a monotonically increasing function of the TES, although not strictly so as the ATAR percentiles are truncated into bins of size 0.05. While the ATAR ranges from 0-99.95 (although ATARs < 30 are only reported as such), the TES ranges from 0-500. Figure 3.1 gives a rough idea of the conversion between the TES (named 'aggregate' in this figure) and the ATAR for the 2008 cohort (2014 HSC year).

[9]In the early 2000s, some courses were prerequisites for enrolment into some university degrees.

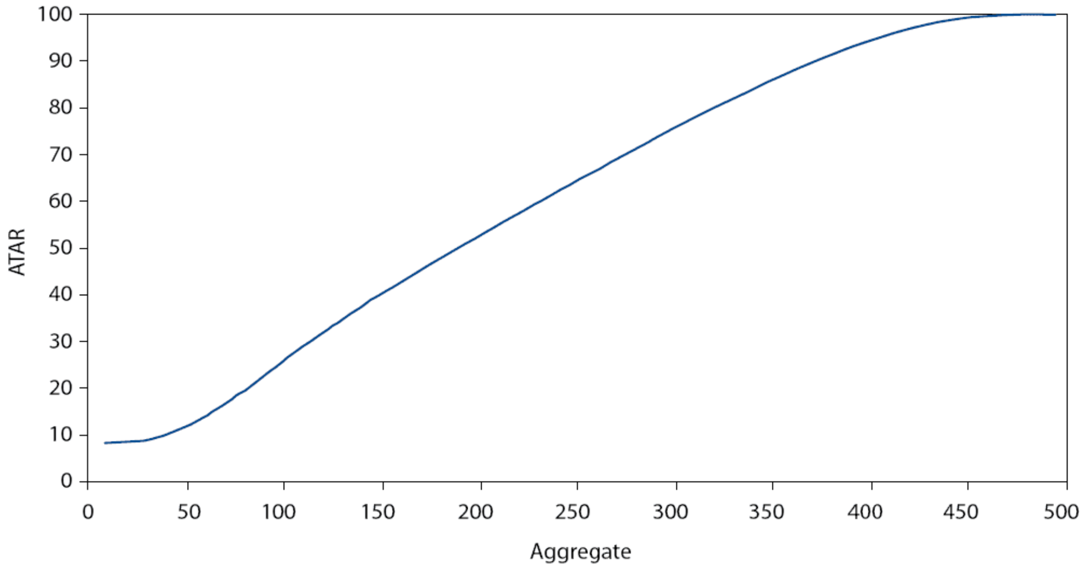


Figure 3.1: "Relationship between aggregate and ATAR, 2014."

Figure and caption reproduced verbatim from UAC 2015.

Chapter 4
Data

The datasets I use in this paper are provided by the Centre for Education Statistics and Evaluation (CESE), a unit within the NSW Department of Education (DOE). I use a student-level panel dataset of 38,585 total observations (19,388 with complete outcome data) comprising enrolment, demographic, selective school, and testing data for all students who applied to selective schools in the years 2007-2009. To illustrate how representative this sample is of the overall student population, 14.5% of all Year 6 students in NSW in 2007 applied to selective schools: 12,649 applicants out of a total population of 60,383 government and 27,017 non-government students in Year 6 (Planning and Innovation, NSW Department of Education, 2008). The included student cohorts were chosen such that both outcome data (HSC results), as well as baseline achievement data (at least Year 7 NAPLAN) are available. This involved matching across NAPLAN and high school enrolment datasets held by CESE, and a selective school dataset held by the High Performing Students Unit (a part of the NSW DOE).I also merge the student-level dataset with a high-school level panel dataset containing school characteristics and demographics.

Data on post-school outcomes such as university records, income and employment is unavailable, but would otherwise be useful in examining the longer-term impacts of selective schools. Similarly, data on non-academic outcomes such as student wellbeing, satisfaction, networking, etc. would be valuable inputs for future research once sufficient data on such measures is collected.[1]

4.1 Student-level dataset

I refer to the '2007 cohort' as all students who took the selective school test in 2007 (and therefore who were almost certainly in Year 6 in 2007, and who typically completed the HSC in 2013), with analogous definitions for the 2008 and 2009 cohorts. As a result, the dataset includes student data observed in the window of 2007-2015, with most students in the dataset contributing 7 years of data spanning Year 6 and their high school careers.

The enrolment data includes the type of primary school attended (whether it was a NSW government school or not), the government high school attended for each year (if any)[2], and the year level of the student for each year (if enrolled in a government high school). A primary school type that is coded as not a NSW government school are most likely private schools, although other schools such as those attended by overseas or interstate applicants (of any kind) are also included in this category.

I assign each student a 'principally attended school' based on attendance patterns. Students who attend the same government school for all six years of high school make up the vast majority of students with government school records, and so the principal school is straightforward to identify. For students with at least three years in government schools but who switch schools, I take the modal school as their principal school. Students with fewer than three years of attendance at

[1] More specifically, such data might come in the form of the DOE-implemented *Tell Them From Me* student, teacher, and parent surveys which commenced in 2015.

[2] The government school attended in a given year is determined by a census of student records which takes place on the first Friday of August of that year.

a government high school are considered educated by a non-government school, and assigned a missing value for their principally attended school. Any ties in the modal school break to the latest school attended, including if that school was a non-government high school.

Non-government schools do not share student data with the DOE, and as a result, no data is available from the high school years of students who took the selective test but went on to enrol at a non-government high school. This missing data excludes such students from the regression analysis.

The demographic information available in the dataset includes gender, postcode, language background other than English (LBOTE), the countries of birth of the student and up to two parents, the schooling, qualifications and occupational bands of up to two parents (see Table 5.3 for further detail), and a measure of school-level socioeconomic status in the form of their primary school's Family Occupation and Education Index (FOEI) (only for government primary schools). I construct an indicator for immigrant status, which takes on the value of 1 if both parents were not born in Australia (or if only one parent reports, and that parent reports not being born in Australia), and 0 otherwise.

The selective school data includes the school preferences expressed by the student (up to four), the moderated scores representing Year 5 achievement in English and Mathematics, the individual test component scores (Mathematics, English, Writing, and General Ability), the selective score ('calculated profile') used for admissions, and the application outcome and student decision (if applicable) for each preference.

Baseline testing data is available in the form of NAPLAN test scores for the reading and numeracy domains for Years 5 (for the 2009 cohort only), and 7. Year 5 NAPLAN scores are a measure of true pre-treatment academic achievement, whereas Year 7 NAPLAN testing technically takes place during treatment (in May of Year 7). However, the time-in-treatment period should be sufficiently short (approximately four months) compared to the overall intended treatment period (6 years) that Year 7 NAPLAN results can be still be considered baseline. The reading and numeracy NAPLAN domains were chosen as they are general in nature. The other English domains of Spelling, Grammar, and Writing were not available for analysis, although Houng and Justman (2013) find that such domains provide little additional information on students' English capability conditional on Reading. In addition, the writing domain experienced major changes in 2011, reducing comparability across years and tests.

Available outcome data includes Year 9 NAPLAN scores (a measure of mid-treatment outcomes), and HSC results, both for individual English, mathematics and science courses, and the overall Tertiary Entrance Score (TES). The HSC course achievement data encompasses subject areas with high participation rates and whose content is commonly assumed knowledge for university courses: English Standard, English Advanced, English Extension 1, English Extension 2; General Mathematics, Mathematics (2 unit), Mathematics Extension 1, Mathematics Extension 2; Biology, Chemistry, Physics, and Earth and Environmental Science. All subject scores range from 0-50.

I construct a measure of participation in HSC courses for each student by assigning a dummy variable indicating whether or not a score is present for that course. The term 'course participation' thus indicates full completion of that course — students who drop out of a course are not coded as participating. Completion of a subject is arguably of more interest than simple participation, although information about the drop rates of subjects (absent from this dataset) could be used to investigate other effects of selective schools.

Peer achievement for a student is calculated as the arithmetic average of NAPLAN scores for a given domain and year across the student's Year 7 schoolmates as at Year 5 and 7, excluding their own score. These measures have been calculated for all students in government schools at Year 7 in my sample, including students who attend non-selective government schools.

4.2 High school-level dataset

The high school-level panel dataset contains yearly data gathered on each school from the same high school year range (2008-2015) as is captured in the student-level dataset.

In addition to a unique school identifier which facilitates linking to the student-level dataset, the school-level characteristics (which did not change over the years considered) include year admission range (e.g. offering Years 7-12, or 11-12), postcode, region (rural or metropolitan), selective type (none, partial, fully), and indicators for other school types (agricultural, boarding, and co-educational).

Aggregate school demographics include enrolment (full-time equivalents), enrolment breakdowns by gender, indigenous, and language background other than English, number of teaching staff (full-time equivalents), total yearly net recurrent income (including government funding, parent contributions, and other sources such as donations), student attendance rate, FOEI, Index of Community Socio-economic Advantage (ICSEA, a similar school-level measure of socioeconomic status), and the percentage of students in each ICSEA quartile.

Government funding for schools varies by demographic factors such as socioeconomic background of students and location (NSW Department of Education, 2015b), and so is unlikely to differ systematically between selective and non-selective schools (beyond that which is attributed to differences in student demographics).[3] School funding also includes funding from sources such as voluntary parent contributions, which may differ across schools.

From the above aggregate school demographics, I construct simple student-teacher ratios and school incomes-per-student, allowing comparison between schools with different levels of enrolments.

To merge the student- and school-level datasets, I take an average across all available years of the school-level dataset variables, and then match one school to each student based on the school they principally attend.[4]

4.3 A note on the FOEI and ICSEA socioeconomic status indices

As mentioned above, FOEI and ICSEA are separate school-level indices of socioeconomic status as they relate to academic performance, and are developed by CESE and the Australian Curriculum Assessment and Reporting Authority (ACARA), respectively. FOEI is only available for government schools; school-level ICSEA values are also available for Catholic and most Independent schools, although since I do not possess data on student-level private school enrolments, ICSEA values cannot be matched to the students in my dataset.

FOEI is constructed from an average of parental education and occupational variables weighted by the extent to which these variables predict academic performance (as measured by NAPLAN) in a multiple regression model estimated by Rickard and Lu (2014) for the DOE. The set of socioeconomic variables included in FOEI includes:

- Parental secondary schooling
- Parental qualifications

[3] Reforms to NSW education funding began in 2014, although the general principles of 'needs-based' funding remain relatively unchanged. See PwC (2013) for more details.

[4] Variable variation within schools over time is generally low, especially compared to variation between schools.

– Parental occupation group

– 10 community-level variables based on students' addresses from the Australian Bureau of Statistics (ABS)

– Aboriginality

– School remoteness classification

FOEI scores are roughly normally distributed, normalised to have a mean of 100, a standard deviation of 50, and a range of -30 to 300, where a lower value indicates higher socioeconomic advantage. For clarity reasons, I reverse the direction of FOEI by multiplying school FOEI scores by -1. Thus, the 'reverse' FOEI ranges from -300 to 30, and a higher reverse FOEI value indicates higher socioeconomic advantage.

The construction of ICSEA is broadly similar and includes the same basic variables as FOEI (Australian Curriculum Assessment and Reporting Authority, 2014).[5] ICSEA values are standardised to have a mean of 1,000 and a standard deviation of 100; the higher the ICSEA score, the more advantaged a school's student population is.

4.4 Constructing the analysis dataset

From my initial student-level dataset, I restrict my sample in the following ways and in the following order:

1. Exclude students who did not attend a government primary school in NSW at time of application.
2. Exclude students who repeated years of high school.
3. Exclude students who are admitted to any selective school in Years 8-12, such as students who switch into selective schools from other selective schools.
4. Exclude students with no recorded 1st preference.
5. Exclude students with selective scores equal to zero.
6. Exclude students enrolled in government schools in Year 12 without a TES.
7. Exclude students enrolled at a government school in Year 12 with no recorded score for any of the 2 unit English courses.

Table 4.1 documents the number of observations dropped at each of the steps above.

I exclude students from non-government primary schools to reduce differential attrition of students out of the public high school system as such students are more likely to also attend non-government secondary schools. This also focuses my analysis on the group of students likely facing reduced school choice and lower socioeconomic status. Following Dobbie and Fryer (2011) and Abdulkadiroğlu et al. (2014), I conjecture that potential school effects would be largest and of more interest for this restricted sample.

In the unrestricted sample (Column 1 of Table 4.1), 62.5% of students coming from non-government primary schools go on to principally attend a non-government high school, compared with 22.3% of those coming from government primary schools. This seems to indicates that those

[5]For more detail on the construction of FOEI and ICSEA, see Rickard and Lu (2014) and Australian Curriculum Assessment and Reporting Authority (2014) respectively.

who come from non-government primary schools enjoy a greater school choice. Moreover, the presence of an offer has a clear effect on the high school type attended: 34.5% of students who are not offered a place at any selective school go on to principally attend non-government secondary schools as compared to 16.6% of those who are offered a place. After restricting the sample (up to and including Column 7 of Table 4.1), the corresponding non-government school attendance rates are 29.7% and 15.6% respectively.

I exclude students who apply to selective schools between 2007-2009, but gain entry to selective schools in years later than Year 7, as this selection process is idiosyncratic as previously mentioned.[6] This includes any student who switches between selective schools. An average of 8.1% of students principally attending fully selective schools are students who have switched into that school. The switching proportion at the most selective school is 15.3%. After restricting the sample (up to and including Column 7 of Table 4.1), 95% of selective school students who attend a selective school in Year 7 stay at that selective school for all 6 years (Figure 4.1).

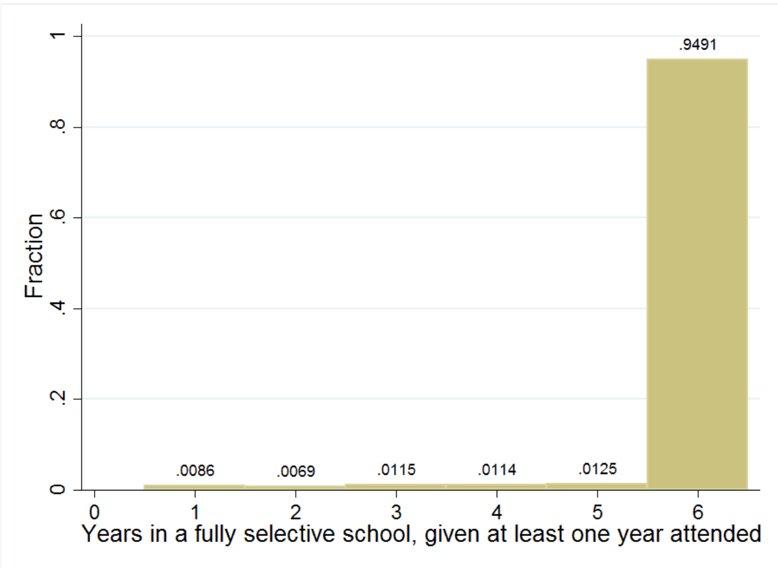


Figure 4.1: Distribution of years that a student attends a fully selective school, conditional on at least one year's attendance (after sample restrictions).

I exclude students with repeated years for simplicity and to reduce the impact of outliers. I exclude students with selective scores recorded as 0, the vast majority thereof having been coded as 'absent from the test without excuse'. I also exclude students who are missing data of some kind that should be present, such as TES if they were in Year 12 at a government school, or if there is no 2 Unit English course score recorded in their expected HSC year (and who otherwise would be expected to have a TES), as at least 2 units of English is compulsory in the HSC.[7]

In Column 8 of Table 4.1, I also indicate the number of observations that are generally available for estimation by excluding all students missing TES data, the chief cause of which is the student attending non-government high schools. Available sample sizes for other outcome variables, such as individual subject scores, are smaller since not all students take all courses. The restrictions above reduces my total analysis sample size from 38,585 to a restricted sample size of 28,635, of which 19,388 observations can potentially be used for estimating selective school effects on the TES. Descriptive statistics and graphs are based on the restricted sample where possible and

[6]Note that this dataset does not contain students who did not apply for entry into selective schools in Year 7, but did apply to individual schools in later years.

[7]This missing data is likely a result of matching failure or errors in data collection.

appropriate, and unless otherwise stated. Any variable standardisation or normalisation process used in the analysis occurs after sample restriction (up to and including Column 7 of Table 4.1) unless otherwise stated.

Table 4.1: Sample restriction process

	Unrestricted sample							Restricted sample available for most descriptive statistics	Indication of sample available for most estimation
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Application year	Total observations	Excluding non-government primary school applicants	Excluding students who repeat year(s)	Excluding students admitted to selective schools in Years 8-12	Excluding students with no 1st preference	Excluding students with selective scores of 0	Excluding students in government schools in Year 12 without a TES	Excluding students with no HSC English score	Excluding all students without a TES
2007	12649	10796	10740	10115	10114	9751	8903	8878	5833
2008	13195	11386	11330	10739	10732	10334	9982	9981	6873
2009	12741	11034	11033	10374	10368	10311	9777	9776	6682
All years	38585	33216	33103	31228	31214	30396	28662	28635	19388

Column numbers 1-7 correspond to the the sample restriction list above.

4.5 Constructing school sharp samples

From the universe of selective high schools, I focus on a set of fully selective high schools, in the sense that these are the only schools for which I estimate treatment effects. I exclude partially or fully selective boarding schools (Farrer Memorial Agricultural High School, Hurlstone Agricultural High School, and Yanco Agricultural High School) and one virtual school (xsel)[8] from consideration, given their additional admission criteria (such as location) and idiosyncratic school environments.

I also exclude partially selective schools from analysis for several reasons. First, fully selective schools dominate partially selective schools in terms of preferences and enrolments. Out of the 41 non-boarding/non-virtual selective schools, the 18 fully selective schools constitute 83.0% of first preferences in the unrestricted sample. On first preferences, all 18 fully selective schools are preferred to all partially selective schools, with the exception of only one partially-selective school. Furthermore, the 9 most preferred fully selective schools make up 51.4% of first preferences and the most first-preferred school represents 8.3% of first preferences. The 23 (non-boarding) partially selective schools only offer a small proportion of selective places, constituting only 8.41% of students' principal schools in my sample. As a result, sample sizes for individual schools would likely be too small for precise estimation of partially selective school effects.

Second, effects of partially selective school attendance might be difficult to interpret, as the treatment experienced by such students is additionally affected by non-selective stream peers and the unobserved degree of mixing between peer groups (assuming the existence of peer effects).

I refer to the remaining set of 18 schools as fully selective high schools henceforth, and these schools are the focus of all subsequent analysis. Eight schools in this set are single sex (four for each gender), and one school is agricultural. These schools are:

[8]The xsel virtual selective school program was superseded by Aurora College in 2015.

- Baulkham Hills High School
- Caringbah High School
- Fort Street High School
- Girraween High School
- Gosford High School
- Hornsby Girls High School
- James Ruse Agricultural High School
- Merewether High School
- Normanhurst Boys High School
- North Sydney Boys High School
- North Sydney Girls High School
- Northern Beaches Secondary College Manly Campus
- Penrith High School
- Smiths Hill High School
- St. George Girls High School
- Sydney Boys High School
- Sydney Girls High School
- Sydney Technical High School (boys school)

I partition the overall sample of students into 'school sharp samples' in a similar fashion as Abdulkadiroğlu et al. (2014). These school sharp samples are 'sharp' in the sense that the selective score of a student in a school sharp sample is directly relevant to whether or not that student receives an offer for that school. That is, the sharp sample of school k contains all the students who were in contention for a place at that school: students who preferenced k first, or who were rejected from their first preference and preferenced k second, and so on for all four preferences, for all cohorts $t \in \{2007, 2008, 2009\}$. Note that students can be in multiple sharp samples: for example, suppose student i preferences school k first, l second, and m third; and suppose they are rejected by k but accepted at l. In such a case, student i appears in the sharp samples for both schools k and l, but not for school m.

73.9% of applicants in the restricted sample list at least two selective school preferences, 61.4% at least three, and 49.4% list a school for all four preferences. 2.56% of students also receive more than one offer, which usually occurs when reserve places open up at a selective school after the initial round of offers due to students withdrawing.[9] In addition to meeting the above criteria for school sharp sample membership, these students would also appear in the sharp samples for all schools for which they received an offer — including the school that initially offered a place (if applicable), and any reserve schools that eventually offered a place.

4.6 Cut-off determination

While the initial stage of the selection mechanism might assign cut-off scores based on availability of places and preferences, no precise data on these cut-off scores exist.[10] Furthermore, while initial cut-offs might hold for the first round, selection committees can offer students places after other students have withdrawn, possibly much later in the year than most students. Since many of the rejected-but-close students may have already accepted places or enrolled in other schools, these late-offer students may have selective scores significantly below the original cut-off score. However, I do not observe if students were offered a place in the first round or afterwards, presenting the problem of selecting an implicit cut-off score for each school and each year.

[9] Another, though likely less prevalent, possibility is that the parents request a change in their preferences due to, say, moving suburb.

[10] Indicative cut-offs are published by the DOE, although they are calculated as the simple minimum score amongst all students who accept a place at a given selective school, and the cut-offs are rounded to whole numbers (NSW Department of Education, 2016).

The candidate cut-offs for a given school in a given year are drawn from the pool of selective scores c_i of students who were offered a place at that school in that year. Variables throughout this paper such as in Equation 4.1 are indexed by school k, individual i, and/or cohort t. I determine the optimal school- and year-specific cut-off for each school sharp sample in each cohort year (denoted by $\tau_{k,t}$) by running regressions of the form described by Equation 4.1 over all candidate cut-offs in said sample. These regressions predict offers at a given school in a given year using variables calculated from each candidate cut-off, and other (cut-off invariant) individual covariates.[11]

$$\text{Offer}_i = \beta_0 + \beta_1 A_i^\tau + \beta_2 \text{Gap}_i^\tau + \beta_3 A_i^\tau \cdot \text{Gap}_i^\tau + \lambda' X_i + \epsilon_i \quad \tau \in \{c_i | \text{Offer}_i = 1\} \tag{4.1}$$

$A_{i,}^\tau$ is a dummy variable that takes on the value of 1 if the student's selective score is above the candidate cut-off value τ. Gap_i^τ is the difference between the student's selective score c_i and the candidate cut-off. X_i is a vector of pre-treatment covariates: gender (not included for single-sex school sharp samples), language background other than English or not, primary school FOEI, and Year 7 NAPLAN Numeracy and Reading scores. The assigned cut-off $\tau_{k,t}$ for each school at each year that I use for all subsequent analysis is the cut-off candidate whose corresponding regression has the highest F-statistic.

The lowest cut-off averaged across the three cohorts is 181.3, the highest is 242.3 (both lowest and highest cut-off schools are the same for all three cohorts), and the standard deviation of cut-offs is 16.7. In comparison, the average student selective score across the cohorts in the sample is 175.8 with a standard deviation of 31.2. Hence, roughly two standard deviations separate the selective scores of students who just make the least-selective cut-off and those who just make the most-selective cut-off. An average of 3.6 students below the calculated cut-offs are offered a place ('crossovers'), and an average of 1.2 students above the calculated cut-offs are not offered a place ('no-shows').

I show the variation in school selectivity in Figure 4.2 in a histogram of student selective scores against an overlay of the assigned cut-off scores for each of the 18 fully selective schools for the 2009 cohort. Selective schools clearly accept students in the upper half of the performance spectrum. In particular, apart from the middle subset of schools which vary roughly evenly in selectivity ('mid-low/mid-high cut-off'), there are three schools that constitute a relatively less-selective subset ('low cut-off'), and one school that is relatively more selective ('high cut-off'). The same overall pattern is found in the 2007 and 2008 cohorts, with school selectivity remaining fairly stable over 2007-2009. The biggest change in this constructed selectivity ranking amongst fully selective schools over the three years is two places, which only one school experienced.

I calculate a generalised degree of selectivity by averaging a school's cut-offs across the three years (Figure 4.3). I then order schools by these averaged cut-offs, which I use to anonymise schools; A indicates the lowest cut-off school and R the highest.

[11] This equation is similar to the first stage of the fuzzy regression discontinuity estimating equation (Equation 6.1) in discussed in more depth in Section 6.1.1.

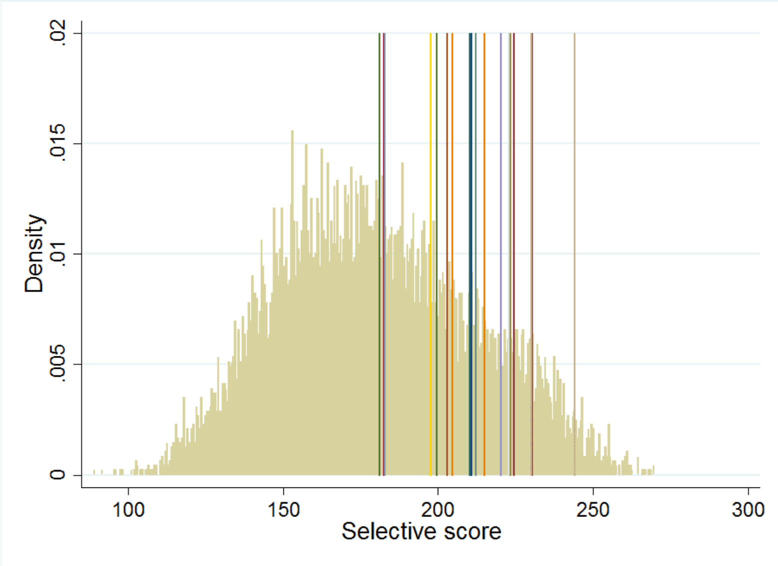


Figure 4.2: Selective scores and assigned fully selective school cut-offs (as vertical lines) in 2009

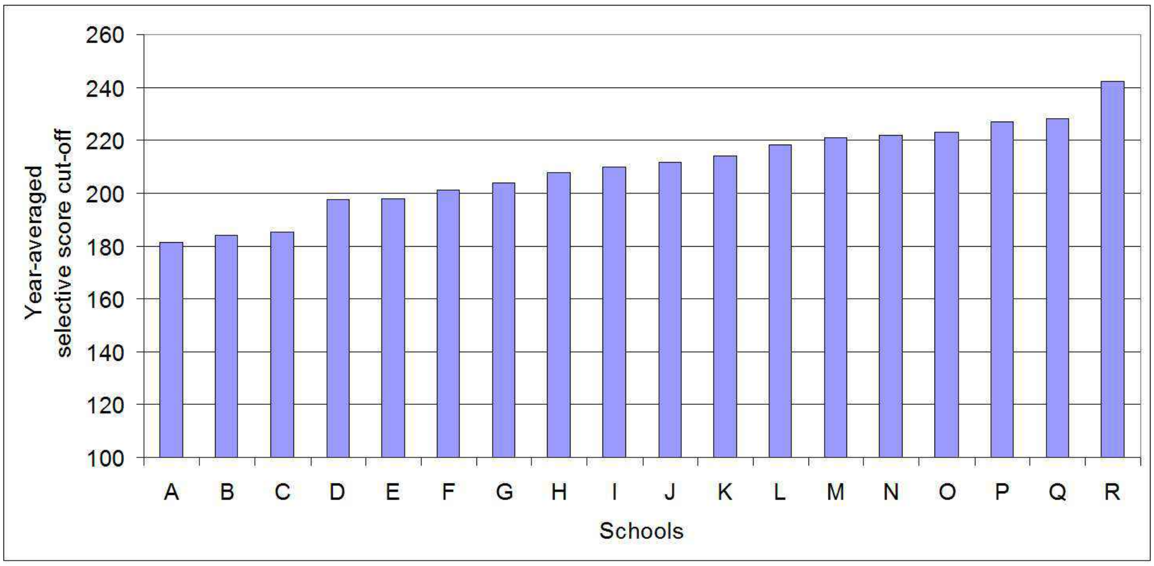


Figure 4.3: Cut-offs for fully selective schools averaged over 2007-2009

4.7 Constructing stacked sharp samples

To improve the precisions of estimates by increasing sample size, I then construct four stacked sharp samples by pooling together school sharp samples with similar levels of selectivity as follows:

- Low cut-off stack (3 schools: A, B, C)
- Mid-low cut-off stack (7 schools: D, E, F, G, H, I, J)
- Mid-high cut-off stack (7 schools: K, L, M, N, O, P, Q)
- High cut-off stack (1 school: R)

I construct the low cut-off stack from the three lowest cut-off schools and the high cut-off 'stack' from the highest cut-off school as these schools constitute outliers in terms of selectivity. The remaining 14 schools, whose cut-offs are roughly evenly spaced out, are divided into equal-sized stacks. The stacking process is similar to the sharp sample construction process described above, where a stacked sharp sample is composed of all students in the sharp sample of all the schools that are members of that stack. Just as students can contribute observations to multiple school sharp samples, students can contribute to multiple stacked sharp samples. Moreover, students can also contribute multiple times to the *same* stacked sharp sample: for example, a student who is in the sharp samples for Schools D and H will contribute two observations to the mid-low cut-off stack. A tabulation of the number of students who contribute multiple times is shown in Table 4.2: for example, 3,409 students contribute two observations to the mid-low stack.

Table 4.2: Incidence of multiple observation contributions in stacked sharp samples

No. of contributions	Stack			
	Low	Mid-low	Mid-high	High
1	4200	10074	9273	2258
2	131	3408	5201	0
3	4	163	1517	0
4	0	4	105	0

Not excluding students with no TES

4.8 Selective score standardisation

The selective scores for students $c_{i,t}$ in the sharp sample of school k are centred around the cut-off $\tau_{k,t}$ corresponding to that student's cohort and school, and standardised by the standard deviation $\sigma_{k,t}$ of the scores of students in that cohort and in that sharp sample to form school-specific standardised selective 'z-score(s)' $Z_{i,k}$.[12] In this way, I am able to pool together cohorts which differ slightly in selective score distributions and cut-offs. This also has the effect of normalising all cut-offs to zero in each sharp sample.

$$Z_{i,k} = \frac{c_{i,t} - \tau_{k,t}}{\sigma_{k,t}} \tag{4.2}$$

[12] Note that the term 'z-scores' is a slight misnomer given that these scores are not mean zero.

The standardised scores for stacked sharp samples are applied in the same manner (that is, scores are still standardised by school and cohort, not by stack). Hence, in the case of a student in the sharp samples for multiple schools in a stacked sharp sample, each school-specific z-score will appear once in the stacked sharp sample.

Chapter 5
Descriptive statistics and graphs

In this chapter, I describe the characteristics of students and schools in my sample and provide visual representations of how these characteristics change at sharp sample cut-offs. This chapter lays the groundwork for the econometric framework described in Chapter 6 and implemented in Chapter 7.

Students who apply to selective schools exhibit self-selection on demographic and achievement measures, being more likely to come from language backgrounds other than English, to come from higher socioeconomic status primary schools, and to achieve higher on baseline tests compared to the average NSW student. Compared to the average NSW comprehensive high school, selective school environments are characterised by significantly higher average academic achievement, higher proportions of students from language backgrounds other than English, slightly lower indigenous proportions, higher student-teacher ratios, and higher overall attendance.

That applicants are self-selected on achievement seems relatively unsurprising, given that applicants have learned about their ability over their time in school. The fact that applicants self-select on demographic variables is likely a combination of institutional/informational biases (e.g., low socioeconomic status families may be less informed about selective schools) and overall structural correlations between achievement and demographics.

Table 5.1 summarises sample means for demographic, enrolment, and baseline test score variables. Table 5.5 summarises sample means for outcome variables. In both tables, the figures in the first two columns are calculated using the entire non-restricted sample (Column 1 in Table 4.1) to illustrate the effect of restricting the sample. The figures in the last five columns are calculated using the restricted sample (up to and including Column 7 of Table 4.1), and supply the statistics I use to explore the data unless otherwise stated.

Table 5.2 summarises school-level differences in characteristics and demographics of the government schools that are attended by students in the sample, with schools weighted by their number of 'principally attending' students in each of the column subgroups. I also provide graphs of the differences in the school environments across the stacked sharp sample cut-offs in order to pinpoint what, precisely, the selective school treatments involve.

5.1 Demographics

5.1.1 Gender

Students in the sample are split roughly evenly on gender lines (49.1% female), with students who are offered a selective school place less likely to be female (46.3%) compared to students who are not offered a place (50.1%). Figure 5.1 shows the slight male-bias in selective scores. The points in cohort-wide scatter plots such as Figure 5.1 represent binned averages, using 100 evenly-spaced bins, with either linear or quadratic smoothed lines. Overall scatter plots of the relationship between variables and non-standardised selective scores are presented for the 2009 cohort only for space reasons (as the selective score distribution differs slightly across cohorts, and I do not

Table 5.1: Student-level means of demographic, enrolment, and baseline test score variables for various subgroups.

	All selective school applicants (unrestricted)	From non-government primary schools	All selective school applicants (restricted)	Students coded as not in government secondary schools	Not offered a fully selective school place	Offered a fully selective school place	Accepted a fully selective school place
Total observations	38585	5369	28635	7019	21367	7268	6229
Demographics							
Female	0.488	0.475	0.491	0.478	0.501	0.463	0.458
LBOTE	0.588	0.541	0.596	0.517	0.563	0.687	0.726
Reporting parent(s) foreign born	0.562	0.524	0.567	0.460	0.526	0.680	0.714
Aboriginality	0.017	0.077	0.011	0.015	0.013	0.005	0.003
Parent 1 completed Year 12	0.752	0.816	0.750	0.749	0.702	0.860	0.865
Parent 2 completed Year 12	0.756	0.784	0.756	0.737	0.705	0.868	0.872
Parent 1 Bachelor degree or above	0.445	0.517	0.442	0.457	0.394	0.542	0.542
Parent 2 Bachelor degree or above	0.508	0.546	0.506	0.513	0.439	0.643	0.643
Parent 1 Not in paid work in last 12 months	0.218	0.149	0.224	0.196	0.231	0.211	0.212
Parent 2 Not in paid work in last 12 months	0.031	0.036	0.029	0.039	0.034	0.020	0.020
Enrolments							
Non-public primary school	0.139	1.000	0.000	0.000	0.000	0.000	0.000
Primary school FOEI (reversed)	-55.10		-54.77	-46.26	-60.95	-36.62	-38.38
Years attending public high school	4.126	1.999	4.338	0.151	4.047	5.194	5.742
Principally attended government high school	0.721	0.375	0.755	0.000	0.713	0.877	0.969
Offered fully selective school place	0.244	0.228	0.254	0.127	0.000	1.000	1.000
Baseline and selective test scores							
Selective score	175.79	174.33	180.31	173.80	167.18	218.90	219.49
Average score of Year 7 peers in Year 5 NAPLAN Reading	527.62	551.60	525.94	497.01	495.87	601.93	606.97
Average score of Year 7 peers in Year 5 NAPLAN Numeracy	542.37	575.13	540.15	498.04	498.46	645.47	653.22
Average score of Year 7 peers in Year 7 NAPLAN Reading	578.33	602.80	577.69	549.58	546.96	651.10	656.24
Average score of Year 7 peers in Year 7 NAPLAN Numeracy	607.88	641.59	606.95	563.53	561.94	714.48	722.75
Year 5 NAPLAN Reading score	556.43		556.46	552.52	538.45	609.95	607.32
Year 5 NAPLAN Numeracy score	573.76		573.22	556.89	546.58	652.32	653.48
Year 7 NAPLAN Reading score	608.03	623.29	607.75	585.54	586.99	657.08	657.11
Year 7 NAPLAN Numeracy score	647.04	665.41	646.40	600.92	614.42	722.34	724.31
Year 9 NAPLAN Reading score	638.77	650.51	638.70	591.74	620.03	681.83	681.67
Year 9 NAPLAN Numeracy score	686.29	702.76	685.86	596.02	654.93	757.22	759.12

The number of observations available for each variable depends on missing data for that variable, including due to survey non-responses. Means calculated with fewer than 100 observations are excluded.

Table 5.2: **High-school-level weighted means for various subgroups.**

	All government high schools in sample	Non-selective government high schools	All selective schools	Fully selective schools
Number of schools	371	326	45	18
Types				
Regional = 1, Metropolitan = 0	0.031	0.026	0.038	0.000
Co-educational = 1	0.717	0.750	0.667	0.561
Demographics				
Enrolments (FTE)	978.21	1008.29	931.26	937.9
Male (%)	50.09	47.09	54.77	53.81
Indigenous (%)	1.83	2.33	1.04	0.21
LBOTE (%)	54.09	45.29	67.83	69.98
FOEI (reversed)	-58.43	-77.59	-28.52	-9.01
ICSEA	1071.16	1035.93	1126.14	1163.27
% of students in bottom ICSEA quartile	18.26	24.27	8.88	3.7
% of students in second ICSEA quartile	17.4	21.39	11.15	7.18
% of students in third ICSEA quartile	26.66	27.74	24.98	23.17
% of students in top ICSEA quartile	37.7	26.62	55.01	65.93
Resources				
Attendance	92.85	91.53	94.9	95.81
Students per teacher	13.93	13.73	14.23	14.52
Teaching staff (FTE)	70.07	72.83	65.75	64.51
Income per student	12316.78	12139.61	12593.32	12049.75
Net recurring income	11757141	11913438	11513179	11278817

FTE: Full-time equivalents

standardise selective scores by year). Analogous plots for the other cohorts reveal similar patterns to the 2009 plots.

The lower representation of females at fully selective schools might reflect the general finding in the education literature is that male students are more likely to produce test results at both the high and the low ends of the scale than female students in science and mathematics, and that the opposite is true for reading (Pope and Sydnor, 2010). In the NSW context, it is possible that the slight gender bias in selective scores arises from male advantage in mathematics dominating the female advantage in reading at the high end of the ability spectrum, and/or that the general ability section more highly correlates with mathematics performance. Pope and Sydnor (2010) also find large geographical variations in this gender differential, indicating that this gender divide likely reflects specific societal contexts and/or educational practices. Another possibility is that, as Azmat, Calsamiglia, and Iriberri (2014) suggest, girls outperform boys in low-stakes tests (such as small, frequent tests throughout the Year 5 school year), but that boys outperform girls in high-stakes tests (such as the selective test). While selective scores are constructed through an equally-weighted combination of school and selective test scores (similar to the construction of HSC scores), the latter phenomenon might dominate.

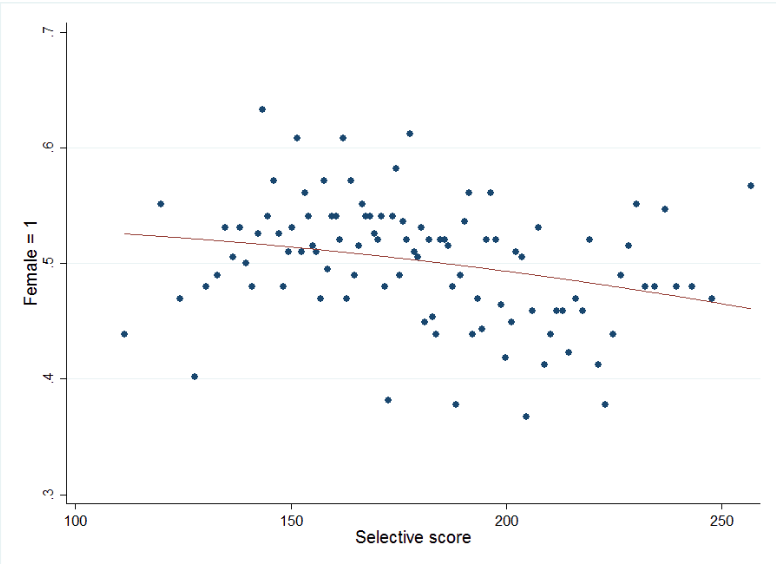
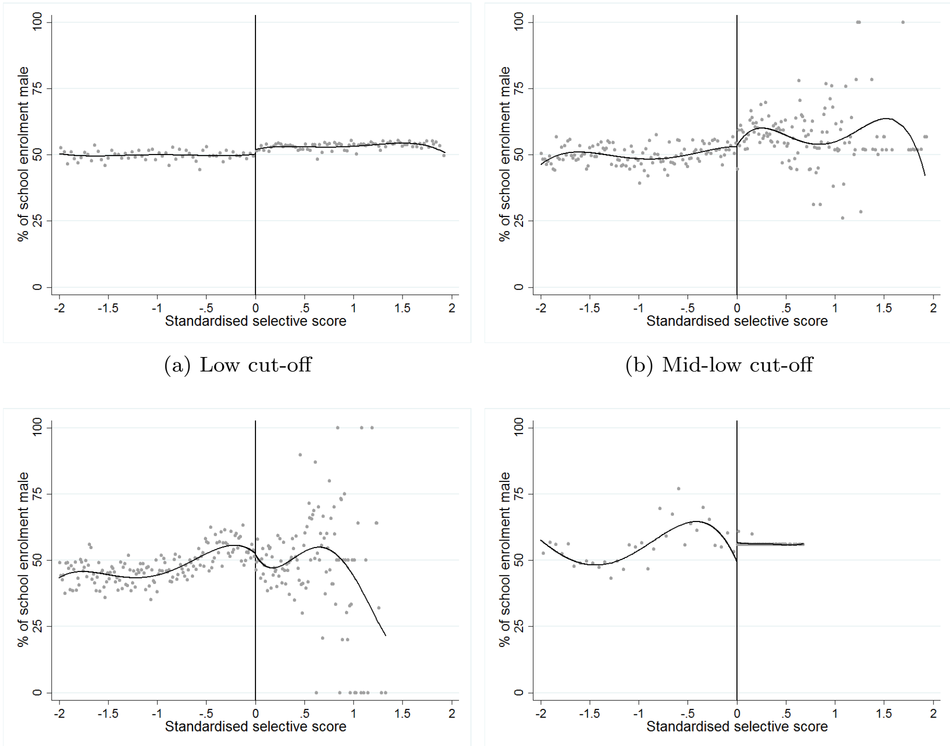


Figure 5.1: Individual gender and selective score for the 2009 cohort

Figure 5.2 shows that the male proportions of students' principally attended schools are relatively continuous across stacked sharp sample cut-offs, indicating that the gender disparities are relatively minor. The points in regression discontinuity plots (i.e., scatter plotswith a vertical line denoting standardised selective score cut-offs) such as Figures 5.2 in this and later sections represent binned averages with bin sizes determined by the optimal data-driven bin selection method described by Calonico, Cattaneo, and Titiunik (2015).[1] I restrict the regression discontinuity plots to observations whose school standardised selective score is within ± 2 standard deviations of their school's cut-off (0, by construction from the normalisation process). In addition, in the case of stacked sharp samples, I drop all observations with standardised selective scores that are distance-zero from the cut-off, as recommended by Fort, Ichino, and Zanella (2016), who demonstrate that not doing so for stacked and standardised regression discontinuity plots may show misleading discontinuities. The smoothed lines are global polynomial estimates of order 4. I specify a selection method that results in evenly-spaced bins separately estimated for observations above and below

[1] Implemented by the Stata package *rdplot*. See Calonico, Cattaneo, and Titiunik (2014a) for more details.

the cut-off, with bin sizes selected such that the within-bin variances mimic the overall variance of units above and and below. As a result, the regression discontinuity plots presented in this paper use different bin sizes on either side of the cut-off.



(a) Low cut-off

(b) Mid-low cut-off

(c) Mid-high cut-off

(d) High cut-off

Figure 5.2: Male proportion of school enrolment in stacked sharp samples

5.1.2 Socioeconomic status and language background

59.6% of students in the restricted sample come from a language background other than English (LBOTE), compared with just 32.1% of all government secondary students as at 2013 (Centre for Eduation Statistics and Evaluation, 2014). LBOTE students tend to exhibit extreme selective scores, especially at the top end, as illustrated in Figure 5.3. Students receiving offers are even more likely to come from a language background other than English: 68.7% of students offered a place are LBOTE students. The same broad conclusions can be drawn when looking at the measure of student immigrant background.

The school enrolment proportion of students with a language background other than English is slightly discontinuous at the stacked sharp sample cut-offs (Figure 5.4), most clearly at the mid-low cut-off stack. As selectivity increases, the school LBOTE proportion of students rises sharply, and exceeds 90% at highly selective schools.

Aboriginal/indigenous students make up only 1.1% of the restricted sample. Across the set of fully selective schools, the indigenous proportion of a student's school decreases discontinuously at the low cut-off stack (Figure 5.5), though indigenous proportions at schools attended by students below the low cut-off tend to be relatively low on average to begin with (4-6%). In general, the indigenous proportion tends to zero as selectivity increases.

Selective school applicants tend to attend primary schools that are relatively advantaged on average. The average government primary school Family Occupation and Education Index

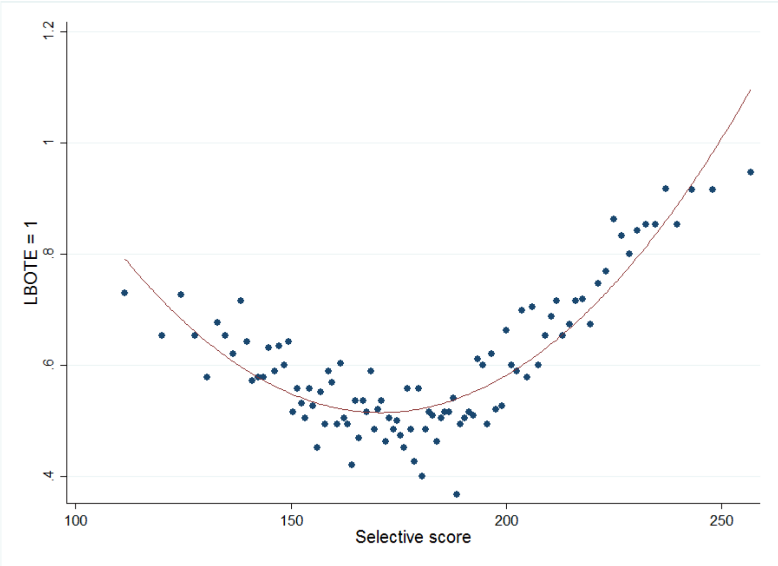
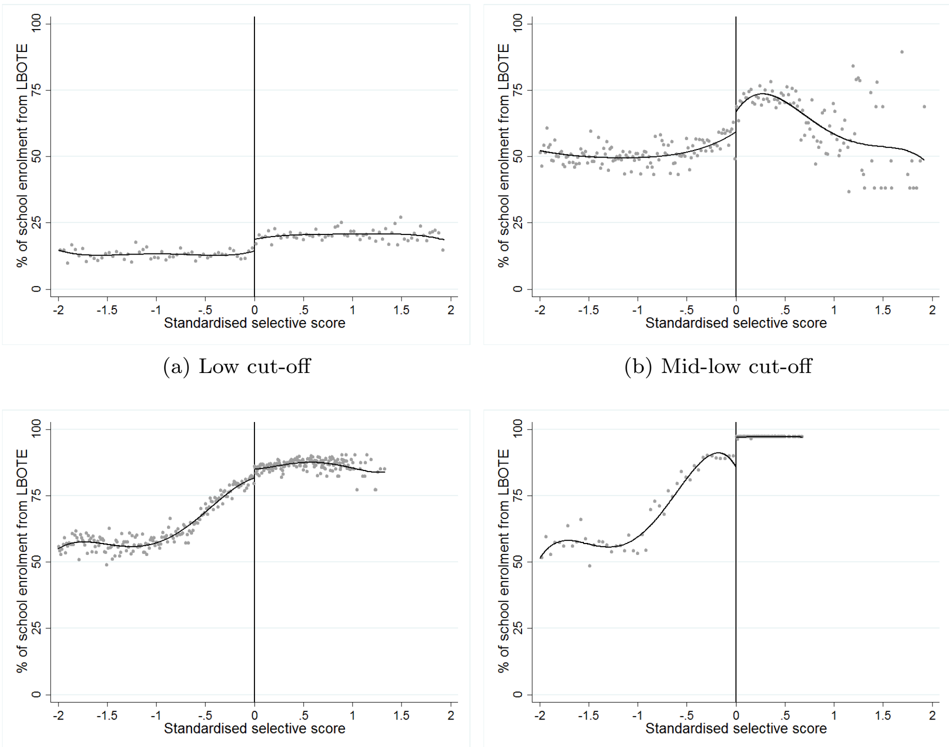


Figure 5.3: Individual LBOTE and selective score for the 2009 cohort



(a) Low cut-off

(b) Mid-low cut-off

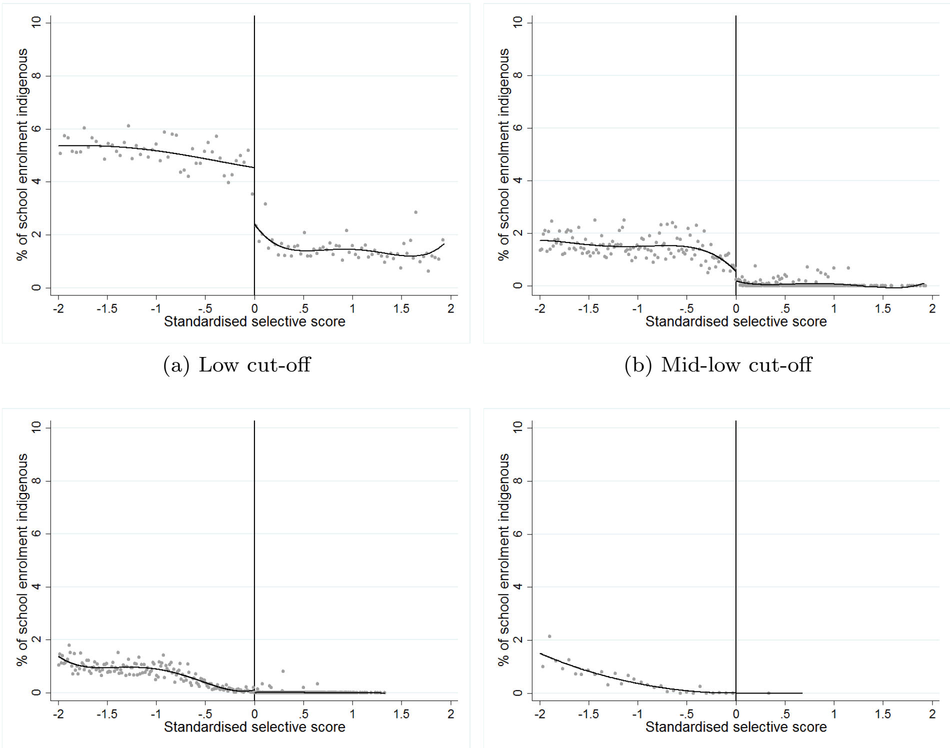
(c) Mid-high cut-off

(d) High cut-off

Figure 5.4: LBOTE proportion of school enrolment in stacked sharp samples

(reversed so that a higher number corresponds to a higher socioeconomic status) of applicants is -55.10 (-54.77 for the restricted sample). This represents a 0.9 standard deviation increase in school socioeconomic status above the state average.[2] Since it is likely that non-government primary schools (for which FOEI values are not calculated) have a relatively high socioeconomic status, selective school applicants as a whole tend to be significantly self-selected on the basis of

[2]Based on the constructed FOEI statewide standard deviation of 50 and the constructed (reversed) FOEI statewide mean of -100 (Rickard and Lu, 2014).



(a) Low cut-off

(b) Mid-low cut-off

(c) Mid-high cut-off

(d) High cut-off

Figure 5.5: Indigenous proportion of school enrolment in stacked sharp samples

socioeconomic advantage.

Moreover, the socioeconomic status of a student's government primary school is highly and positively correlated with the selective score ($\rho = 0.32$), which I illustrate in Figure 5.6) for the 2009 cohort. Consequently, of applicants coming from government primary schools, those offered a selective school place tend to come from higher socioeconomic status schools (with the average primary school reversed FOEI for such students being -36.62) compared to those who are not (with an average reversed FOEI of -60.95), a difference of around half a standard deviation.

As a result, unsurprisingly, fully selective schools are significantly more advantaged than the average government school on both available measures of socioeconomic status, FOEI and ICSEA (Table 5.2). In fully selective schools, 3.7% of students are in the bottom overall ICSEA quartile, compared with 24.3% for comprehensive high schools, representing a significant difference in the representation of the most disadvantaged students in the selective system. This pattern of higher average socioeconomic status is reflected in the discontinuous increases at the low and mid-low cut-off stacks, measured by ICSEA (Figure 5.7) and FOEI (Figure 5.8).

Student-level socioeconomic characteristics in the form of parental occupation and education variables are self-reported in the selective school application form, although reporting rates are relatively low. Socioeconomic characteristics tend to be more commonly reported for the first parent than for the second parent. Data on the first parent's occupation is missing for 41.56% of the sample, compared to 46.29% for the second parent. The respective missing rates for parent schooling are 38.26% versus 45.04%, and for qualifications, 42.68% versus 47.13%.

It appears that Parent 2 tends to be the primary income earner, given the large difference in non-employment rates between the two parents. Non-employment rates, however, are fairly consistent across the subgroups, with the biggest difference being between restricted applicants and applicants from non-government primary schools, indicating that families in the latter case

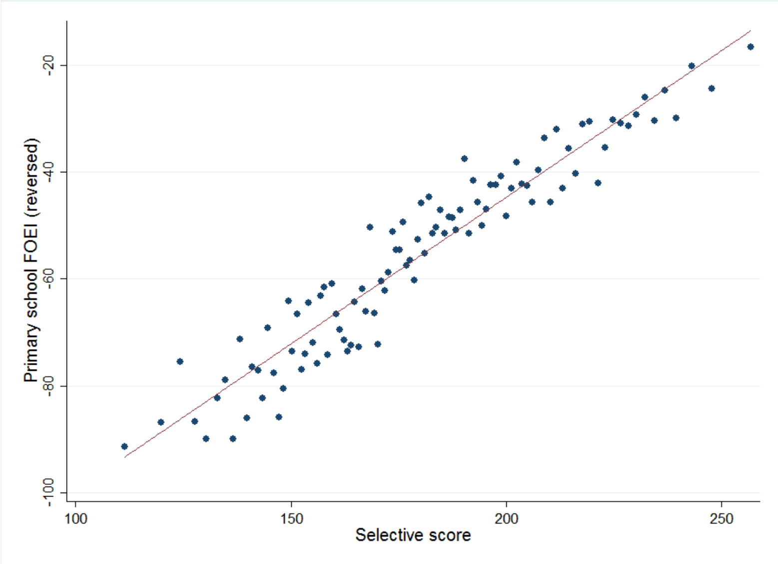
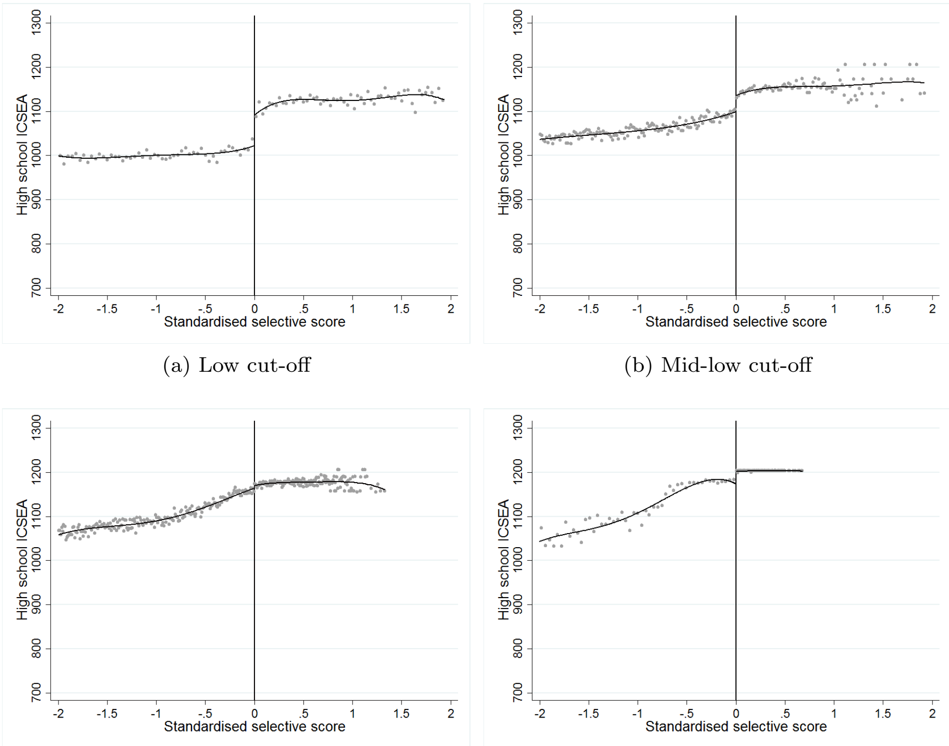


Figure 5.6: Selective school applicants' primary school FOEI and selective score for the 2009 cohort.



(a) Low cut-off

(b) Mid-low cut-off

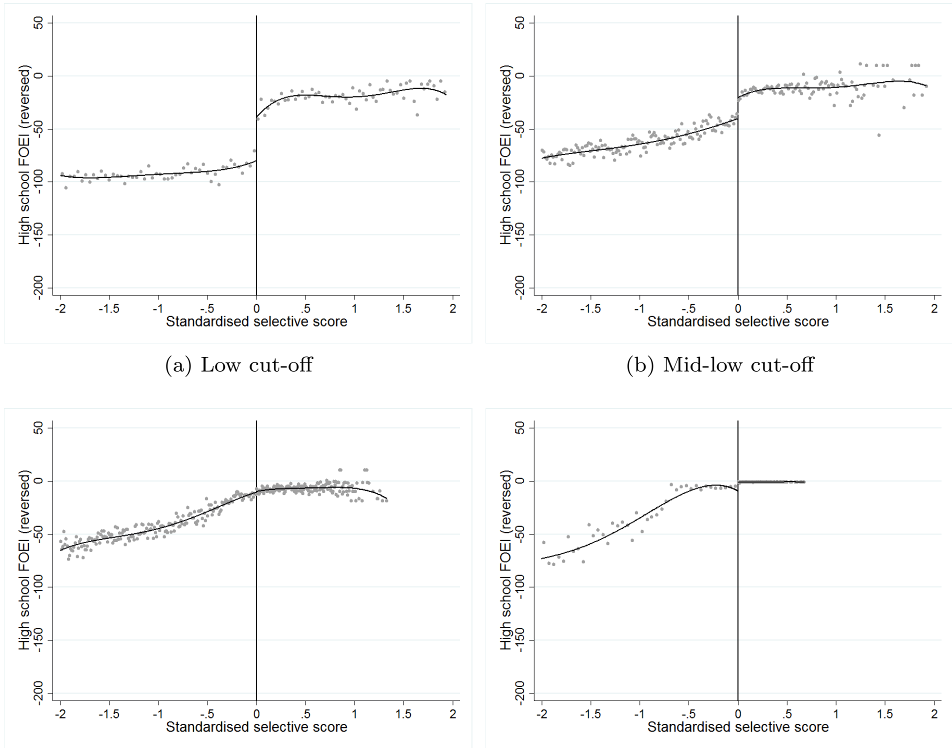
(c) Mid-high cut-off

(d) High cut-off

Figure 5.7: High school ICSEA measure of socioeconomic status in stacked sharp samples

are more likely to be dual-income (Table 5.1).

In Table 5.3, I more finely examine the distribution of parental characteristics for the reporting parents in the (restricted) sample. 60.6% of Parents 2 work in what could be termed professional or 'white-collar' occupations (occupation bands 4 and 5), which usually require a university degree. In comparison, 36% of all employed people aged 15 years or over in NSW in 2011 were in similar 'white-collar' occupations (Australian Bureau of Statistics, 2013). Approximately three-quarters of



(a) Low cut-off

(b) Mid-low cut-off

(c) Mid-high cut-off

(d) High cut-off

Figure 5.8: High school FOEI measure of socioeconomic status in stacked sharp samples

first and second parents who respond have completed Year 12 or equivalent, although the parents of students who are offered a selective school place are about 16 percentage points more likely to have done so. At least 50.63% of applicants have at least one parent who holds a Bachelor or higher-level degree. Again, there is a large disparity in Bachelor-qualified parents for students who are offered and not offered a selective school place: a roughly 15 and 20 percentage point differential for first and second parents respectively. In general, parental socioeconomic status, as measured by schooling, qualifications, and employment rates, is higher for those who are offered a selective school place compared to those who are not.

Table 5.3: Parental demographics definitions and distribution

	Band	Definition presented in survey	% of reporting Parents 1	% of reporting Parents 2
Parent occupation	1	Not in paid work in last 12 months	22.44	2.95
	2	Machine operators, hospitality staff, assstants, labourers and related workers	15.34	18.42
	3	Tradesmen/women, clerks and skilled office, sales and service staff	22.07	18.02
	4	Other business managers, arts/media/sportspersons and associate professionals	20.78	28.91
	5	Senior management in large busines organisation, government administration and defence, and qualified professionals	19.37	31.69
Parent schooling	1	Year 9 or equivalent or below	5.28	4.82
	2	Year 10 or equivalent	14.75	15.15
	3	Year 11 or equivalent	4.95	4.43
	4	Year 12 or equivalent	75.01	75.61
Parent qualifications	4	No non-school qualification	14.88	11.49
	5	Certificate I to IV (including trade certificate)	20.56	21.39
	6	Advanced diploma/Diploma	20.33	16.49
	7	Bachelor degree or above	44.23	50.63

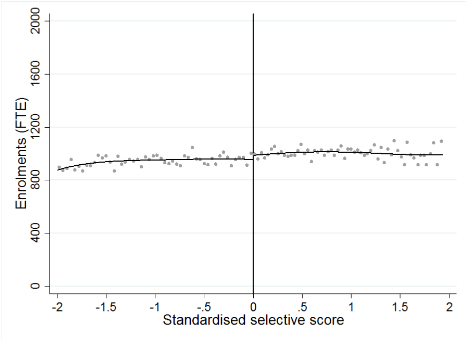
5.1.3 Non-demographic aggregate school characteristics

There is relatively little evidence of systematic discontinuities in school sizes (enrolments) across cut-offs (Figure 5.9), likely indicating that school sizes are determined by other factors across schools (e.g. for historical or geographical reasons) and relatively uncorrelated to selectivity.

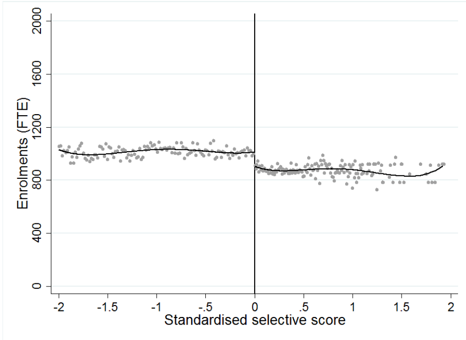
School income per student exhibits no general pattern across the stacked sharp samples, except for possibly the high cut-off stack (i.e., School R), which is a general reflection of the mechanical link between school demographics and the objectively determined government funding scheme (Figure 5.10). The higher funding at School R might reflect higher voluntary parent contributions at this school. In addition discontinuities in high cut-off stack RD plots may simply be reflective of the singleton defnition of the high cut-off stack.

Perhaps counterintuitively, student-teacher ratios tend to actually increase across some stack cut-offs (Figure 5.11), although the differences are relatively modest. In general, the student-teach ratios at schools attended by applicants appear to be relatively constant.

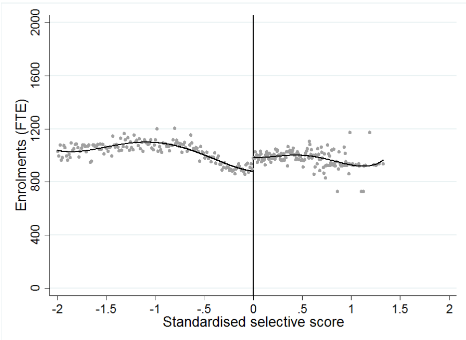
School attendance rates increase discontinuously at cut-offs, lending some credence to the belief that selective school students benefit from less-disruptive peers (Figure 5.12). However, while this pattern may suggest a link between 'bad apples' and attendance, there may be other student- or school-specific reasons for lower attendance at comprehensive schools that are unrelated to disruptive behaviour.



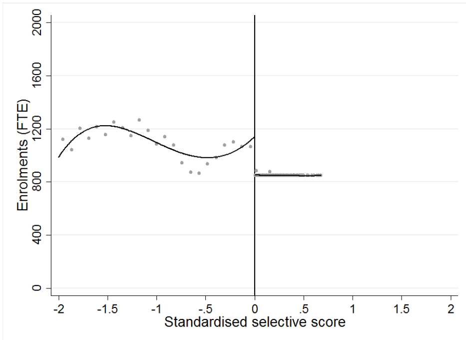
(a) Low cut-off



(b) Mid-low cut-off

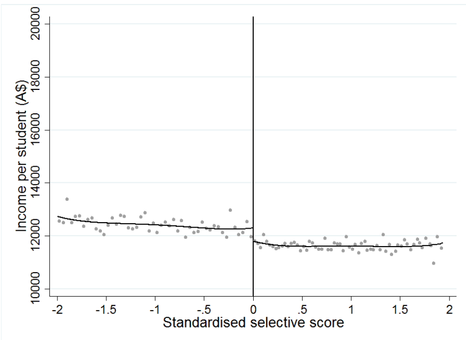


(c) Mid-high cut-off

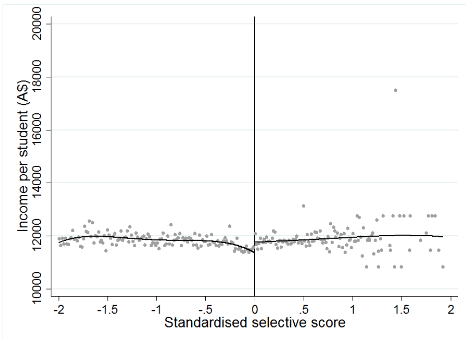


(d) High cut-off

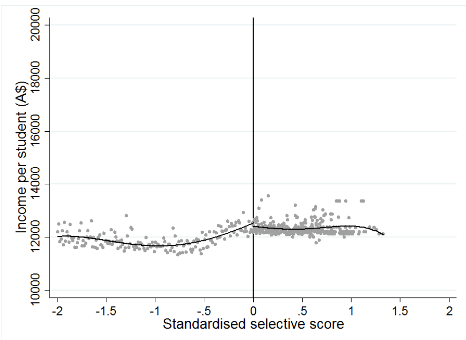
Figure 5.9: Total school enrolment in stacked sharp samples



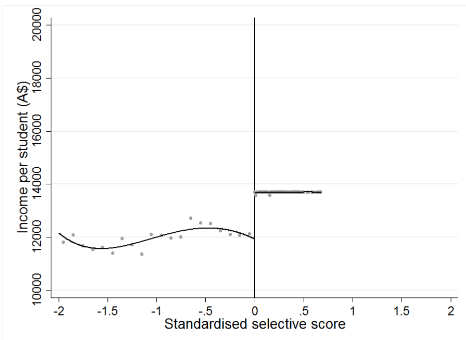
(a) Low cut-off



(b) Mid-low cut-off

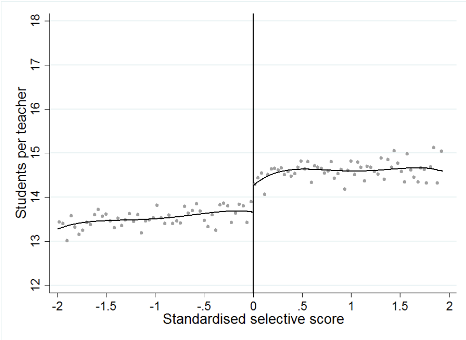


(c) Mid-high cut-off

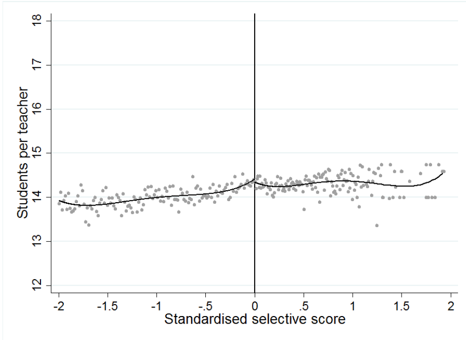


(d) High cut-off

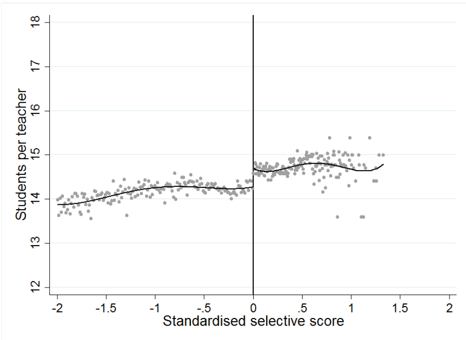
Figure 5.10: School income per student in stacked sharp samples



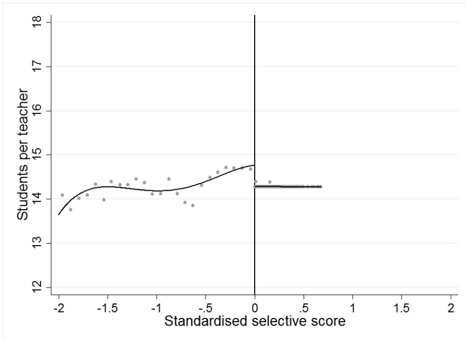
(a) Low cut-off



(b) Mid-low cut-off

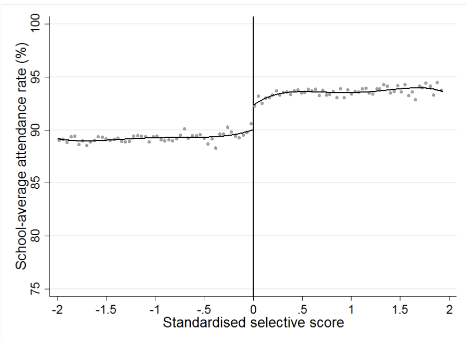


(c) Mid-high cut-off

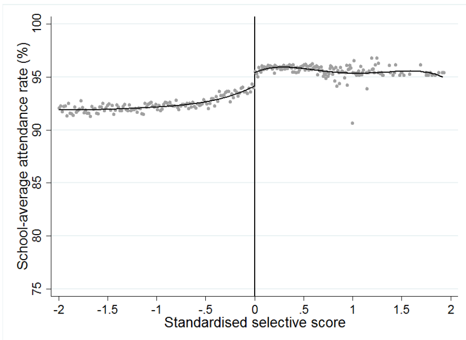


(d) High cut-off

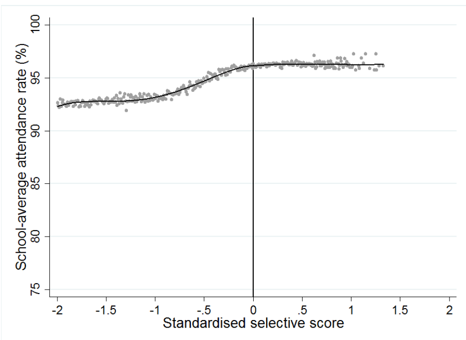
Figure 5.11: Student-teacher ratios in stacked sharp samples



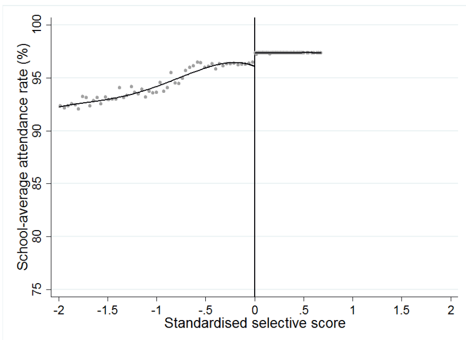
(a) Low cut-off



(b) Mid-low cut-off



(c) Mid-high cut-off



(d) High cut-off

Figure 5.12: School attendance rates in stacked sharp samples

5.2 Baseline academic achievement (NAPLAN scores)

Table 5.4: Average NAPLAN achievement comparisons.

	NSW average across all students and relevant cohort years	All applicants in restricted sample	Not offered any selective school place	Offered a fully selective school place
Year 5 NAPLAN Reading score	494.6 (74.9)	556.46	530.39	609.95
Difference to NSW average in standard deviations		.83	.48	1.54
Year 5 NAPLAN Numeracy score	487.8 (72.4)	573.22	534.94	652.32
Difference to NSW average in standard deviations		1.18	.65	2.27
Year 7 NAPLAN Reading score	544.5 (70.0)	607.75	580.82	657.08
Difference to NSW average in standard deviations		.9	.52	1.61
Year 7 NAPLAN Numeracy score	550.2 (77.6)	646.40	603.75	722.34
Difference to NSW average in standard deviations		1.24	.69	2.22

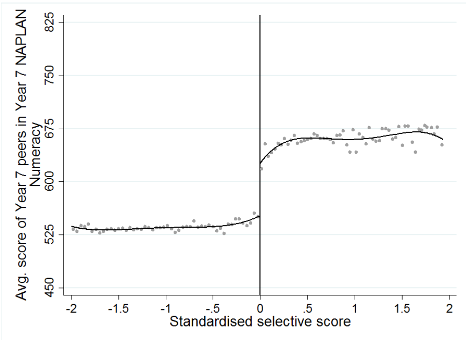
Brackets denote state-wide standard deviations averaged over relevant cohort years. NSW cohort averages of NAPLAN means and standard deviations are author's calculations based on Australian Curriculum Assessment and Reporting Authority (2008; 2009; 2010). Note that NAPLAN means and standard deviations over time are fairly consistent.

Table 5.4 compares average baseline NAPLAN scores for the state and the sample. Selective school applicants are significantly higher-achieving than the average NSW student, with average baseline differentials of 0.90 and 1.24 state-wide averaged standard deviations for Year 7 Reading and Numeracy, respectively. Selective school applicants thus represent an academically self-selected student population. This self-selection is common to other similar contexts where application is voluntary, for example, Abdulkadiroğlu et al. (2014) find that applicants to Boston exam schools achieved more than 0.7 standard deviations higher on baseline (pre-treatment) mathematics scores. As expected, those who are offered a fully selective school place score even higher, with such students scoring 1.61 and 2.22 state-wide averaged standard deviations higher on Year 7 NAPLAN Reading and Numeracy tests respectively.

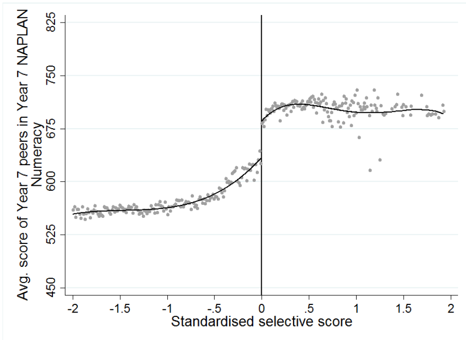
Furthermore, NAPLAN scores are comparable across years — a score represents the same level of achievement regardless of the testing year. The state-wide average Year 9 NAPLAN Reading score across 2010, 2011, and 2012 was 576, with an average standard deviation of 66.4; the average Numeracy score was 584.23 with an average standard deviation of 71.6. The Year 7 NAPLAN scores achieved by students who are offered a selective school place thus translate to a level of achievement that is significantly higher than the average Year 9 student: these students perform at more than 2 year levels ahead of the average student.

The achievement differentials found above are robust when comparing Year 5 NAPLAN scores for the 2009 cohort (tests taken in 2008), which are pure pre-treatment academic measures. The consistency in achievement differentials between selective school applicants (and those who are successful) and all NSW students in both Year 5 and 7 NAPLAN scores suggests that using Year 7 NAPLAN (which is available for all cohorts) as my measure of baseline achievement is justified, despite technically occurring during treatment.

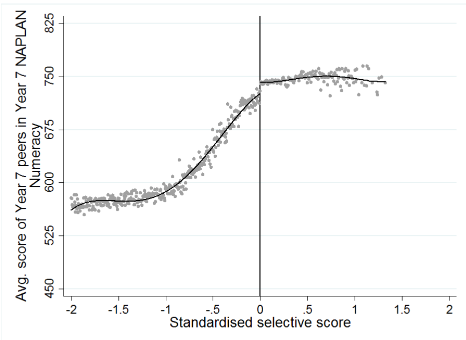
Figures 5.13 and 5.14 show the discontinuous differentials in average peer achievement across the stacked sharp samples. The differences in average peer achievement is most dramatic at the low cut-off selective schools, where average peer performance in both Year 7 NAPLAN Numeracy and Reading increases discontinuously by approximately 100 points at the cut-off (differences of approximately 1.3 and 1.4 state-wide averaged standard deviations respectively). We also see significant discontinuities in peer achievement at the mid-low and high cut-off stacks. In the latter case, there is an increase of 0.4 and 0.9 standard deviations in average peer performance in Numeracy and Reading respectively. This effect is less pronounced for the mid-high cut-off stacks, as the counterfactual for students in these school sharp samples is often a less-selective school which nonetheless exhibit similarly high levels of peer achievement.



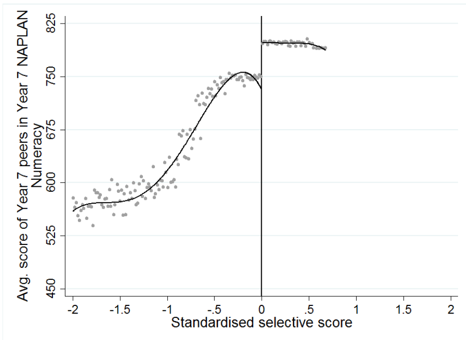
(a) Low cut-off



(b) Mid-low cut-off

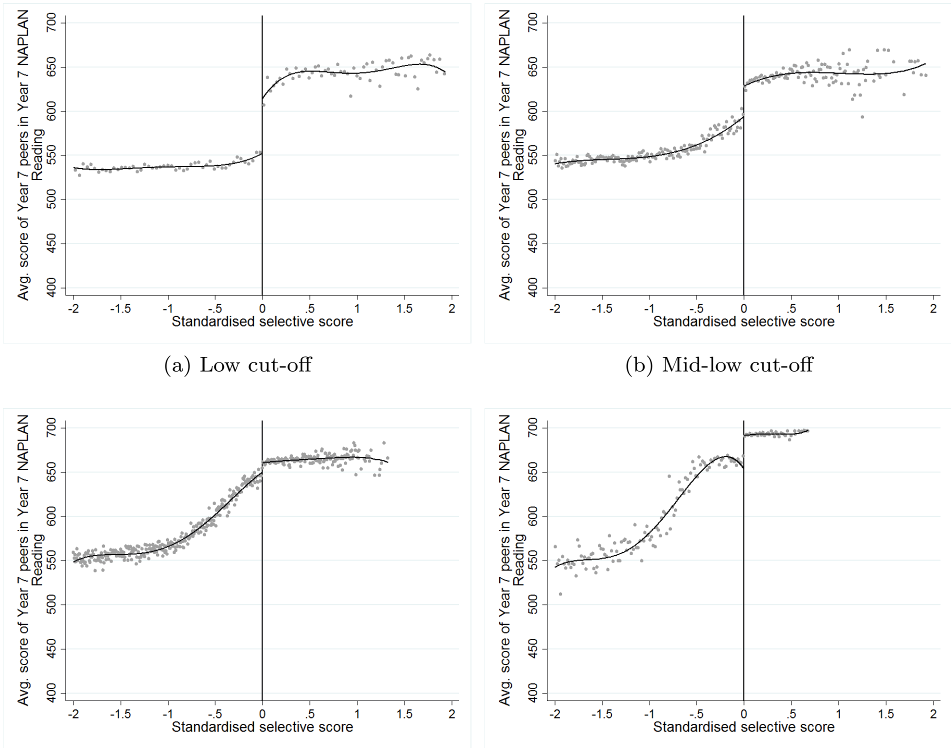


(c) Mid-high cut-off



(d) High cut-off

Figure 5.13: Average peer achievement in Numeracy (NAPLAN Y7) in stacked sharp samples



(a) Low cut-off

(b) Mid-low cut-off

(c) Mid-high cut-off

(d) High cut-off

Figure 5.14: Average peer achievement in Reading (NAPLAN Y7) in stacked sharp samples

5.3 Student outcomes

Table 5.5 summarises sample means for the outcome variables: HSC achievement and course participation. Students who are offered a selective school place score a TES of 388.88 on average, compared to 297.44 for those who are not offered a place. Using a restricted sample TES standard deviation of 88.64, this represents a difference in achievement of about one standard deviation. As expected, the same pattern of higher achievement for students offered a selective school place exists across all the subjects in Table 5.5.

Students who are offered a selective school place are also more likely to take the higher level 2 Unit mathematics and English courses and their extension courses. Anecdotally, many selective schools tend not to offer the lower-level 2 unit courses (English Standard and Mathematics General) as alternate options to the higher-level 2 unit courses (English Advanced and Mathematics (2 unit)).[3] The increase in subject participation is most pronounced for English Advanced, which is taken by 96% of students who are offered a place, as compared to 61% for students who are not offered a place. In contrast, the increase of 9 percentage points in Mathematics (2 unit) participation seems *prima facie* less dramatic because, unlike English, the Mathematics Extension 2 course is a possible substitute for Mathematics (2 unit). The 39 and 26 percentage points increase in the likelihood of participating in Mathematics Extension 1 and 2, respectively, indicates that students offered a place are significantly more likely to take advanced mathematics courses. There are less dramatic increases of 22 and 4 percentage points in participation rates in English Extension 1 and 2.

[3] As an example, North Sydney Girls High School, a fully selective school, does not offer either of the lower-level 2 unit courses (North Sydney Girls High School, 2016).

Table 5.5: Student-level means of outcome variables for various subgroups.

	All selective school applicants (unrestricted)	From non-government primary schools	All selective school applicants (restricted)	Students coded as not in government secondary schools	Not offered a fully selective school place	Offered a fully selective school place	Accepted a fully selective school place
Total observations wth a TES	23035	1677	19388	36	13244	6144	5829
HSC achievement							
Tertiary entrance score	329.39	345.95	326.42		297.44	388.88	391.09
English Standard score	35.07	35.13	35.53		35.47	36.70	36.90
English Advanced score	41.67	42.12	41.64		40.59	43.11	43.18
English Extension 1 score	42.58	43.02	42.54		40.98	43.64	43.68
English Extension 2 score	40.06	40.78	40.02		38.72	41.59	41.50
Mathematics General score	36.77	37.43	37.45		37.01	41.79	42.00
Mathematics (2 unit) score	41.07	41.84	40.95		39.47	43.61	43.77
Mathematics Extension 1 score	42.50	42.70	42.39		40.17	44.26	44.35
Mathematics Extension 2 score	42.99	43.05	42.94		40.42	44.19	44.26
Biology score	39.13	39.52	39.16		38.13	42.34	42.44
Chemistry score	40.55	40.99	40.42		38.67	42.73	42.83
Physics score	39.26	40.02	39.13		37.17	41.79	41.89
Earth and Environmental Science score	39.77		40.11		39.80		
HSC subject participation rates							
English Standard participation	0.276	0.207	0.278		0.386	0.045	0.038
English Advanced participation	0.723	0.793	0.721		0.613	0.955	0.962
English Extension 1 participation	0.174	0.212	0.173		0.104	0.320	0.327
English Extension 2 participation	0.054	0.077	0.054		0.043	0.078	0.077
Mathematics General participation	0.261	0.211	0.266		0.361	0.062	0.056
Mathematics (2 unit) participation	0.481	0.502	0.488		0.459	0.552	0.551
Mathematics Extension 1 participation	0.380	0.436	0.373		0.249	0.640	0.654
Mathematics Extension 2 participation	0.166	0.199	0.162		0.078	0.342	0.352
Biology participation	0.309	0.302	0.313		0.346	0.243	0.241
Chemistry participation	0.367	0.417	0.362		0.302	0.493	0.499
Physics participation	0.308	0.333	0.307		0.259	0.413	0.416
Earth and Environmental Science participation	0.018	0.014	0.018		0.023	0.005	0.005
Participation in any science(s)	0.651	0.669	0.654		0.619	0.731	0.735

In the case of individual course scores, the number of observations available depends on the number of students who participated in that course. Means calculated with fewer than 100 observations are excluded.

73% of students who are offered a place take at least one science course, compared to 62% of all students in the sample. Students who are offered a selective school place are 11 percentage points less likely to take Biology, but 19 and 15 percentage points more likely to take Chemistry and Physics, respectively, compared to students who are not offered a place. The higher participation rates in Chemistry are of interest because Chemistry is usually commonly assumed knowledge for science and engineering university degrees, whereas other sciences tend not to be. Participation rates in Earth and Environment Science are close to zero across all subgroups.

Chapter 6
Econometric framework

6.1 Regression discontinuity designs

Students who are offered a selective school place perform almost one standard deviation higher on the HSC on average than students who are not offered a place. However, it is possible that any achievement differential between selective and non-selective school students is due to differences in baseline student characteristics, including unobservables such as (baseline) motivation, rather than exposure to the selective school environment (e.g., higher-achieving peers).

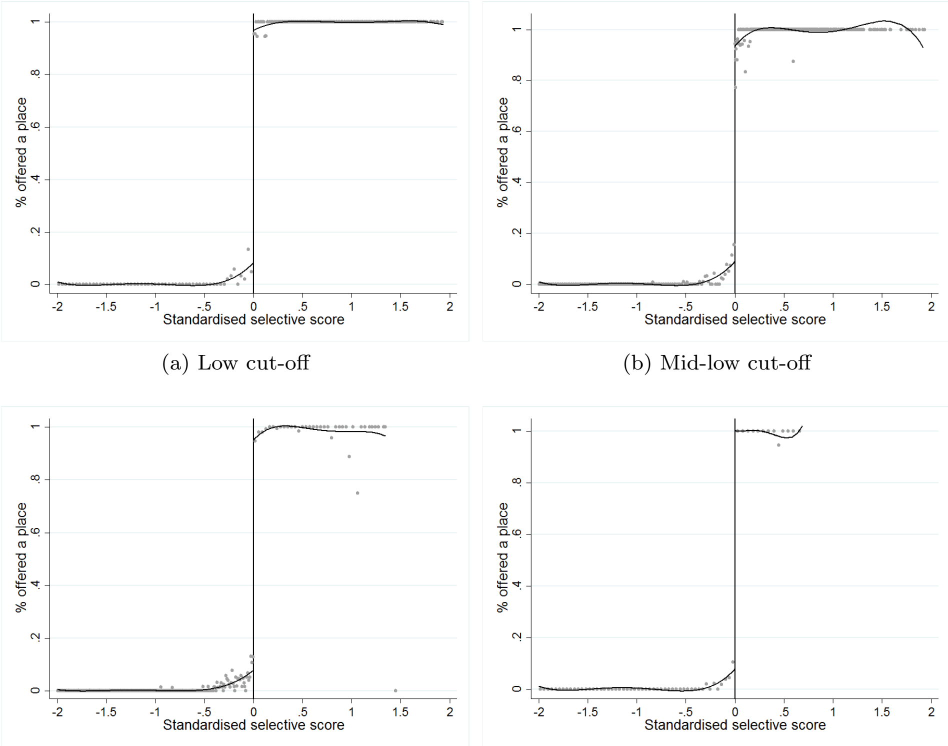
To circumvent this selection problem, I exploit regression discontinuity (RD) designs generated by school capacity constraints and the selective school assignment mechanism. In an RD design, cut-offs in a 'forcing variable' (in this case, the selective score) determine the probability of treatment assignment (an offer to attend a selective school) (Lee and Lemieux, 2010). The intuition behind this identification strategy is that selective school offers are as good as randomly assigned for students scoring close to a given cut-off because of noise in the selective score. Using RD designs, I estimate causal effects of offers to attend selective schools on student achievement based on this local randomisation assumption: students in a narrow bandwidth on either side of the cut-off are similar in characteristics — and most crucially, unobserved characteristics — other than being offered a selective school place. The RD designs I exploit in this paper are mostly of the 'fuzzy' type, meaning that the forcing variable does not perfectly determine treatment assignment, as there exists a small proportion of students below the cut-off receiving an offer, and vice versa. Fuzzy RD designs lead to an instrumental variables set-up in which the discontinuity is the instrument for treatment status (Lee and Lemieux, 2010).

In all regressions involving stacked sharp samples, and as with the regression discontinuity plots, I drop all observations whose standardised selective scores are zero following the recommendations by Fort et al. (2016), who demonstrate that not doing so might otherwise lead to spurious treatment effects in stacked samples with normalised forcing variables.

The main outcome variables of interest are students' post-treatment academic achievement: results in the overall Higher School Certificate (as measured by the TES) and their performance and participation in individual subjects. In order to address the possibility of heterogeneous treatment effects, I analyse treatment effects on the TES for subsamples by student characteristics. I also examine the treatment effect on mid-treatment outcomes in the form of Year 9 NAPLAN results, although as mentioned in Chapter 4, it is likely that these results are truncated at the top level.

Figure 6.1 motivates the use of regression discontinuity as the identification strategy, indicating the acute, but nonetheless imperfect sharpness of offers with respect to the standardised selective score across all stacks (and therefore all fully selective schools).

Figure 6.2 shows the relative fuzziness of acceptances (note that the absence of an offer is coded as a non-acceptance), indicating that applicants frequently exercise their choice of school. For applicants who receive an offer, 76.9% accept. Acceptance rates are lowest for students in the low and mid-low cut-off stacks, likely because the schools therein are more likely to be lower in students' preference lists.



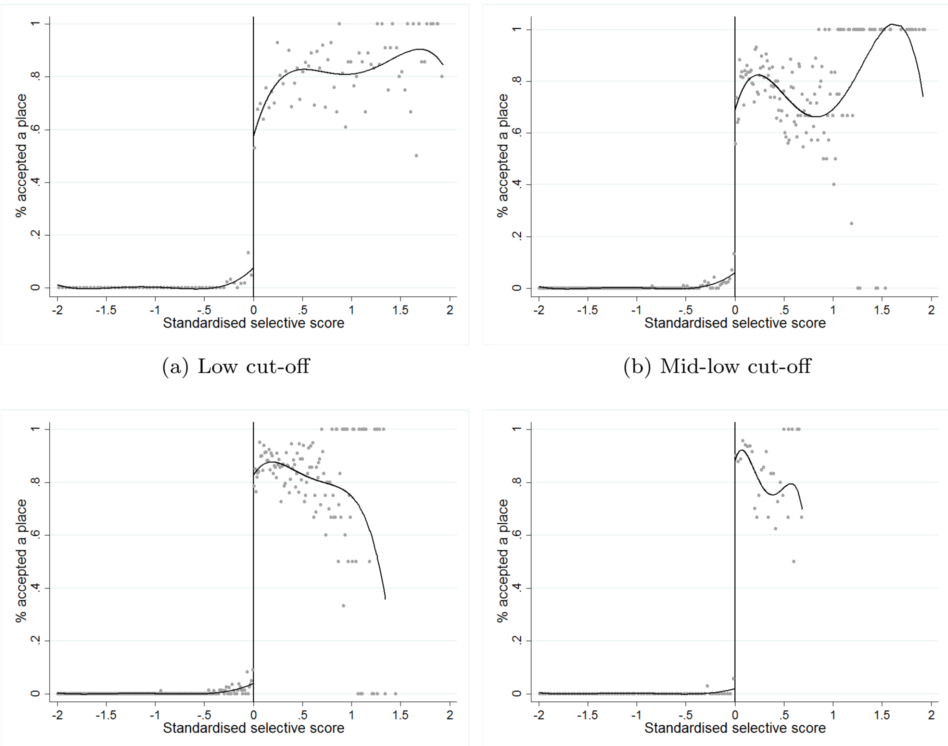
(a) Low cut-off

(b) Mid-low cut-off

(c) Mid-high cut-off

(d) High cut-off

Figure 6.1: Offer rates for schools within each stacked sharp sample



(a) Low cut-off

(b) Mid-low cut-off

(c) Mid-high cut-off

(d) High cut-off

Figure 6.2: Acceptance rates for schools within each stacked sharp sample

6.1.1 Non-parametric estimation

I implement a non-parametric strategy that has become the standard choice in RD designs (Calonico, Cattaneo, and Titiunik, 2014b), which limits samples to the observations close to the cut-offs, with the size of this window called the bandwidth (the total bandwidth is the sum of the bandwidths above and below the cut-off). The drawback of non-parametric estimation is that the bandwidth must be sufficiently large to encompass enough subjects to be informative, but sufficiently small such that the local randomisation assumption holds. Hahn, Todd, and van der Klaauw (2001) suggest simply assuming linearity in the bandwidth and running local linear regressions on either side of the cut-off. These regressions are often weighted by a selected kernel, such as a triangular kernel which assigns greater weight to observations closer to the cut-off, where the assumption of local randomisation is stronger. Lee and Lemieux (2010) favour the use of rectangular kernels for simplicity, and note that results are generally robust to kernel choice, but other practitioners tend to follow Imbens and Kalyanaraman (2012) in using triangular (Cohodes, 2015) and similarly downweighting kernels (Abdulkadiroğlu et al., 2014; Dustan et al., 2015). Following more modern studies, I use a triangular kernel.

I run a system of non-parametric 2SLS local linear regressions on each fully selective school and stacked sharp sample as follows:

$$\text{First stage:} \quad \text{Treatment}_{i,k} = \alpha_0 + \alpha_1 A_{i,k} + \alpha_2 Z_{i,k} + \alpha_3 A_{i,k} \cdot Z_{i,k} + \lambda' X_i + \epsilon_{i,k} \tag{6.1}$$

$$\text{Second stage:} \quad Y_i = \beta_0 + \beta_1 \widehat{\text{Treatment}}_{i,k} + \beta_2 Z_{i,k} + \beta_3 A_{i,k} \cdot Z_{i,k} + \theta' X_i + \eta_{i,k} \tag{6.2}$$

where $Z_{i,k}$ is the school-specific standardised selective score (from Equation 4.2, equivalent to the standardised distance to the cut-off), $A_{i,k}$ is a dummy variable taking a value of 1 if $Z_{i,k}$ is above the school-specific cut-off (due to selective score normalisation, all cut-offs are constructed to be 0), and Y_i is an outcome variable such as TES or a subject participation dummy. $A_{i,k}$ is the excluded instrument for $\mathit{Treatment}_{i,k}$. As usual, variables are indexed by school k and individual i. For school sharp samples, k is single-valued; for stacked sharp samples, k takes on a different value for each school in the stack. To improve precision, I also include a vector of pretreatment covariates X_i (gender, language background other than English or not, primary school FOEI, Year 7 NAPLAN Numeracy and Reading scores, and selective test year dummies) . The gender covariate is dropped from estimations using school sharp samples where that school is single sex (but not dropped for estimations using stacked sharp samples). $\mathit{Treatment}_{i,k}$ captures six alternate measures of the treatment, each of which are fuzzily related to the forcing variable:

(a) Selective school offer

(b) Selective school acceptance

(c) Standardised average achievement of Year 7 peers in Year 7 NAPLAN Reading

(d) Standardised average achievement of Year 7 peers in Year 7 NAPLAN Numeracy

(e) Standardised average achievement of Year 7 peers in Year 5 NAPLAN Reading

(f) Standardised average achievement of Year 7 peers in Year 5 NAPLAN Numeracy

Estimates using treatment indicator (a) compare students receiving and not receiving an offer in a given sharp sample, and therefore fall under an intention-to-treat framework. These estimates are the main focus of this paper. Treatment indicator (b) compares students who accept an offer with students who either were did not receive an offer, or who did receive an offer but did not accept. Regressions using treatment indicator (b) estimate treatment effects on compliers in part, as in nearly all cases those who accept a selective school place do enrol in at least one year of that selective school, although students still have the ability to move schools in later years. Treatment indicators (c-f) isolate the causal peer effects on student outcomes. In all cases, the estimate of β_1 represents the usual causal local average treatment effect on the outcome variable.

Standard errors are clustered by primary school FOEI (which proxies primary school attended).[1] The duplication of observations in stacked sharp samples (where a student who is in the sharp sample for multiple schools in that stack can appear multiple times) provides some motivation to cluster additionally on individual students (as do Abdulkadiroğlu et al. (2014) for example). As is general practice when given the choice of clustering between multiple nested variables, I cluster on the higher-level variable, namely primary school FOEI (Cameron and Miller, 2015).

I select the X_i covariates based on economic logic and their degree of availability among the sample. I do not include parental background variables due to relatively low reporting rates (approximately 38-47%). Calonico, Cattaneo, Farrell, and Titiunik (2016) outline the implications of adding covariates to RD designs, and find that both point estimation and inference are improved with covariates under the assumption that there is no RD treatment effect at the cut-off for the additional covariates. I depict the continuity of covariates in Section 6.4.3, and the results of more formal tests are reported in Appendix A. In such tests, I find little evidence for violation of the covariate continuity assumption.

Non-parametric, bias-corrected estimation, robust RD standard error estimation, and optimal data-driven bandwidth selection in this paper follow the method described first by Calonico et al. (2014b) and later updated by Calonico, Cattaneo, Farrell, and Titiunik (2016), and as implemented in the Stata package *rdrobust*.[2] The common MSE-optimal bandwidth selector I use selects a single bandwidth for observations both above and below the cut-off by minimising the asymptotic mean-square error of the point estimator. These relatively new methods (see Cohodes (2015) for a recent implementation) allow me to select optimal bandwidths and correct for bias arising from large bandwidths, in a setting with pre-treatment covariates and two-sided imperfect compliance. Results are generally unaffected by using a uniform kernel or other bandwidth selectors, although some results are marginally more significant in some sharp samples in the latter case (see Section 7.7).

6.2 Interpreting effects across a selectivity spectrum with varying counterfactuals

The relatively large number of selective schools allows me to examine selective school effects for schools with varying degrees of selectivity and peer achievement, as well as stack samples to improve the precision of estimates. In this section, I explore the nature of treatment and control groups across the sharp samples, and I also note various caveats to the interpretation of the estimation results.

[1] Each FOEI is given to 5 decimal places, and so is likely to uniquely identify schools. 1,140 unique primary school FOEI values are observed in the unrestricted sample, compared to 1,617 government primary schools that existed in NSW in 2014 Centre for Eduation Statistics and Evaluation (2015). The difference represents primary schools with no applicants amongst their student body, which is likely the case for schools in regional areas of NSW with few or no selective schools.

[2] For more details on the implementation of *rdrobust*, see Calonico, Cattaneo, and Titiunik (2014a) and Calonico, Cattaneo, and Farrell (2016).

6.2.1 Interpreting stacked sharp sample effects

The stacked sharp sample construction process combines treatment effects for multiple selective schools, but each selective school has idiosyncrasies, relating to say, school culture, informal institutions, or the suburb they are located in. We must then be careful in recognising the fact that no single treatment is experienced by students above the cut-off in stacked sharp samples (nor one single counterfactual below it, as in any sharp sample). The existence of the four-school preference list and the large number of heterogeneous selective schools conspire to vary the types of treatment and counterfactual schools across the sharp samples, which affect the interpretation estimated treatment effects.

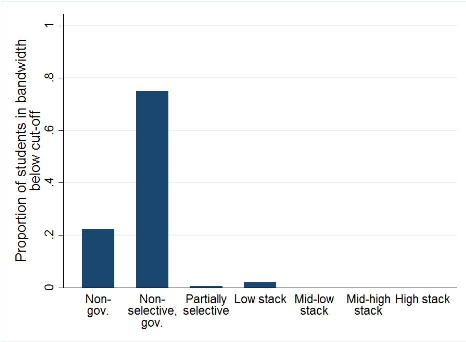
Figure 6.3 provides a more detailed indication of the types of schools principally attended by students within a bandwidth of 0.5 standard deviations below each stacked sharp sample cut-off, a window chosen as it encompasses most bandwidths selected in the fuzzy RD regressions. Panel 6.3 (a) shows again that the comprehensive high school counterfactual is most valid for the low cut-off stack, and that very few students apply to multiple low cut-off schools (and who meet the cut-off for one, but not the other(s)). The low cut-off stack treatment effect is then arguably of most interest policy-wise, as it most precisely estimates the effects of being offered a place at a selective schools on student achievement, as compared to compared to attending a government, non-selective school. Panel 6.3 (d) shows that the counterfactual school for the high cut-off stack is overwhelming a less-selective school. The treatment effect estimated for the high cut-off stack is then also relatively well defined, that is, the effect of attending an exceptionally selective school versus attending a less-selective school.

On the other hand, Panels 6.3 (b) and especially 6.3 (c) show that many students apply to multiple schools within the mid-low and mid-high cut-off schools, and that those who fail to reach a cut-off at a higher-preferenced school often gain entrance to a lower-preferenced, less-selective school. Panels 6.3 (c) and 6.3 (d) in particular indicate that the overwhelming counterfactual for mid-high and high cut-off selective schools is a less selective school. Treatment effects estimated for the mid-low and mid-high cut-off stacks are then a mix of the effect of being offered a selective school place, as compared to a government, non-selective school, as well as compared to a less-selective school.

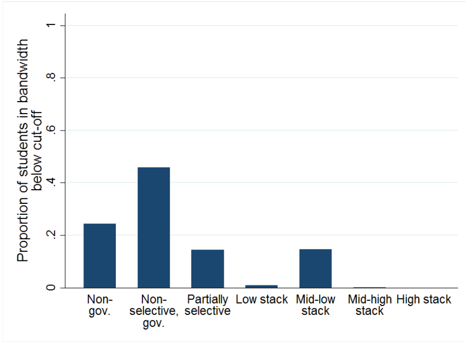
6.2.2 Interpreting school sharp sample effects

Understanding the relevant counterfactual is also key to correct interpretation of the individual school sharp samples. In addition, care must be taken when comparing treatment effects across the selectivity spectrum for the 18 individual selective schools. Strictly speaking, each school sample represents students who are offered and not offered a place at that school. While I have ordered schools by selectivity, this ordering should not be thought of as strictly implying a well-defined hierarchy, nor that this ordering allows us to cleanly interpret each successive school effect as effects incrementing over selectivity, or school or peer quality. In addition, the construction of the selectivity order by averaging calculated cut-offs over the three cohort years masks some, albeit minor, time variation in relative selectivity.

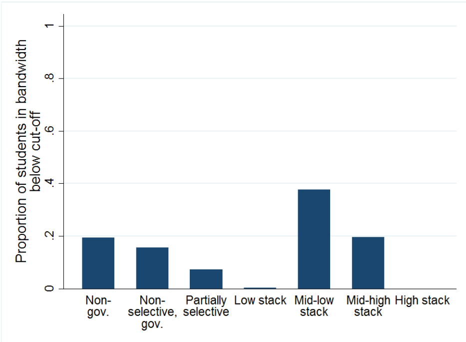
We might be able to reinforce the incremental peer achievement interpretation if we assume that students preference schools based on accurate perceptions of academic prowess. If they do so, then a student in any sharp sample faces the possibility of being offered a school with better performing peers than their alternative(s). However, although DOE publications on selective school cut-offs NSW Department of Education (2016) or unofficial rankings based on HSC achievement (such as by Ting and Bagshaw (2015)) might give some information to students, this information is clearly



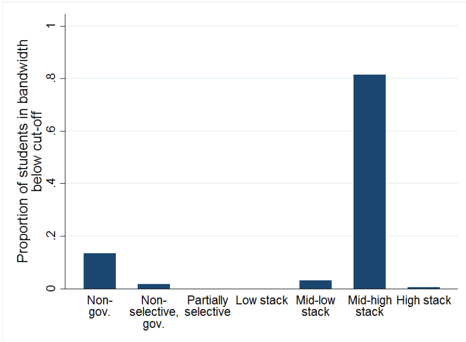
(a) Low cut-off



(b) Mid-low cut-off



(c) Mid-high cut-off



(d) High cut-off

Figure 6.3: Proportion of counterfactual school types principally attended by applicants below the cut-off and within an illustrative bandwidth in stacked sharp samples

These graphs are based on observations in each stacked sharp sample with standardised selective scores below the cut-off, and within an illustrative bandwidth of 0.5 (i.e., $-0.5 < Z_{i,k} < 0$). The 'partially selective' school type only includes students who accepted a place in the selective stream at a partially selective school. Students who attend the non-selective stream of a partially selective school are recorded as attending the 'public, non-selective' school type.

an imperfect measure of peer achievement. In addition, there are many other reasons for school preferences other than peer quality, such as location or social bonds, which somewhat reduces the validity of this argument.

Furthermore, the selectivity spectrum does not exactly match with baseline peer achievement levels, and so it is difficult to finely differentiate average peer quality between fully selective schools that are relatively similar in terms of selectivity. Figure 6.4 indicates as selectivity increases, there is a roughly incremental but non-monotonic increase in peer achievement in Numeracy, whereas Figure 6.5 indicates that peer achievement in Reading is much less correlated with selectivity. The relative closeness of average individual NAPLAN scores might also be indicative of selective school students 'topping out' the NAPLAN test as mentioned earlier. The ordering of selective schools is thus an imperfect indication of the relative levels of academic achievement, at least for schools that are relatively similar (e.g., between schools that belong to the same stack).

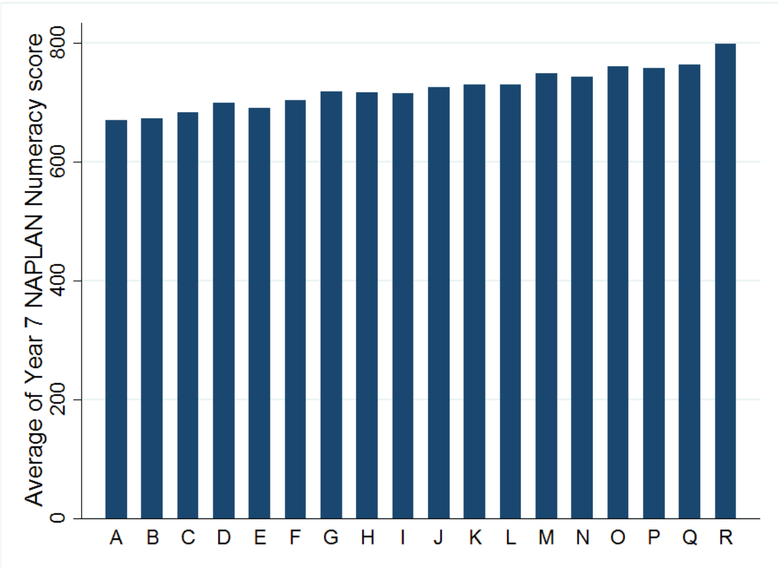


Figure 6.4: Average of individual Year 7 NAPLAN Numeracy scores by principal school attended

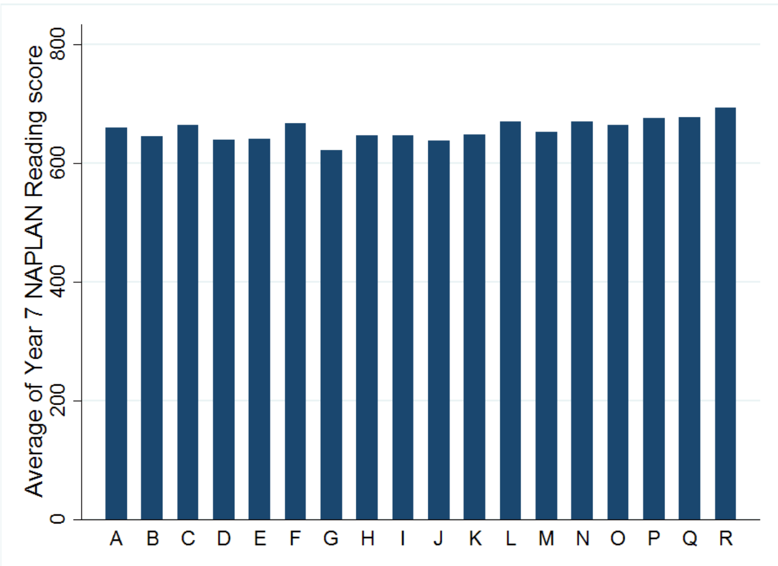


Figure 6.5: Average of individual Year 7 NAPLAN Reading scores by principal school attended

6.3 Estimating peer effects

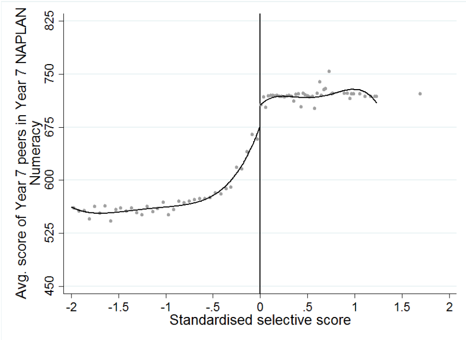
The RD designs using selective school offers or accepted offers as treatment indicators estimate overall treatment effects for selective schools or stacks, but does not identify the specific causal mechanisms by which these effects operate. However, as documented in Chapter 5, one of the clearest differences between selective schools and comprehensive schools is the level of achievement of their respective student populations. Following a similar method employed by Abdulkadiroğlu et al. (2014), I exploit RD designs for each school and stacked sharp sample where being above or below the cut-off is an instrument for the average ability of a student's Year 7 peers as measured by Year 5 and 7 NAPLAN test scores (excluding own score) to identify causal peer effects on academic achievement. The specifications, including covariates, kernel choice, and bandwidth selection process are identical to the regressions involving school offers, acceptances, or principally attended school as above.

By testing peer achievement effects in Reading and Numeracy (separately), I examine the possibility of domain-specific peer effect channels. For example, Arcidiacono, Foster, Goodpaster, and Kinsler (2012) find peer effects for university students that are largest for social science courses, and smallest for mathematics and science courses. However, any domain-specific peer effects I find are only suggestive without identification of the causal effect of both measures of peer quality (say, through additional valid instruments) in the same regression. I estimate peer effects from both Year 5 and Year 7 NAPLAN tests for robustness, although naturally the sample sizes for the regressions using Year 5 NAPLAN test scores are smaller since only one cohort possesses such data.

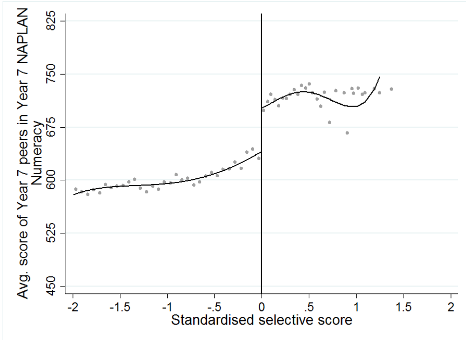
Unlike other treatment indicators, the 'average peer quality treatments' might not be strongly influenced by being above or below the cut-off, as many students on either side of some school or stacked sharp sample cut-offs face very similar levels of peer achievement. For example, peer achievement in Numeracy is relatively continuous at the cut-offs for Schools N and Q, as can be seen in Figure 6.6, Panels (c) and (d). Treatment effects estimated for these sharp samples are thus subject to the usual bias stemming from weak instruments.

Another potential source of bias in estimating peer effects is the extent to which the instruments (being above or below the cut-off) affect omitted education inputs in addition to peer quality. Specifically, if selective schools have better unmeasured inputs, the 2SLS peer effect estimates will also capture the effects of these inputs, and be upwardly biased. Abdulkadiroğlu et al. (2014) present two arguments as to the effect of this 2SLS omitted variable bias on their results. On one hand, since Boston/New York exam schools have better unmeasured inputs to academic achievement, such as richer course offerings, more modern facilities, and a challenging college-oriented curriculum, the results they find are simply positively biased. On the other hand, if such omitted inputs are a direct consequence of peer characteristics (in one of the authors' examples, exam school curricula might be challenging because exam school students are high-achieving), then the 2SLS estimates capture the total impact of randomly assigning peer groups, a similar argument to one made by Sacerdote (2011).

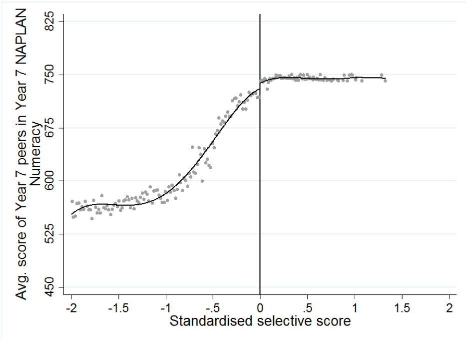
I argue that these peer effects are more readily identifiable in the NSW context than in the Boston and New York contexts of Abdulkadiroğlu et al. (2014), as there are relatively fewer differences in other school environment factors that could bias estimation. In contrast to U.S. school districts which appear to exhibit inequity of access to advanced high school coursework (Klugman, 2013), course menus and curricula in NSW tend to be much more consistent across the public school system. Differences in course menu composition are likely to arise more from the scale of specific advanced courses on offer, such as Mathematics Extension classes, due to relatively



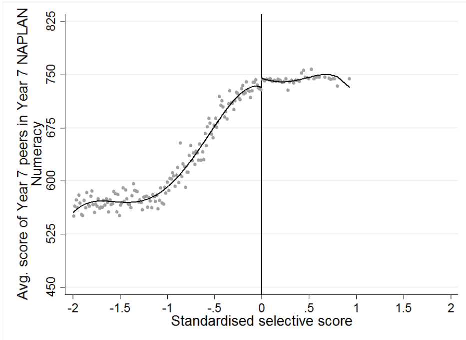
(a) Mid-low cut-off (School G)



(b) Mid-high cut-off (School J)



(c) Mid-high cut-off (School N)



(d) Mid-high cut-off (School Q)

Figure 6.6: Average peer achievement in Numeracy (NAPLAN Y7) in some school sharp samples in the mid-low and mid-high cut-off stacks

higher demand and supply in selective schools. Government funding levels, which make up the lion's share of school funding, are relatively consistent and formula-driven, compared to Boston and New York exam schools which can have sizable college-style endowments.[3] In addition, the wider availability of selective high schools, some of which are exceptionally selective and others which are less so, means that peer effects can be estimated over a relatively wider range of peer achievement.

6.4 Threats to validity

In general, the internal validity of an RD design depends on the following key assumptions (Lee and Lemieux, 2010):

1. Individuals are not able to precisely manipulate the forcing variable.

2. The distribution of observed baseline covariates do not change discontinuously at the threshold, and that there is no plausible argument that unobserved baseline covariates would do so.

Failure to meet these assumptions invalidates the comparison between the control groups (students below the cut-off) as counterfactuals to the treatment groups (students above the cut-off). In the following sections, I argue that the institutional environment is unlikely to give rise to precise manipulation, and provide formal test results and visual evidence of the validity of the RD designs.

[3]For example, one fundraising campaign by the Bronx High School of Science, a New York exam school, aimed to raise $20 million in private donations (Bronx Science Alumni Association, 2011).

6.4.1 The admission mechanism and forcing variable manipulation

In the admissions process, both sides of the selective school 'market' — the schools and the applicants[4] — potentially have incentives to precisely manipulate an applicant's selective score or a school's cut-off. I argue that this is unlikely to occur on either side within the institutional environment considered here.

Since applicants elect to apply for selective schools, they presumably prefer selective to comprehensive schools, and given that the selection mechanism elicits truthful preferences as argued earlier, they would logically prefer a higher selective score to a lower score.[5] Applicants thus clearly have an incentive to increase their selective score if given the opportunity to do so.

However, it is unlikely that applicants are able to precisely manipulate the forcing variable. The selective test and students' performance in Year 5 are noisy measures of ability and dependent on many factors outside the precise control of the student, for example, random environment conditions on test day. In addition, although cut-offs are published, applicants are unaware of the specific conversion mechanism of their performance inputs into the selective score. Applicants would also only be aware of the cut-offs for previous years, as contemporaneous year cut-offs are only published after the selection process, and so precise manipulation of scores close to cut-offs is implausible.

Additional admission factors such as illness/misadventure, disability, etc., should not affect manipulability if they are legitimate, since any adjustment is intended only to correct scores to the level that they would otherwise be. While applicants may potentially use these types of admission factors fraudulently, this still does not present a significant violation of internal validity, especially as such factors are still evaluated, and scores adjusted (if deemed necessary), by selection committees in manners generally unknown or unpredictable for applicants. In practice, such adjustments are usually made in a mechanical fashion. For example, if a student with less than 48 months of English-language education performs worse on the General Ability part of the selective test than is predicted by their other results, the average score of other students with the same English and mathematics scores is substituted.[6] As a result, even if additional admission factors were abused, this abuse is unlikely to constitute significantly precise forcing variable manipulation.

The remaining possibility for forcing variable manipulation by applicants involves manipulating additional admission factors *ex-post* by lodging an appeal after initial outcomes have been determined and communicated. Although applicants frequently raised appeals over the 2007-2009 test years, the appeals panels rarely upheld frivolous appeals which might represent attempts to manipulate the admissions process, e.g. test absence due to holidays, lack of familiarity with the process, etc.[7]

On the other side of the admissions process, selection committees cannot manipulate selective scores, except to adjust for additional admission factors where present as previously mentioned, e.g., certified illness on test day. While a computer generates the cut-off based on the projected supply of student places, selection committees may face incentives to manipulate the threshold through subjective admission. Such incentives might arise from, say, a desire to admit students with certain characteristics. To some extent, selection committees could theoretically alter cohort

[4] In this section, 'applicant' refers to the collective applicant-side stakeholders, including students, parents, and even theoretically the student's teachers/principal, since they submit school scores and special consideration requests in some cases.

[5] It is likely that this is the case as the highest possible preferenced school that the student qualifies for is offered and this is communicated by the DOE to families. This contrasts with other methods, say, the former Boston public school matching system where it is was 'safer' to prefer a lower-ranked school rather than a higher-ranked school even if applicant preferred the higher-ranked school (see Abdulkadiroğlu et al. (2005)).

[6] Private correspondence with staff of the High Performing Students Unit (NSW DOE), 2nd June 2016.

[7] Private correspondence with staff of the High Performing Students Unit (NSW DOE), 25th October 2016.

sizes, offer or not offer places to students on the 'wrong' side of the threshold on some non-objective or non-declared basis, or adjust scores based on additional admission factors beyond objectively reasonable compensation. As an example of admission manipulation, Urquiola and Verhoogen (2009) document manipulation by Chilean private schools of admission numbers in response to class-size caps to avoid the costs of adding an additional classroom.

In general, I argue that manipulation by selection committees is relatively unlikely, as they are composed of multiple mutually monitoring individuals with somewhat disparate incentives - such as a school director, the school's principal, a parent representative, and potentially other DOE staff members. In addition, as I only examine government schools, traditional economic motivations (e.g., the profit motive as in the Chilean case) to adjust the admission process are likely to be less of a concern. As a result, although there is an element of subjectivity in the admissions process at the margins, systematic corruption of the process is improbable.

In general, the assumption of random sorting (or equivalently the absence of discontinuities in non-forcing variables) is testable for observed variables in an RD framework. In order to check whether this assumption holds, I check the smoothness of observation density and covariates over school sharp sample cut-offs in the following sections.

6.4.2 Tests for discontinuities in the selective score density function

If there is non-random sorting of students at selective school cut-offs because of precise forcing variable manipulation, then we might expect disproportionately more students to appear on one side of the cut-off. This then allows the internal validity of RD designs to be at last partially assessed. Visual evidence for density discontinuities at cut-offs is relatively ambiguous, as shown in histograms in Figure 6.7. The high cut-off stack school appears to exhibit the greatest density discontinuity, followed by the mid-low cut-off stack. However, differences in sample size across stacked sharp samples, and choices of bandwidth and bin size influence the interpretation of these histograms. In order to better evaluate the significance of these potential discontinuities, I run a series of formal tests.

Cattaneo et al. (2016b) propose a test (the 'CJM' test henceforth) for forcing variable manipulation based on the continuity of the forcing variable density function, which builds on a similar test first proposed by McCrary (2008) (the 'McCrary' test henceforth). We would expect that if students can precisely manipulate scores, then more students will appear just above the cut-off and therefore receive treatment. If school(s) manipulate admissions, then the imbalance could be either way depending on the potential incentive. For example, they may wish to reduce the number of students attending to increase average cohort ability, in which case there would be disproportionately more students appearing below the cut-off. Disproportionately more students may appear above the cut off if, say, schools systematically 'game' the threshold by over-reporting their initial capacity in order to more finely screen potential students. Given that the overall density of selective scores is decreasing in the region containing selective school cut-offs (that is, the right-hand side of a roughly normal distribution), densities that are higher below the cut-off than above would not be overly surprising. I conjecture on this basis that densities that are discontinuously higher above the cut-off are more likely to be evidence of precise forcing variable manipulation.

The CJM test smooths out an empirical distribution function using local polynomial techniques, the advantage of which is a reduction in subjective selection of test parameters. In addition, the CJM test takes advantage of a MSE-optimal bandwidth selector with bias-corrected estimation of the type described by Calonico et al. (2014b), which is the same bandwidth selector used in the main models in this paper. In contrast, the McCrary test requires pre-binning of the data, or pre-estimation of the density near the cut-off, uses a simpler bandwidth selector (see McCrary (2008) for further details), and the bias arising from the size of the selected bandwidth is minimised by undersmoothing (i.e., arbitrarily shrinking the bandwidth) rather than bias-correction. Ultimately, the CJM test possesses improved size and power properties as compared to other tests, such as the McCrary test, and I mainly guide my analysis on the results thereof.

I run CJM tests using the Stata package *rddensity*[8] on both the unrestricted sample and the restricted sample available for estimation (i.e., excluding students with no TES outcome variable data, see Table 4.1). As usual, I drop observations with zero standardised selective scores in stacked sharp samples following Fort et al. (2016). I also provide McCrary tests for comparability and completeness' sake. I create sharp samples and standardised selective scores for the unrestricted sample as I do with the restricted sample to run these tests.[9] By testing the entire unrestricted sample, I evaluate the likelihood of non-random sorting through precise forcing variable manipulation by parties in the admission process, which would invalidate the assumptions necessary for the RD design. As I am only interested here in the potential presence of manipulation in the admissions process, I do not create stacked sharp samples for the unrestricted sample. By testing the restricted sample, I evaluate the concern of estimation bias arising from unbalanced treatment and control groups.

For completeness' sake, I also run both unrestricted and restricted versions of the CJM test (Cattaneo et al., 2016b). The unrestricted test separately estimates the distribution of observations on either side of the cut-off, and evaluates any discontinuity at the cut-off. The restricted test assumes that the cumulative distribution functions and higher-order derivatives are identical on either side of the cut-off, and then evaluates any discontinuity at the cut-off. The restricted test is more sensitive to discontinuities due to the stronger assumptions it makes, whereas the unrestricted test makes minimal assumptions.

The CJM and McCrary test results for each school and stacked sharp sample are summarised in Table 6.1. Informally, we would expect approximately one result in each column to be statistically significant by chance, given a 5% significance level. In the unrestricted sample, the unrestricted CJM test shows little evidence of forcing variable manipulation; the restricted CJM test indicating three schools with discontinuities that are significant at least at a 5% level. These results confirm the discussion above arguing that manipulation of the admissions process by students and/or schools is unlikely in this context. As a result, any discontinuities found in the restricted samples are likely to reflect the effect of differential attrition as opposed to forcing variable manipulation *per se*.

In the restricted sample, the unrestricted test shows one school, as well as the mid-high cut-off stack, with discontinuities that are significant at the 5% level; the restricted test shows significant discontinuities at eight schools, as well as at the mid-low and mid-high cut-off stacks. This increase in sharp samples which fail the test is relatively unsurprising as the sample restrictions imposed are unlikely to have symmetrical effects around the cut-offs — for example, private school primary students, who are excluded in this sample, are more likely to appear below school cut-offs than above (see Figure 6.13).

The McCrary test results generally detects more significant discontinuities than the unrestricted CJM tests. In particular, the McCrary test detects discontinuities for the mid-low and high, but not the mid-high cut-off stack. The differences in methodology, including the additional

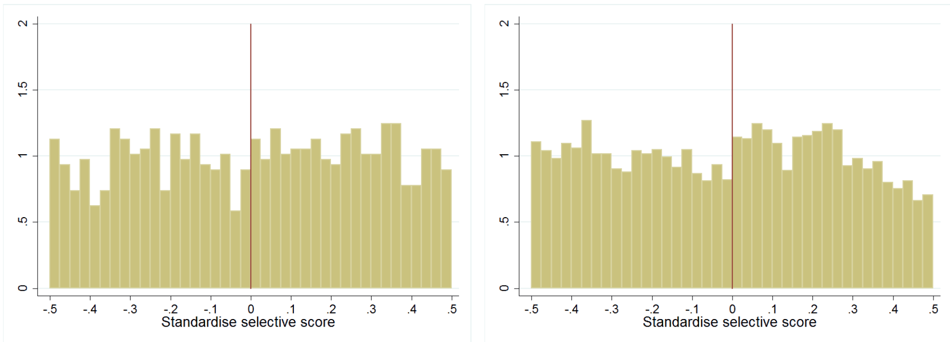
[8]For more details on the implementation of *rddensity*, see (Cattaneo, Jansson, and Ma, 2016a).

[9]Note that as the means and standard deviations of selective scores in each sharp sample differ between the restricted and unrestricted samples, standardised selective scores also differ.

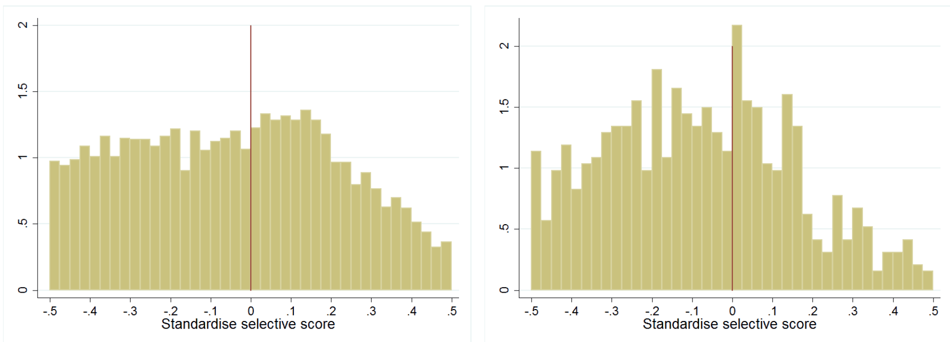
subjective assumptions on pre-binning and bandwidth selection described above likely lead to the discrepancies between the McCrary and CJM tests.

On the basis of the unrestricted CJM tests on the restricted samples, estimates for the mid-high cut-off stack and School K appear to be at most risk of invalidation due to violations in the RD design assumption of no precise manipulation. More schools, as well as the mid-low cut-off stack, might also be at risk on the basis of the restricted CJM tests and McCrary tests, although this is likely to be a result of the increased number of restrictive assumptions or parameter settings as aforementioned. The low and high cut-off stacks appear to satisfy most tests.

In Appendix B, I show the results of unrestricted CJM tests on each of the samples available for each outcome variable (Table B.1), as well as for subsamples (Table B.2). The results generally indicate that density discontinuity concerns are highest for the mid-high, and to a lesser extent, the mid-low cut-off stacks. As found above, the low and high cut-off stacks appear to satisfy the tests for density discontinuity.



(a) Low cut-off (b) Mid-low cut-off



(c) Mid-high cut-off (d) High cut-off

Figure 6.7: Histograms within an illustrative bandwidth in stacked sharp samples

Bin size is 0.025. Bandwidth used in each stacked sharp sample is $\gtrless 0.5$. Students with no TES are excluded.

Table 6.1: Forcing variable manipulation testing based on density discontinuity following Cattaneo et al. (2016b)

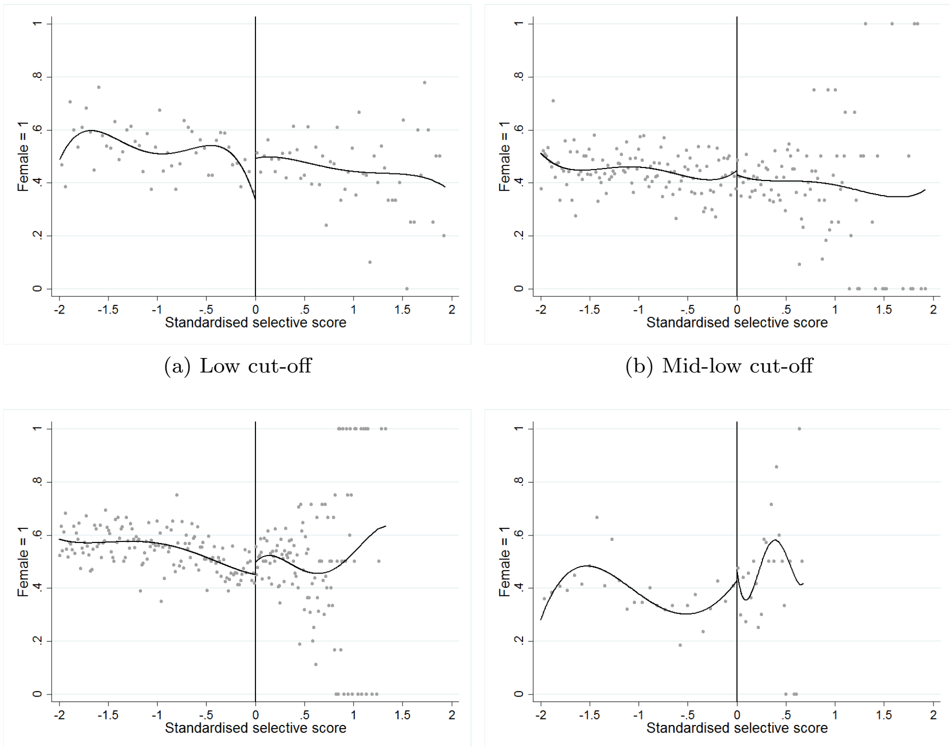
School/stack	Cattaneo et al. (2016b) tests				McCrary (2008) test	
	Unrestricted sample, unrestricted test	Restricted sample, unrestricted test	Unrestricted sample, restricted test	Restricted sample, restricted test	Unrestricted sample	Restricted sample
A	-.016	.085	.034	.161**	.109	.259
	(.088)	(.123)	(.057)	(.078)	(.202)	(.264)
B	.011	.013	-.006	.040	.181	.279
	(.052)	(.086)	(.045)	(.059)	(.168)	(.221)
C	-.039	.031	-.013	.058	.060	.356
	(.088)	(.110)	(.051)	(.069)	(.185)	(.259)
D	-.081	-.062	-.039	.063	-.043	.133
	(.068)	(.098)	(.037)	(.050)	(.154)	(.216)
E	.025	.170	.049	.180***	.214	.364*
	(.080)	(.115)	(.045)	(.058)	(.147)	(.186)
F	.060	.111	-.004	.106**	.510**	.705*
	(.074)	(.097)	(.038)	(.050)	(.236)	(.362)
G	.084	.152	.104**	.240***	.312*	.717***
	(.065)	(.100)	(.043)	(.060)	(.183)	(.249)
H	.074	.078	.067*	.149***	.302**	.441*
	(.061)	(.099)	(.040)	(.057)	(.150)	(.226)
I	.120*	.137	.028	.054	.297*	.357*
	(.063)	(.089)	(.038)	(.055)	(.161)	(.197)
J	-.031	-.137	.003	.134*	-.009	-.014
	(.086)	(.136)	(.045)	(.069)	(.172)	(.234)
K	-.006	.168**	.050	.150***	.057	.401**
	(.047)	(.067)	(.031)	(.056)	(.140)	(.170)
L	.023	-.101	.054	.193***	.090	.384*
	(.060)	(.118)	(.046)	(.065)	(.156)	(.220)
M	.043	.060	.105***	.117*	.278*	.227
	(.076)	(.129)	(.038)	(.067)	(.145)	(.170)
N	-.094	.058	.115***	.292***	.184	.638***
	(.066)	(.118)	(.035)	(.064)	(.153)	(.193)
O	.004	.020	.050	.071	-.021	.014
	(.067)	(.118)	(.042)	(.074)	(.144)	(.182)
P	-.012	-.045	.048	.055	-.124	-.014
	(.067)	(.110)	(.035)	(.057)	(.180)	(.221)
Q	-.081	-.134	.006	.023	-.005	.029
	(.070)	(.112)	(.034)	(.054)	(.154)	(.169)
R	.106	.311	.011	.078	.137	.485**
	(.117)	(.206)	(.072)	(.090)	(.158)	(.196)
Low		.019		.047		.287
		(.070)		(.044)		(.177)
Mid-low		.074*		.108***		.294***
		(.041)		(.025)		(.095)
Mid-high		.098***		.078***		.135
		(.035)		(.030)		(.082)
High		.311		.078		.485**
		(.206)		(.090)		(.196)

Standard errors in parentheses. Cattaneo et al. (2016b) estimates are differences in local, order 2 polynomial density estimators using a triangular kernel, in a bandwidth selected by a common MSE-optimal bandwdth selector. CJM standard errors are jackknifed. The unrestricted CJM test allows for distribution estimators with different parameters on either side of the cut-off, the restricted CJM test forces identical distributions. McCrary (2008) test statistics are log differences in height of the densities at the cut-off, estimated in a bandwidth selected by a common MSE-optimal bandwidth selector. The definitions of restricted and unrestricted samples are given in Table 4.1, where the restricted sample is that which is 'available for estimation' (i.e., excluding students with no TES). * significant at 10%, ** significant at 5%, *** significant at 1%.

6.4.3 Tests for continuities in baseline covariates

In this section, I show visual evidence for the continuity of baseline covariates included in the RD regressions: gender (Figure 6.8), language background other than English (Figure 6.9), Primary school FOEI (Figure 6.10), and baseline NAPLAN test scores (Figures 6.11 and 6.12. I test for discontinuities more formally in Appendix A by replacing the outcome variable in the main model regressions (Equations 6.1 and 6.2) with each covariate in turn. In such placebo tests, I include no other covariates.

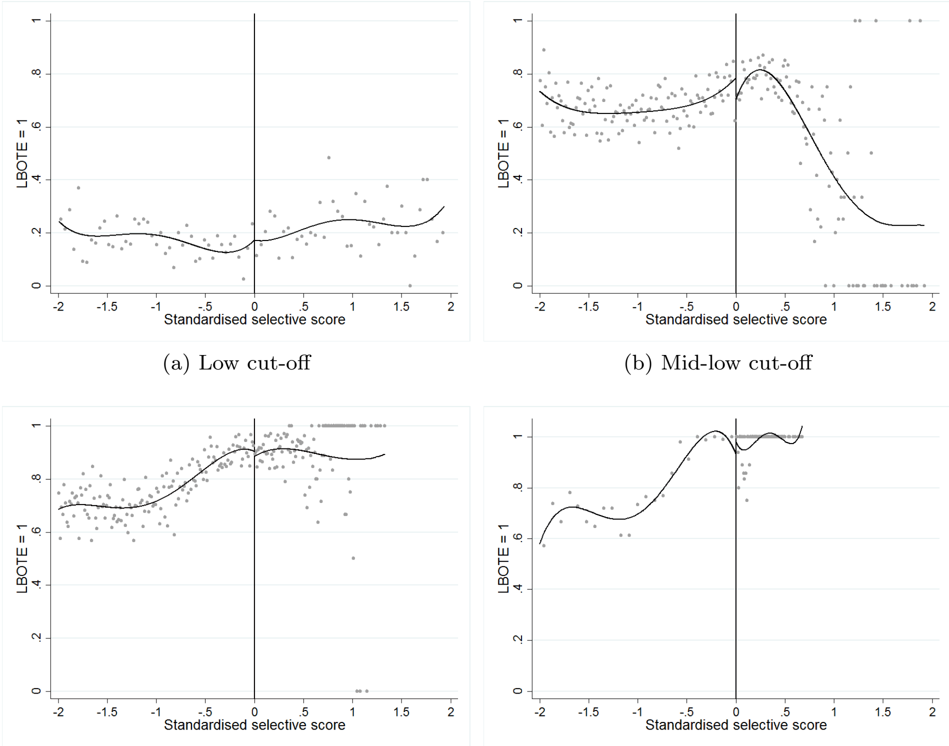
Overall, covariates appear continuous across cut-offs, with formal placebo test results producing placebo effects of only scattered statistical significance, broadly in line with the number of false positives expected at a 5% significance level. The pattern of continuous covariates across cut-offs lends credibility to the RD design, even in cases where we might suspect that there is self-sorting either before or after sample restriction. Continuity of the covariates also justifies their inclusion in the regression equations to improve estimation precision.



(a) Low cut-off (b) Mid-low cut-off

(c) Mid-high cut-off (d) High cut-off

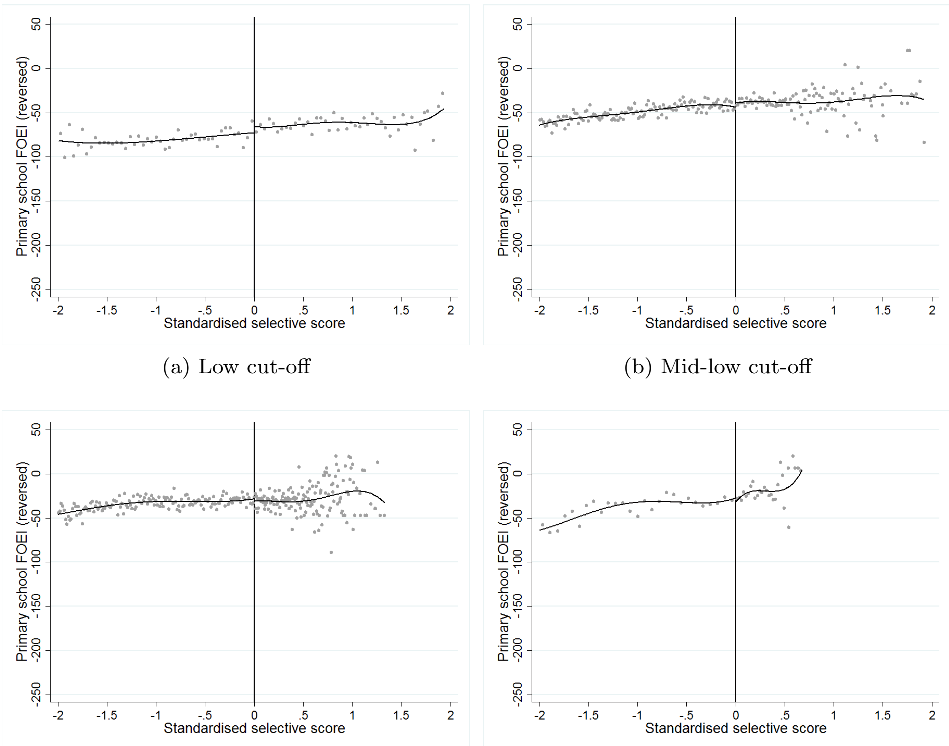
Figure 6.8: Proportion of students that are female in stacked sharp samples



(a) Low cut-off (b) Mid-low cut-off

(c) Mid-high cut-off (d) High cut-off

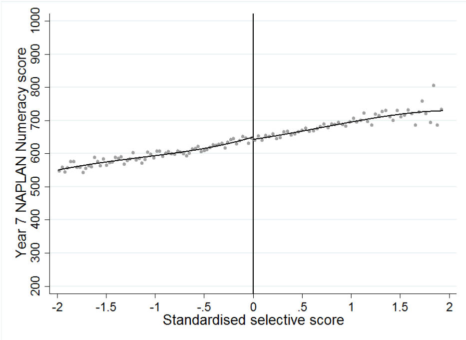
Figure 6.9: Proportion of students from language backgrounds other than English in stacked sharp samples



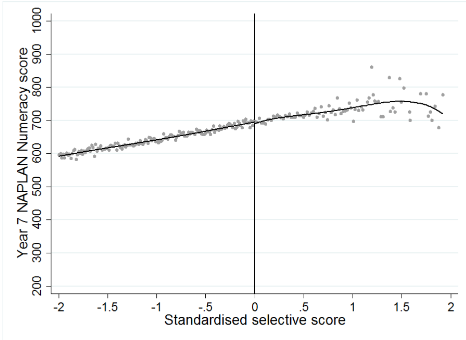
(a) Low cut-off (b) Mid-low cut-off

(c) Mid-high cut-off (d) High cut-off

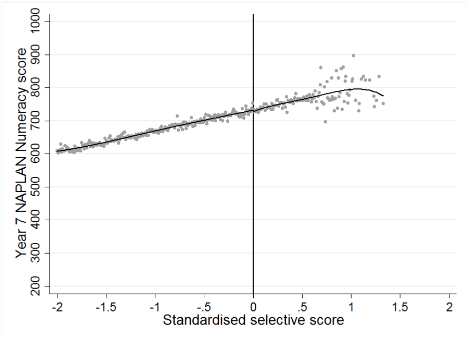
Figure 6.10: Primary school FOEI in stacked sharp samples



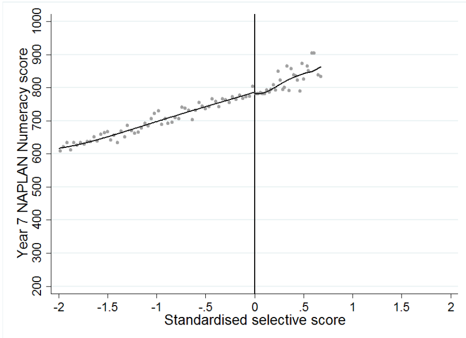
(a) Low cut-off



(b) Mid-low cut-off

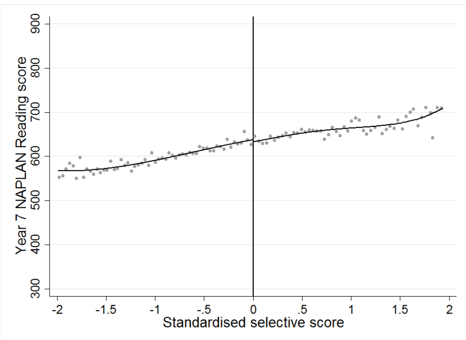


(c) Mid-high cut-off

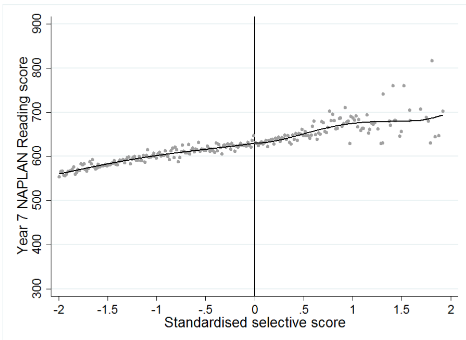


(d) High cut-off

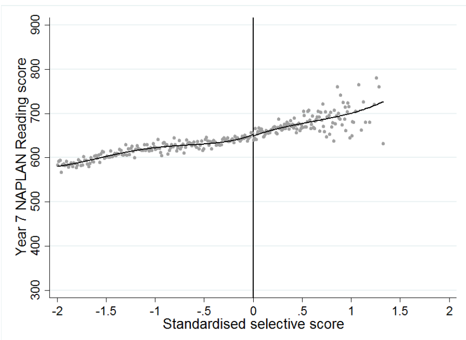
Figure 6.11: Individual baseline Numeracy (NAPLAN Y7) in stacked sharp samples



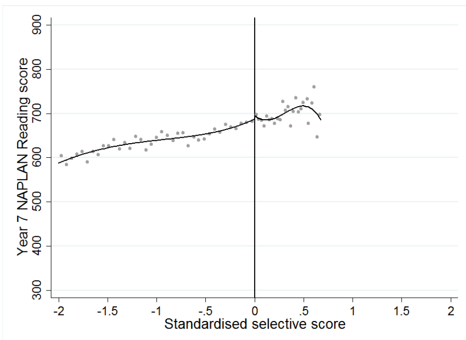
(a) Low cut-off



(b) Mid-low cut-off



(c) Mid-high cut-off



(d) High cut-off

Figure 6.12: Individual baseline Reading (NAPLAN Y7) in stacked sharp samples

6.4.4 Differential attrition

The absence of linked data for students who attend non-government high schools potentially creates a selection bias, as potential or actual non-government school students differ in characteristics from government school students. Since applicants are 16.4 percentage points more likely to exit the public school system if they were not offered a place in a fully selective school than if they were, this means that student characteristics in either treatment and non-treatment groups may be different (e.g. family income).

I exclude non-government primary school applicants to mitigate this problem as such students are more likely than government primary school applicants to later attend a non-government high school. I explore the characteristics of the 13.9% of applicants who apply from a primary school external to the NSW public school system, in order to evaluate the possible implications of dropping them from the sample. The proportion of applicants coming from non-government primary schools is considerably lower than in previously studied contexts such as Boston, where 45% of exam school applicants come from private schools (Abdulkadiroğlu et al., 2014), further showing the significant discrepancies across educational jurisdictions.

Non-government primary school students have slightly lower selective scores on average compared to all students, as documented in Table 5.1. Despite the fact that non-government primary schools are likely to be relatively advantaged, students who are offered a selective school place are slightly less likely to come from such schools (13.4%) compared to students who are not offered a place (14.2%). The probability of attending a private primary school is quadratically related to the selective score, such that both high- and low-scoring students tend to come from government schools (Figure 6.13)

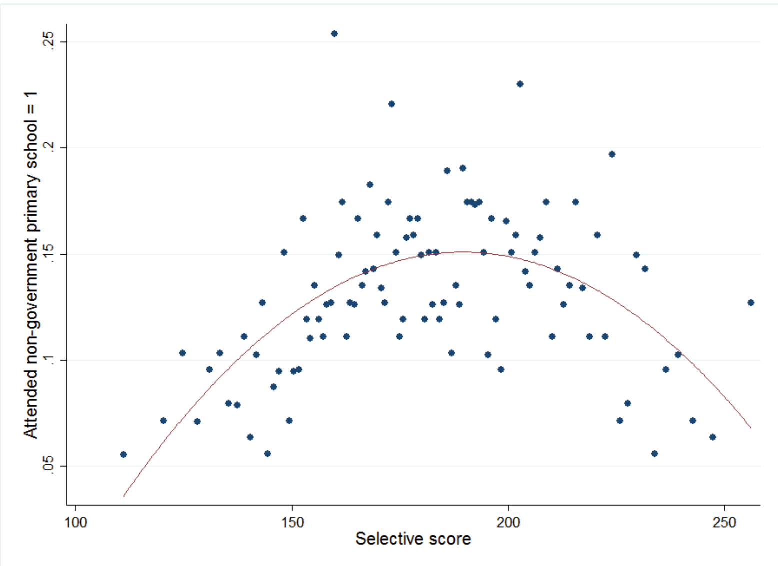


Figure 6.13: Proportion of applicant primary schools that are non-NSW government administered and selective score for the 2009 cohort

However, excluding non-government primary school applicants does not completely eliminate differential attrition bias. Table 6.2 shows rates of attrition out of the public school system rates in a bandwidth of 0.5 standard deviations around the cut-off. Attrition is determined by the principally attended school variable (i.e., if the school is a government or non-government school). The low and mid-low cut-off stacks exhibit the greatest rates of differential attrition. Attrition rates are higher below the cut-off than above in all sharp samples except at the most selective school (School R/high cut-off stack). The attrition rate is on average 10 percentage points higher below the school sharp sample cut-offs than above, indicating that students below the cut-offs are more likely to leave the public school system than those who do. This indicates that a sizable number of students enjoy and exercise the option to attend private schools, and that more students do so if they are unsuccessful in meeting the cut-off(s).

Table 6.2: Attrition rates within an illustrative bandwidth around sharp sample cut-offs

School or stacked sharp sample	Rates of attrition out of public school system	
	Below cut-off	Above cut-off
A	0.28	0.13
B	0.22	0.10
C	0.19	0.08
D	0.24	0.08
E	0.25	0.10
F	0.38	0.26
G	0.24	0.11
J	0.25	0.16
H	0.22	0.11
I	0.17	0.14
K	0.21	0.08
L	0.29	0.13
M	0.16	0.10
O	0.24	0.19
N	0.16	0.06
P	0.25	0.13
Q	0.09	0.07
R	0.13	0.16
Low	0.23	0.10
Mid-low	0.24	0.14
Mid-high	0.19	0.11
High	0.13	0.16

Bandwidth used in each sharp sample is $\gtrless 0.5$ standardised selective scores.

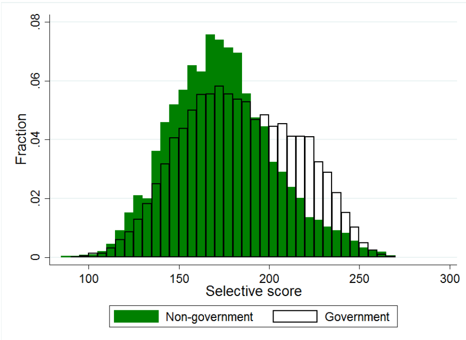
By the local randomisation assumption, students close to the cut-off have similar underlying ability, and so any bias arising from differential attrition operates through differences in other inputs to educational achievement. To understand the likely magnitude and direction of the bias arising from differential attrition, I examine the differences in baseline variables between applicants who principally attend a government school and those who do not.

I complement the descriptive statistics in Table 5.1 with histograms to better examine distributional differences. Government high school students tend to achieve higher in baseline Year 5 NAPLAN test scores[10] and higher selective scores (Figure 6.14). As broadly characterised previously, private high school students tend to have parents with higher-band occupations, though their levels of schooling and qualifications are otherwise fairly similar (Figure 6.15) Private high school students are also less likely to come from language backgrounds other than English, and less likely to be female (Figure 6.16).

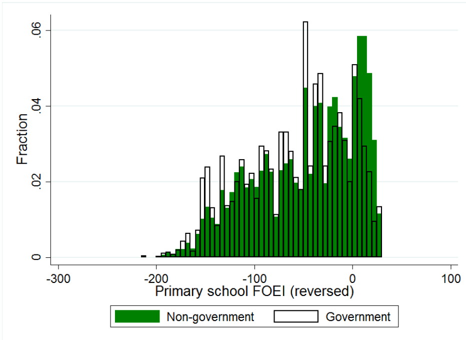
Differential attrition could result in an overestimation of treatment effects on education achievement, since it is likely that non-government high school students enjoy better education

[10] I do not consider Year 7 NAPLAN scores, reflecting the lack of high-school data for private high school students.

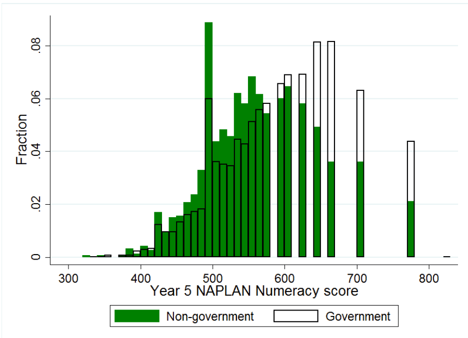
inputs than government (non-selective) high school students in my sample, for example, higher family incomes. However, the direction of bias could also be reversed. Parents who choose to send their children to private high schools after they fail to gain entry to their preferred school(s) might do so because their counterfactual local government high school is perceived to be of low quality. If this is the case, then high-quality, non-selective government high schools may be overrepresented in the control group of students, and so treatment estimates would be underestimated. I argue that the bias arising from the absence of students with better education inputs is likely to be larger than the bias arising from the self-selection of government high schools by quality, especially as it seems unlikely that parents who are sufficiently wealthy to enjoy wide school choice would also be located in suburbs with low-quality schools.



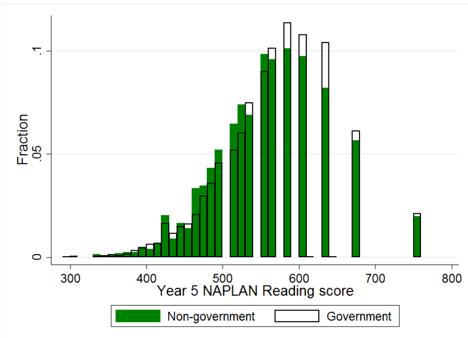
(a) Selective score



(b) Primary FOEI

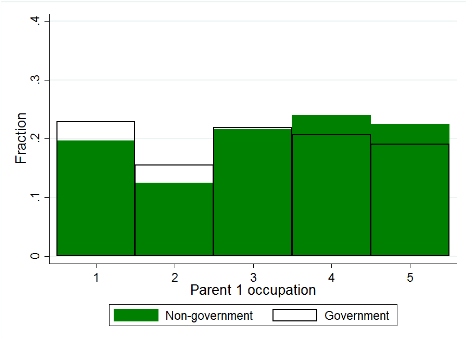


(c) Individual NAPLAN Numeracy (Year 5)

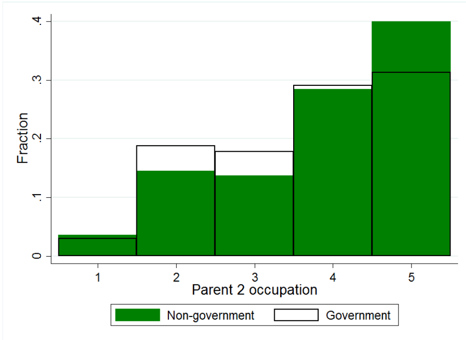


(d) Individual NAPLAN Numeracy (Year 5)

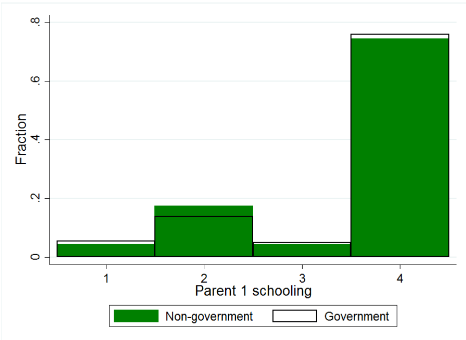
Figure 6.14: Various academic achievement and socioeconomic status indicators by type of principal high school attended



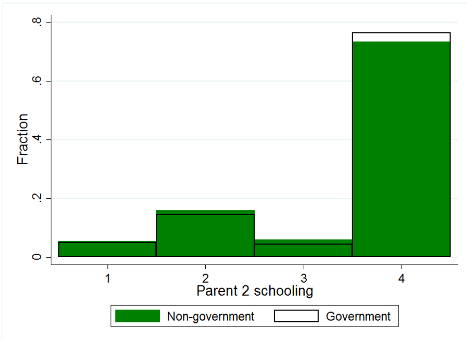
(a) Occupation band of parent 1



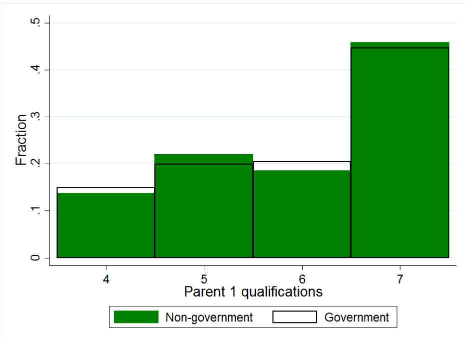
(b) Occupation band of parent 2



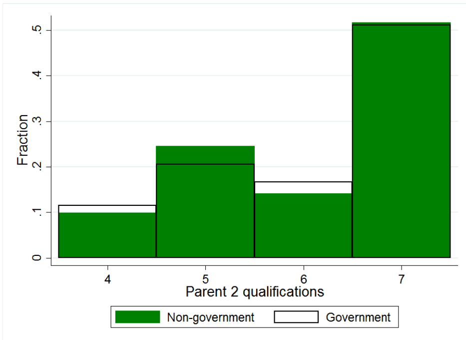
(c) Schooling level of parent 1



(d) Schooling level of parent 2



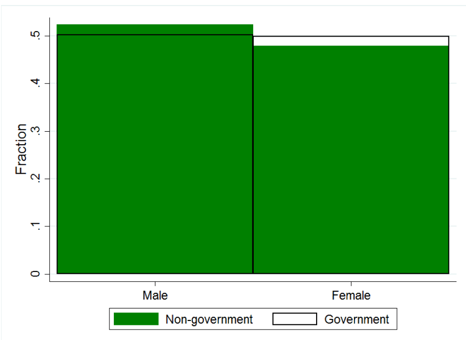
(e) Qualification level of parent 1



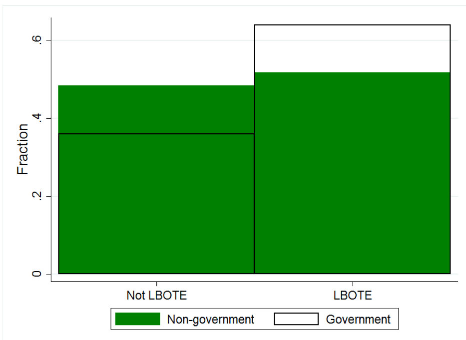
(f) Qualification level of parent 2

Figure 6.15: Parental socioeconomic indicators by type of principal high school attended

See Table 5.3 for indicator definitions.



(a) Student gender



(b) Student primary school FOEI

Figure 6.16: Student characteristics by type of principal high school attended

Chapter 7
Results

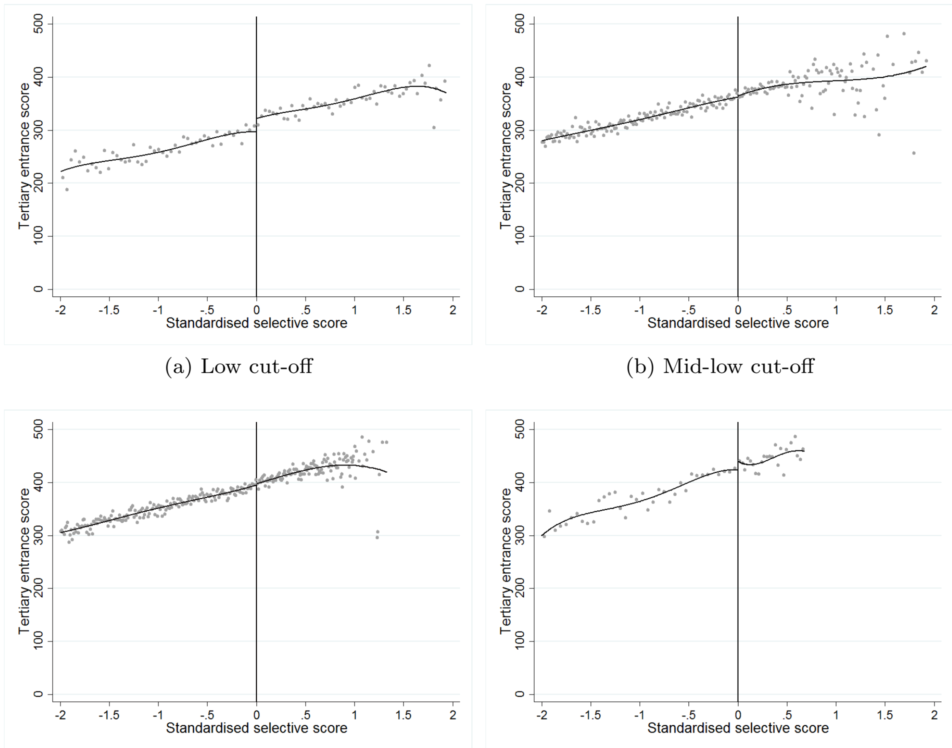
For space reasons, I only show the full regression results for the TES outcome variable using the 'offer' treatment indicator, subsample analysis, and peer effects estimates. I provide complete regression results and graphs for outcome variables other than the TES in Appendix C , and for alternate regression specifications in Appendix D. Three regression estimates are reported for each model, as estimated by the *rdrobust* package implemented by Calonico et al. (2014b). The 'conventional' results contain the conventional point estimates and (non-robust) standard errors; 'bias-corrected' results contain point estimates that have been bias-corrected for the bandwidths selected by the *rdrobust* common-bandwidth MSE-optimal bandwidth selector, with conventional standard errors; 'robust' results contain bias-corrected point estimates with robust standard errors clustered on a proxy for the primary school attended. I focus on 'robust'-estimated results in my discussion of results, as Calonico et al. (2014b) show that conventional point estimates and conventional standard errors are generally invalidated due to bandwidth-size bias as well as standard heteroskedasticity concerns. Calonico, Cattaneo, and Farrell (2016) show that robust confidence intervals are never worse than standard confidence intervals in terms of coverage error (i.e., the accuracy of a confidence interval's range). The results reported in the summary of results for all outcome variables for the entire sample in Table 7.8 and for all subgroups in Table 7.9 are these 'robust' results.

I also report other figures such as the bandwidth $\gtrless 0$, which is the largest distance to the cut-off in standardised selective scores for which an observation is included in the regression (on either side of the cut-off), the number of observations within said bandwidths, and the total number of observations available. I also include the t-statistic of the single excluded instrument (i.e. the dummy variable $A_{i,k}$ which equals 1 if the standardised selective score $Z_{i,k}$ is greater than or equal to 0) from the first-stage regression. The general rule of thumb proposed by Staiger and Stock (1997) for sufficient relevance of an instrument is an F-statistic of all excluded instruments in the first-stage regression that is greater than 10, or equivalently, a t-statistic of $\sqrt{10} \approx 3.2$ for a single excluded instrument. No first-stage instrument t-statistic is reported for regressions on samples (such as some subsamples) that exhibit full treatment assignment compliance, i.e., where the RD design is 'sharp'.

7.1 Final academic achievement: TES

As the primary outcome variable of interest, I plot students' TES in the stacked sharp samples in Figure 7.1 as well as in some specific school sharp samples in Figures 7.2 and 7.3. Figure 7.1 shows the relatively small or non-existent treatment effects for the stacked sharp samples. Treatment effects at individual school sharp samples in Figures 7.2 and 7.3 appear similarly small.

Perhaps surprisingly, TES results appear to jump slightly, but not dramatically, at the low cut-off ((Figure 7.2), even though the comprehensive high school counterfactual is most valid with these schools, and school characteristics such as peer achievement levels (Figure 5.13) or socioeconomic status (Figure 5.8) do tend to exhibit significant discontinuities at the cut-off that might lead us to expect improved outcomes. The high cut-off school sharp sample also captures a discontinuity in peer achievement levels, but again evidence for a discontinuity in students' TES is only slight. There is little evidence of discontinuities in students' TES at the middle cut-off schools, where peer achievement differentials are less pronounced.



(a) Low cut-off

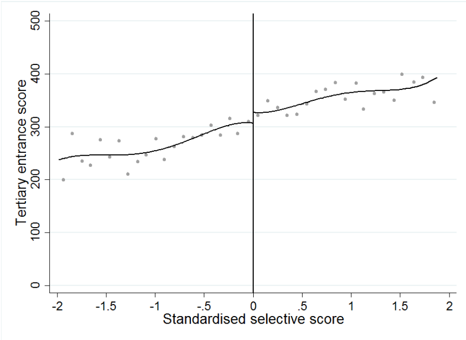
(b) Mid-low cut-off

(c) Mid-high cut-off

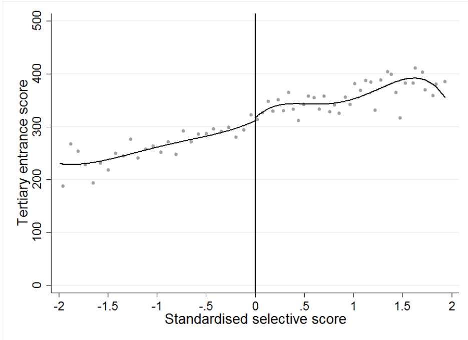
(d) High cut-off

Figure 7.1: TES in stacked sharp samples

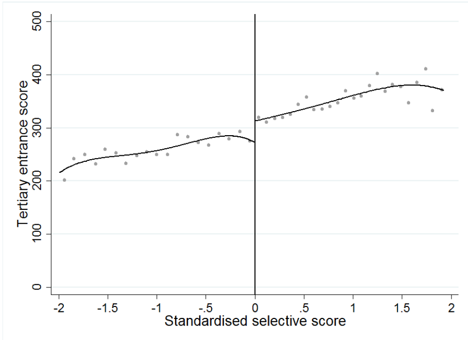
Non-parametric estimates reveal that these treatment effects are generally statistically insignificant (Table 7.1). No clear pattern to the sign of the selective school effects on TES is evident, with positive estimates for 10 out of 18 schools scattered across the selectivity spectrum. Only the effects at Schools F and Q are significant at the 5% level; C and K at the 10% level. The effect for the low cut-off stack, which Figure 7.1 indicated was the most promising candidate for a significant treatment effect, translates to an increase of 0.19 standard deviations (based on the restricted sample TES standard deviation of 88.64), but the effect is statistically insignificant. The high cut-off stack effect translates to a statistically insignificant increase in TES of 0.15 standard deviations. The mid-low and mid-high cut-off stack effects are economically and statistically insignificant.



(a) Low cut-off (School A)

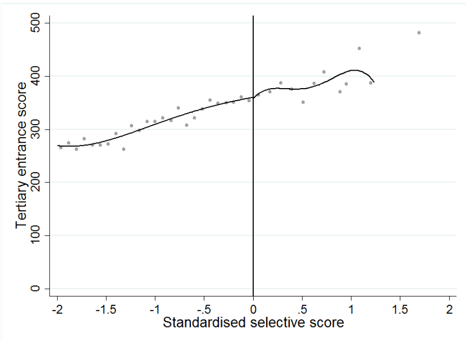


(b) Low cut-off (School B)

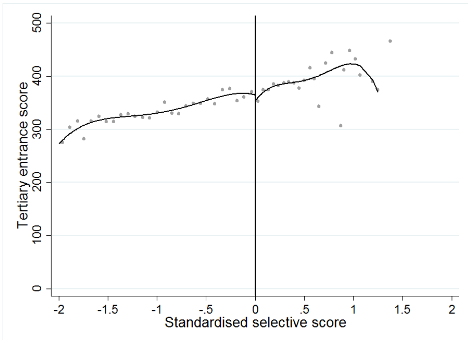


(c) Low cut-off (School C)

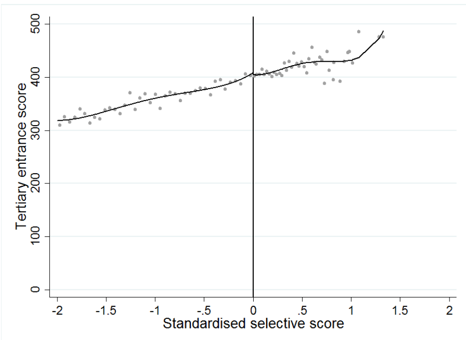
Figure 7.2: TES in the school sharp samples constituting the low cut-off stack



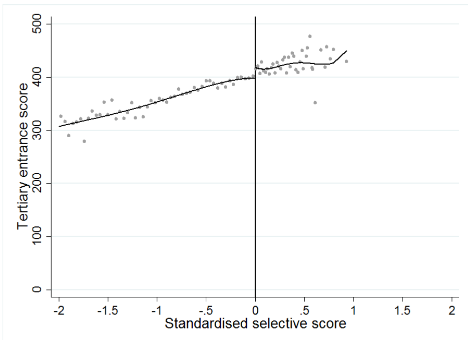
(a) Mid-low cut-off (School G)



(b) Mid-high cut-off (School J)



(c) Mid-high cut-off (School N)



(d) Mid-high cut-off (School Q)

Figure 7.3: TES in some school sharp samples representing the mid-low and mid-high cut-off stacks

Table 7.1: Fuzzy regression discontinuity estimates of effects of an offer on TES

Specifications	A	B	C	D	E	F	G	H	I	J	K
Conventional	6.680	-7.798	40.73**	3.639	-7.832	47.53**	-3.260	-14.23	-11.65	2.817	-16.12
	(14.17)	(14.83)	(19.91)	(13.63)	(9.083)	(19.72)	(16.62)	(10.68)	(13.58)	(14.93)	(10.16)
Bias-corrected	9.728	-14.57	43.39**	2.491	-11.76	52.69***	-6.221	-19.09*	-16.99	1.218	-20.37**
	(14.17)	(14.83)	(19.91)	(13.63)	(9.083)	(19.72)	(16.62)	(10.68)	(13.58)	(14.93)	(10.16)
Robust	9.728	-14.57	43.39*	2.491	-11.76	52.69**	-6.221	-19.09	-16.99	1.218	-20.37*
	(16.73)	(17.15)	(23.52)	(15.51)	(10.30)	(22.35)	(19.62)	(12.83)	(15.68)	(17.97)	(12.28)
Observations	697	1,234	1,022	1,511	1,533	1,098	1,408	1,728	2,222	1,494	2,711
First-stage instrument t-statistic	12.877	19.969	15.518	11.374	18.339	6.368	7.891	11.99	14.767	16.031	21.773
Bandwidth $\gtrless 0$	0.861	0.380	0.495	0.392	0.464	0.444	0.338	0.428	0.196	0.397	0.305
Effective obs. < 0	190	153	174	177	220	121	131	211	139	209	268
Effective obs. ≥ 0	207	166	190	214	251	146	199	236	137	239	299

Specifications	L	M	N	O	P	Q	R	Low	Mid-low	Mid-high	High
Conventional	7.047	9.327	-1.694	5.600	-8.297	19.86**	12.73*	19.04*	2.278	0.283	12.73*
	(8.162)	(6.719)	(8.625)	(9.782)	(8.846)	(8.231)	(7.324)	(11.12)	(4.577)	(4.672)	(7.324)
Bias-corrected	7.454	9.114	-1.959	8.565	-13.65	22.12***	13.45*	16.89	0.379	-0.378	13.45*
	(8.162)	(6.719)	(8.625)	(9.782)	(8.846)	(8.231)	(7.324)	(11.12)	(4.577)	(4.672)	(7.324)
Robust	7.454	9.114	-1.959	8.565	-13.65	22.12**	13.45	16.89	0.379	-0.378	13.45
	(9.638)	(7.948)	(10.91)	(12.26)	(10.60)	(9.266)	(9.105)	(13.34)	(5.500)	(5.264)	(9.105)
Observations	1,805	2,216	2,340	1,610	1,805	3,081	1,623	2,945	10,977	15,548	1,623
First-stage instrument t-statistic	21.293	36.536	24.512	16.251	17.465	18.571	23.062	23.872	34.993	39.649	23.062
Bandwidth $\gtrless 0$	0.471	0.517	0.318	0.413	0.427	0.226	0.221	0.470	0.468	0.220	0.221
Effective obs. < 0	243	474	244	271	221	264	234	441	1580	1200	234
Effective obs. ≥ 0	247	363	268	239	262	241	215	499	1665	1350	215

Each column reports regression discontinuity results for a school or stacked sharp sample. Note, School R and the high cut-off sharp samples are identical. 'Bias-corrected' and 'Robust', but not 'Conventional' specification point estimates are bias-corrected for the selected bandwidth following Calonico et al. (2014b). Estimates are non-parametrically estimated using local linear regressions and a triangular kernel. Optimal bandwidths are calculated for each sharp sample based on the common-bandwidth MSE-optimal selector following Calonico et al. (2014b) and Calonico et al. (2016) as described in the text. Observations is the total number of observations available for estimation. The sum of effective observations on either side of the cut-off are those used for estimation. Standard errors are shown in parentheses. Only standard errors for the 'Robust' specification are robust and clustered on Primary FOEI. * significant at 10%, ** significant at 5%, *** significant at 1%.

7.2 Mid-treatment academic achievement: Year 9 NAPLAN

Visual evidence indicate that mid-treatment effects appear to be broadly negligible in either Reading or Numeracy, as shown in Figures C.1 and C.2 respectively. Evidence of an effect on Reading is scattered across schools, with only School C exhibiting a statistically significant effect, and over two-thirds of schools showing negative point estimates (Table C.2). This general pattern is similar for Numeracy, although the mid-low stack effect is positive and statistically significant (Table C.3). The negative point estimates are puzzling in the light of the generally positive (though largely insignificant) estimates found for TES, although this may be explained by the conjecture discussed in Section 3.3.1 that NAPLAN is relatively low-stakes, and furthermore relatively easy for high-achieving students, which may elicit low effort.

7.3 Participation and performance by subject

7.3.1 English

There is little visual evidence for a discontinuity in participation rates in English Advanced (as opposed to English Standard), except perhaps at the mid-low cut-off stack as shown in Figure C.3. 4 individual schools exhibit significant effects (Table C.4). The results are economically significant, with increases in probability of between 10-34 percentage points although one school appears to have a negative effect (-32 percentage points) on participation probability. In addition, participation increases by 7 percentage points in the mid-low stack, with the increase significant at the 5% level.

The positive effects found here, as well as the high participation rates seen in Figure C.3, could represent purely mechanical treatment effects given the mandatory requirement of 2 units of English combined with the usual selective school practice of only offering English Advanced.

I also examine the impacts of selective school on participation rates and achievement in the Extension 1 courses, which also capture treatment effects on Extension 2 courses. Participation in English Extension 1 seems to increase somewhat discontinuously across three of the four cut-off samples (Figure C.4). The effect is statistically significant at the 1% level for the low stack (an increase of 18 percentage points), as well as at the 10% level for the mid-low stack (an increase of 7 percentage points) (Table C.5). Part of this increased participation could again be attributed to the higher participation rates in English Advanced, as taking English Advanced is required in order to take English Extension 1.

There is little visual evidence for treatment effects on English Advanced or English Extension 1 scores (Figures C.5 and C.6). This is confirmed by the regression estimates for achievement in either course, which are almost completely statistically insignificant. (Tables C.6 and C.7).

7.3.2 Mathematics

There appears to be little evidence of a discontinuities in participation rates in Mathematics (2 unit) across the four stacks, and participation seems to actually decrease at more-selective stacks (Figure C.7). While the latter fact may seem surprising *prima facie*, this corresponds to the fact that participation in Mathematics Extension 2 replaces participation in Mathematics (2 unit). As a result, this seemingly counter-intuitive result actually reflects the higher propensity of students with high selective scores to undertake the highest level of mathematics in the HSC. This can be seen in the increasing participation rates in Mathematics Extension 1, the prerequisite for Mathematics Extension 2, across the sharp samples, although these increases are not discontinuous (Figure C.8). The absence of treatment effects on course participation rates is again confirmed when examining the formal model estimates, which are generally insignificant for participation rates in Mathematics (2 unit) (Table C.8), although the 11% increase in participation in Mathematics (2 unit) for the mid-low stack is marginally significant. No stacked sharp samples exhibit a significant treatment effect on participation in Mathematics Extension 1, and only School F does so out of the school sharp samples (Table C.9).

As with English courses, treatment effects on scores in HSC mathematics courses appear to be slight or non-existent, with the possible exception of the low cut-off stack (Figures C.9 and C.10). The lack of treatment effect on Mathematics (2 unit) scores is confirmed by the formal model estimates, where only the School O effect is statistically significant (Table C.10). However, Schools A, H, and the the high cut-off stack school R exhibit statistically significant effects for Mathematics Extension 1 scores, although the effect is negative for School H (Table C.11).

7.3.3 Sciences

The evidence for treatment effects on participation in HSC science courses is relatively mixed. Of particular interest is participation in Chemistry, which is widely considered assumed knowledge for many science and engineering degrees at Australian universities. The visual evidence for discontinuities in Chemistry participation is mixed (Figure C.11), with a possible discontinuous decrease for the mid-low cut-off stack, but an increase at the high cut-off school. Statistically and economically significant results are found for School I (increase in participation rates of 42 percentage points) and the high cut-off stack school (R) (increase in participation rates of 28 percentage points) (Table C.12). Effects on participation rates in Physics are generally insignificant (Table C.13). For Biology, statistically significant effects are found for Schools F, G, and K, as well as the mid-high cut-off stack (an increase in participation rates of 9 percentage points) (Table C.14). Unsurprisingly given the above results, significant treatment effects for participation rates in science courses as a whole are mostly absent (Table C.15).

As with effects on mathematics and English course scores, there is little visual evidence for a treatment effect on Chemistry or Biology scores (Figures C.15 and C.17). However, there appear to be discontinuities in Physics scores at the low and high cut-off stacks (Table C.16). The fuzzy regression discontinuity results confirm the general absence of significant effects on Chemistry and Biology scores (Tables C.16 C.18), except for a few schools. The large high cut-off school effect on Chemistry participation found above appears to have no accompanying effect on Chemistry scores. Results for treatment effects for Physics confirm the statistical and economic significance of the discontinuities found above: Physics scores at the low and high (i.e., School R) cut-off stack are 2.6 and 2.8 points higher above the cut-off, respectively, translating to gains of approximately half a standard deviation. Schools-specific results for Physics are otherwise relatively scattered, with the four estimates significant at a 5% level split between those that are positive and negative.

7.4 Subsample analysis

In the following sections, I explore the results of regressions on subsamples based on student achievement and demographics to explore possible heterogeneous treatment effects on TES.[1] I only estimate treatment effects for stacked sharp samples to preserve reasonably large sample sizes. A summary of the results is shown in Table 7.9, while complete subsample estimation results can be found in the corresponding subsections which follow.

7.4.1 High and low achievers

The noisiness of the selective score allows for the possibility that high-achieving students to achieve a selective score that is nonetheless marginal for a given school by chance. The same logic applies for students who are relatively low-achieving. I use baseline NAPLAN scores as measures of student ability separate from the selective score, specifically, the Numeracy domain, as it seems to be domain with the most correspondence with the selectivity spectrum (Figure 6.4).

For each stacked sharp sample, I consider two sets of high achievers: those who score above the median Year 7 NAPLAN Numeracy scores amongst students in that sharp sample, and those who score above the third quartile in that sharp sample. I consider only one set of low achievers: those who score below the median NAPLAN score, as the sharp subsample of those who score below the first quartile is too small for valid estimation.

The results in Table 7.2 indicate no significant effects on high achievers, except for the marginally significant high cut-off stack effect using the above-third-quartile ability measure.

Similarly, only one marginally significant, positive effect for low achievers is found for the mid-low stack cut-off stack (Table 7.3). The low sample sizes for the low-achievers subsample suggest that it is relatively unlikely for low achievers to unexpectedly score highly and therefore appear in a narrow bandwidth around the selective school cut-offs.

[1] Another strategy to determine heterogeneous treatment effects would be to interact covariates with the forcing variable, but this practice generally leads to inconsistent estimators (Calonico et al., 2016).

Table 7.2: Fuzzy regression discontinuity estimates of effects of an offer on TES for high-achieving students

Above sharp sample median

Specifications	Low	Mid-low	Mid-high	High
Conventional	16.00	-2.633	0.816	12.63
	(12.59)	(5.080)	(4.793)	(8.669)
Bias-corrected	13.10	-4.064	0.490	13.23
	(12.59)	(5.080)	(4.793)	(8.669)
Robust	13.10	-4.064	0.490	13.23
	(14.62)	(5.772)	(5.503)	(10.30)
Observations	1,591	5,803	8,347	854
First-stage instrument t-statistic	21.259	33.133	39.435	18.779
Bandwidth $\gtrless 0$	0.422	0.425	0.223	0.185
Effective obs. < 0	261	1193	1136	173
Effective obs. ≥ 0	355	1440	1313	186

Above sharp sample third quartile

Specifications	Low	Mid-low	Mid-high	High
Conventional	22.75	-6.063	5.017	19.30**
	(18.02)	(7.128)	(5.432)	(9.686)
Bias-corrected	22.34	-7.751	5.718	20.42**
	(18.02)	(7.128)	(5.432)	(9.686)
Robust	22.34	-7.751	5.718	20.42*
	(21.20)	(7.960)	(6.310)	(11.26)
Observations	852	2,965	4,370	470
First-stage instrument t-statistic	13.003	23.017	28.9	16.307
Bandwidth $\gtrless 0$	0.511	0.417	0.223	0.174
Effective obs. < 0	102	648	726	103
Effective obs. ≥ 0	222	966	946	118

Each column reports regression discontinuity results for a school or stacked sharp sample. Note, School R and the high cut-off sharp samples are identical. 'Bias-corrected' and 'Robust', but not 'Conventional' specification point estimates are bias-corrected for the selected bandwidth following Calonico et al. (2014b). Estimates are non-parametrically estimated using local linear regressions and a triangular kernel. Optimal bandwidths are calculated for each sharp sample based on the common-bandwidth MSE-optimal selector following Calonico et al. (2014b) and Calonico et al. (2016) as described in the text. Observations is the total number of observations available for estimation. The sum of effective observations on either side of the cut-off are those used for estimation. Standard errors are shown in parentheses. Only standard errors for the 'Robust' specification are robust and clustered on Primary FOEI. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 7.3: Fuzzy regression discontinuity estimates of effects of an offer on TES for low-achieving students

Below sharp sample median

Specifications	Low	Mid-low	Mid-high	High
Conventional	19.36	32.99*	-25.53	24.75***
	(24.78)	(19.16)	(24.45)	(8.807)
Bias-corrected	18.31	37.56**	-36.89	19.27**
	(24.78)	(19.16)	(24.45)	(8.807)
Robust	18.31	37.56*	-36.89	19.27
	(30.25)	(22.48)	(28.85)	(12.07)
Observations	1,393	5,297	7,398	792
First-stage instrument t-statistic	8.224	6.494	5.216	N/A
Bandwidth $\gtrless 0$	0.381	0.241	0.184	0.0866
Effective obs. < 0	138	105	78	7
Effective obs. ≥ 0	94	100	49	11

Each column reports regression discontinuity results for a school or stacked sharp sample. Note, School R and the high cut-off sharp samples are identical. 'Bias-corrected' and 'Robust', but not 'Conventional' specification point estimates are bias-corrected for the selected bandwidth following Calonico et al. (2014b). Estimates are non-parametrically estimated using local linear regressions and a triangular kernel. Optimal bandwidths are calculated for each sharp sample based on the common-bandwidth MSE-optimal selector following Calonico et al. (2014b) and Calonico et al. (2016) as described in the text. Observations is the total number of observations available for estimation. The sum of effective observations on either side of the cut-off are those used for estimation. Standard errors are shown in parentheses. Only standard errors for the 'Robust' specification are robust and clustered on Primary FOEI. * significant at 10%, ** significant at 5%, *** significant at 1%.

7.4.2 Gender

The results of analysis on female- and male-only subsamples in Table 7.4 indicate that only females benefit at the high cut-off stack and only males benefit at the low cut-off stack in terms of TES. These results are economically significant, the female high cut-off effect and the male low cut-off effect translating to an increase of 0.20 and 0.41 standard deviations, respectively. Furthermore, all point estimates for male students are positive, whereas all point estimates except at the high cut-off stack for female students are negative. Nonetheless, most of these effects are statistically insignificant.

Table 7.4: Fuzzy regression discontinuity estimates of effects of an offer on TES for students from language backgrounds that are English or not English

Female

Specifications	Low	Mid-low	Mid-high	High
Conventional	0.0605	0.546	-6.737	14.96**
	(13.55)	(6.126)	(6.025)	(7.194)
Bias-corrected	-6.255	-2.414	-8.317	17.76**
	(13.55)	(6.126)	(6.025)	(7.194)
Robust	-6.255	-2.414	-8.317	17.76**
	(16.19)	(7.120)	(6.785)	(8.874)
Observations	1,484	4,891	8,578	661
First-stage instrument t-statistic	13.981	22.068	23.041	N/A
Bandwidth $\gtrless 0$	0.469	0.479	0.229	0.348
Effective obs. < 0	211	671	577	131
Effective obs. ≥ 0	245	697	725	108

Male

Specifications	Low	Mid-low	Mid-high	High
Conventional	32.71***	4.174	7.034	15.05
	(12.52)	(6.476)	(5.167)	(9.765)
Bias-corrected	36.39***	3.898	7.563	15.95
	(12.52)	(6.476)	(5.167)	(9.765)
Robust	36.39**	3.898	7.563	15.95
	(14.87)	(7.824)	(5.979)	(11.86)
Observations	1,465	6,088	6,963	960
First-stage instrument t-statistic	26.798	30.891	43.398	13.848
Bandwidth $\gtrless 0$	0.745	0.488	0.361	0.257
Effective obs. < 0	354	962	1061	169
Effective obs. ≥ 0	376	1013	933	141

Each column reports regression discontinuity results for a school or stacked sharp sample. Note, School R and the high cut-off sharp samples are identical. 'Bias-corrected' and 'Robust', but not 'Conventional' specification point estimates are bias-corrected for the selected bandwidth following Calonico et al. (2014b). Estimates are non-parametrically estimated using local linear regressions and a triangular kernel. Optimal bandwidths are calculated for each sharp sample based on the common-bandwidth MSE-optimal selector following Calonico et al. (2014b) and Calonico et al. (2016) as described in the text. Observations is the total number of observations available for estimation. The sum of effective observations on either side of the cut-off are those used for estimation. Standard errors are shown in parentheses. Only standard errors for the 'Robust' specification are robust and clustered on Primary FOEI. * significant at 10%, ** significant at 5%, *** significant at 1%.

7.4.3 Primary school socioeconomic status and student language background

Selective school effects might depend on a student's socioeconomic and immigrant status background. For example, selective schools could provide a more enriching learning environment for students who might otherwise attend a relatively disadvantaged non-selective school.

The socioeconomic status of the primary school attended serves as a proxy for the student's own socioeconomic status. Furthermore, it tells us what kind of school environment the student has experienced, and is predictive of the type of counterfactual school that the student might attend in the absence of a selective school offer. To identify students who attended a relatively advantaged or disadvantaged primary school, I split the sample along two different (reversed) FOEI values. The first pair of advantaged/disadvantaged subsamples is created by dividing students based on if their primary school is above or below the (reversed) FOEI state average of -100. However, this produces relatively small sample sizes for the disadvantaged subsample, as applicants tend to come from relatively advantaged primary schools. The second pair divides students based on if their primary school is above or below the (unrestricted) sample mean of -55.10.

The results shown in Table 7.6 indicate that there are insignificant treatment effects for students from relatively advantaged primary schools, except at the high cut-off stack. The high cut-off stack effect is statistically significant at the 10% level in the above state-average subsample, and at the 5% level in the above sample-average subsample.

Students from disadvantaged primary schools who are offered places in the low cut-off stack seem to benefit substantially in terms of TES. These low cut-off stack treatment effects are economically significant, translating to increases in TES of 0.40 and 0.34 standard deviations for the below state-average and sample-average subsamples respectively. Both estimated effects are statistically significant at the 5% level. There is also a significant and negative effect for the high cut-off stack, although this is likely due to biases arising from very low sample sizes.

Density discontinuity test results in Table B.2 indicate that there is a significant discontinuity in density for the low cut-off stack in the below state-average subsample, which may cast doubt on the validity of the significant effect found. However, no significant discontinuity in density is found in the below sample-average subsample, where a similar and significant effect is also found.

I also proxy for minority and/or immigrant background by dividing students by self-reported language background (either English or not). I find a marginally significant selective school effect for students from a language background other than English in the high cut-off stack (Table 7.7). Note that the size of the English language background subsample for the high cut-off stack is too small to produce regression results. Overall, there is little evidence for selective school effects that differ by language background.

Table 7.5: Fuzzy regression discontinuity estimates of effects of an offer on TES for students coming from advantaged primary schools

(Reversed) FOEI above state average (-100)

Specifications	Low	Mid-low	Mid-high	High
Conventional	15.61	-0.223	-1.847	15.13**
	(12.38)	(5.258)	(4.988)	(7.549)
Bias-corrected	12.31	-1.953	-2.869	17.22**
	(12.38)	(5.258)	(4.988)	(7.549)
Robust	12.31	-1.953	-2.869	17.22*
	(14.95)	(5.937)	(5.563)	(9.026)
Observations	2,170	9,338	13,283	1,340
First-stage instrument t-statistic	19.429	33.038	36.692	23.826
Bandwidth $\gtrless 0$	0.533	0.382	0.228	0.258
Effective obs. < 0	367	1167	1127	240
Effective obs. ≥ 0	449	1339	1253	211

(Reversed) FOEI above sample average (-55.10)

Specifications	Low	Mid-low	Mid-high	High
Conventional	8.838	-2.759	0.357	17.60**
	(18.00)	(5.896)	(4.920)	(7.581)
Bias-corrected	6.386	-4.753	-0.466	19.15**
	(18.00)	(5.896)	(4.920)	(7.581)
Robust	6.386	-4.753	-0.466	19.15**
	(21.47)	(6.688)	(5.457)	(9.193)
Observations	980	6,665	11,026	1,213
First-stage instrument t-statistic	12.238	27.323	34.964	22.53
Bandwidth $\gtrless 0$	0.575	0.356	0.246	0.250
Effective obs. < 0	175	795	1055	223
Effective obs. ≥ 0	255	957	1138	198

Each column reports regression discontinuity results for a school or stacked sharp sample. Note, School R and the high cut-off sharp samples are identical. 'Bias-corrected' and 'Robust', but not 'Conventional' specification point estimates are bias-corrected for the selected bandwidth following Calonico et al. (2014b). Estimates are non-parametrically estimated using local linear regressions and a triangular kernel. Optimal bandwidths are calculated for each sharp sample based on the common-bandwidth MSE-optimal selector following Calonico et al. (2014b) and Calonico et al. (2016) as described in the text. Observations is the total number of observations available for estimation. The sum of effective observations on either side of the cut-off are those used for estimation. Standard errors are shown in parentheses. Only standard errors for the 'Robust' specification are robust and clustered on Primary FOEI. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 7.6: Fuzzy regression discontinuity estimates of effects of an offer on TES for students coming from disadvantaged primary schools

(Reversed) FOEI below state average (-100)

Specifications	Low	Mid-low	Mid-high	High
Conventional	30.14**	-21.66	13.49	7.805
	(15.20)	(30.95)	(10.54)	(6.376)
Bias-corrected	35.59**	-26.57	16.98	9.536
	(15.20)	(30.95)	(10.54)	(6.376)
Robust	35.59**	-26.57	16.98	9.536
	(17.16)	(34.60)	(12.88)	(14.50)
Observations	775	1,639	2,265	283
First-stage instrument t-statistic	50.292	4.729	29.42	N/A
Bandwidth $\gtrless 0$	0.743	0.333	0.252	0.167
Effective obs. < 0	217	79	131	22
Effective obs. ≥ 0	137	95	151	15

(Reversed) FOEI below sample average (-55.10))

Specifications	Low	Mid-low	Mid-high	High
Conventional	27.62***	0.190	-0.323	-8.434
	(10.45)	(8.407)	(6.255)	(6.095)
Bias-corrected	30.44***	-4.159	-0.733	-19.27***
	(10.45)	(8.407)	(6.255)	(6.095)
Robust	30.44**	-4.159	-0.733	-19.27**
	(12.55)	(9.970)	(7.802)	(9.647)
Observations	1,965	4,312	4,522	410
First-stage instrument t-statistic	39.057	19.759	44.916	N/A
Bandwidth $\gtrless 0$	0.700	0.548	0.410	0.109
Effective obs. < 0	483	597	496	17
Effective obs. ≥ 0	393	542	455	21

Each column reports regression discontinuity results for a school or stacked sharp sample. Note, School R and the high cut-off sharp samples are identical. 'Bias-corrected' and 'Robust', but not 'Conventional' specification point estimates are bias-corrected for the selected bandwidth following Calonico et al. (2014b). Estimates are non-parametrically estimated using local linear regressions and a triangular kernel. Optimal bandwidths are calculated for each sharp sample based on the common-bandwidth MSE-optimal selector following Calonico et al. (2014b) and Calonico et al. (2016) as described in the text. Observations is the total number of observations available for estimation. The sum of effective observations on either side of the cut-off are those used for estimation. Standard errors are shown in parentheses. Only standard errors for the 'Robust' specification are robust and clustered on Primary FOEI. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 7.7: Fuzzy regression discontinuity estimates of effects of an offer on TES for students from language backgrounds that are English or not English

Language background other than English

Specifications	Low	Mid-low	Mid-high	High
Conventional	19.48	3.722	-1.362	14.32**
	(20.13)	(5.070)	(4.478)	(7.038)
Bias-corrected	21.51	3.098	-2.389	15.96**
	(20.13)	(5.070)	(4.478)	(7.038)
Robust	21.51	3.098	-2.389	15.96*
	(23.90)	(6.099)	(5.070)	(8.708)
Observations	572	7,838	12,476	1,394
First-stage instrument t-statistic	10.427	32.572	42.86	23.897
Bandwidth $\gtrless 0$	0.656	0.486	0.247	0.226
Effective obs. < 0	88	1189	1225	239
Effective obs. ≥ 0	120	1348	1343	208

Language background is English

Specifications	Low	Mid-low	Mid-high
Conventional	18.21	-5.296	18.98
	(12.50)	(11.54)	(16.25)
Bias-corrected	14.92	-8.262	23.24
	(12.50)	(11.54)	(16.25)
Robust	14.92	-8.262	23.24
	(14.96)	(12.77)	(18.58)
Observations	2,377	3,141	3,065
First-stage instrument t-statistic	21.245	14.159	8.538
Bandwidth $\gtrless 0$	0.456	0.405	0.290
Effective obs. < 0	371	358	147
Effective obs. ≥ 0	396	315	158

Each column reports regression discontinuity results for a school or stacked sharp sample. Note, School R and the high cut-off sharp samples are identical. 'Bias-corrected' and 'Robust', but not 'Conventional' specification point estimates are bias-corrected for the selected bandwidth following Calonico et al. (2014b). Estimates are non-parametrically estimated using local linear regressions and a triangular kernel. Optimal bandwidths are calculated for each sharp sample based on the common-bandwidth MSE-optimal selector following Calonico et al. (2014b) and Calonico et al. (2016) as described in the text. Observations is the total number of observations available for estimation. The sum of effective observations on either side of the cut-off are those used for estimation. Standard errors are shown in parentheses. Only standard errors for the 'Robust' specification are robust and clustered on Primary FOEI. * significant at 10%, ** significant at 5%, *** significant at 1%.

7.5 Summary of results by outcome variable; subsample

Table 7.8: Results summary, full analysis sample

Outcome Variable	Stacked sharp sample			
	Low	Mid-low	Mid-high	High
Tertiary entrance score	16.893	.379	-.378	13.447
	(13.335)	(5.500)	(5.264)	(9.105)
Year 9 NAPLAN Reading score	4.237	-2.709	-4.59	-16.341
	(6.143)	(4.306)	(4.527)	(10.705)
Year 9 NAPLAN Numeracy score	-2.662	11.695**	-4.701	-6.836
	(6.430)	(5.651)	(4.735)	(11.879)
Mathematics (2 unit) score	1.518	-.241	.084	.143
	(1.383)	(.734)	(.564)	(.789)
Mathematics Extension 1 score	3.005	-.373	.026	1.034**
	(2.152)	(.624)	(.577)	(.515)
English Advanced score	.893	-.494	.244	-.225
	(.739)	(.386)	(.301)	(.665)
English Extension 1 score	2.337	-.167	.198	.772
	(2.008)	(.902)	(.704)	(.822)
Chemistry score	-.154	.115	-.337	.596
	(1.327)	(.542)	(.438)	(.732)
Physics score	2.608**	-.809	.297	2.825**
	(1.300)	(.558)	(.490)	(1.393)
Biology score	-.449	-.773	.261	-1.828
	(1.012)	(.807)	(.623)	(1.394)
Mathematics (2 unit) participation = 1	-.055	.108*	.031	-.016
	(.099)	(.060)	(.051)	(.092)
Mathematics Extension 1 participation = 1	.031	.057	-.006	-.016
	(.075)	(.057)	(.041)	(.052)
English Advanced participation = 1	-.086	.066**	.021*	-.003
	(.079)	(.028)	(.012)	(.003)
English Extension 1 participation = 1	.179***	.072*	.03	-.119
	(.066)	(.042)	(.043)	(.108)
Participation in any science(s) = 1	0.095	-.097*	.036	.115
	(.086)	(.054)	(.040)	(.076)
Chemistry participation = 1	.009	-.073	-.033	.277**
	(.090)	(.065)	(.050)	(.112)
Physics participation = 1	.002	-.009	.035	.124
	(.083)	(.057)	(.048)	(.117)
Biology participation = 1	.053	-.064	.091**	-.025
	(.080)	(.049)	(.044)	(.078)

Treatment estimates are bias-corrected with robust standard errors clustered on primary school FOEI in parentheses. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 7.9: Results summary by subsamples

Subsample	Stacked sharp sample			
(Outcome variable: TES)	Low	Mid-low	Mid-high	High
High achievers (above median)	13.029	-4.064	.49	13.228
	(14.616)	(5.771)	(5.503)	(10.304)
High achievers (above third quartile)	22.289	-7.752	5.718	20.422*
	(21.194)	(7.959)	(6.310)	(11.264)
Low achievers (below median)	18.28	37.54*	-36.906	
	(30.242)	(22.476)	(28.853)	
Male	36.421**	3.898	7.517	15.933
	(14.923)	(7.824)	(5.999)	(11.861)
Female	-6.064	-2.416	-8.321	17.764**
	(16.144)	(7.119)	(6.783)	(8.874)
From advantaged primary school (above FOEI state average)	12.315	-1.953	-2.869	17.216*
	(14.948)	(5.936)	(5.563)	(9.026)
From disadvantaged primary school (below FOEI state average)	35.59**	-26.573	16.982	
	(17.155)	(34.601)	(12.881)	
From advantaged primary school (above FOEI sample average)	6.386	-4.753	-.466	19.151**
	(21.471)	(6.687)	(5.456)	(9.192)
From disadvantaged primary school (below FOEI sample average)	30.444**	-4.159	-.733	
	(12.552)	(9.970)	(7.802)	
Language background other than English	20.61	3.097	-2.42	15.952*
	(23.663)	(6.099)	(5.074)	(8.714)
Language background is English	14.868	-8.263	23.232	
	(14.951)	(12.768)	(18.587)	

Treatment estimates are bias-corrected with robust standard errors clustered on primary school FOEI in brackets. Regressions estimated with fewer than 50 total effective observations are excluded. * significant at 10%, ** significant at 5%, *** significant at 1%.

7.6 Peer effects estimation

For these regressions only, I standardise peer achievement by subtracting the restricted sample mean and dividing by the restricted sample standard deviation for ease of interpretation.

Tables 7.10 and 7.11 show treatment effect estimates of a one-standard-deviation increase in a student's average baseline peer ability in Year 7 NAPLAN Reading and Numeracy respectively. Tables D.1 and D.2 (in appendix) show the corresponding estimates for peer ability measured by Year 5 NAPLAN test results.

Many of the excluded instruments are weak, reflecting the similarities of many treatment and counterfactual peer achievement levels, especially in the mid-low and mid-high cut-off schools. As expected, the instruments for low and high cut-off schools are relatively strong, reflecting the large jumps in peer ability compared to their counterfactual schools. I do not discuss results of regressions with weak instruments (in practice, with instrument t-statistics that are less than 3.2), due to the susceptibility of such results to bias.

In general, reading or numeracy peer effects on TES appear to mostly mirror the selective school effects found above, with a general pattern of statistical insignificance. Statistically significant and positive Year 7 Reading peer effects are found for the low cut-off stack, but the corresponding Numeracy peer effect is insignificant. The estimated low cut-off stack effect for a one-standard deviation in a student's average peer achievement in Year 7 NAPLAN Reading is an economically significant increase in an individual's TES of 16.61, or 0.19 standard deviations. Effects measured using Year 5 NAPLAN tests are higher in magnitude than those using Year 7 NAPLAN, although this is likely to be due to the observed higher volatility of Year 5 NAPLAN test scores. Both Year 5 Numeracy and Reading peer effects are significant at the 5% level for the low cut-off stack, and at the 10% level for the high cut-off school.

Table 7.10: **Reading (standardised Year 7 NAPLAN) peer effects on TES**

Specifications	A	B	C	D	E	F	G	H	I	J	K
Conventional	8.175	2.571	24.33**	6.366	-1.999	56.28**	-23.60	-30.26	8.972	0.371	-26.83
	(11.65)	(11.65)	(11.24)	(6.615)	(6.303)	(24.34)	(115.7)	(19.46)	(27.15)	(25.19)	(21.82)
Bias-corrected	11.61	-0.420	27.10**	4.786	-5.082	68.17***	-40.52	-37.87*	-2.551	-7.051	-43.60**
	(11.65)	(11.65)	(11.24)	(6.615)	(6.303)	(24.34)	(115.7)	(19.46)	(27.15)	(25.19)	(21.82)
Robust	11.61	-0.420	27.10*	4.786	-5.082	68.17**	-40.52	-37.87	-2.551	-7.051	-43.60*
	(14.10)	(13.42)	(13.84)	(7.853)	(7.177)	(27.29)	(134.8)	(23.40)	(33.00)	(30.72)	(26.06)
Observations	697	1,234	1,022	1,511	1,533	1,098	1,408	1,728	2,222	1,494	2,711
First-stage instrument t-statistic	4.391	5.997	7.275	10.173	11.09	2.492	0.762	4.108	2.353	3.652	4.178
Bandwidth $\gtrless 0$	0.647	0.491	0.714	0.696	0.633	0.434	0.359	0.322	0.344	0.348	0.563
Effective obs. < 0	136	193	251	317	305	118	143	151	249	175	472
Effective obs. ≥ 0	170	197	256	303	280	144	209	191	215	215	358

Specifications	L	M	N	O	P	Q	R	Low	Mid-low	Mid-high	High
Conventional	6.475	23.85	-5.240	19.34	-260.2	4,414	24.58*	15.72**	3.968	0.705	24.58*
	(11.11)	(40.37)	(32.82)	(29.69)	(243.0)	(55,549)	(14.08)	(7.041)	(5.599)	(16.50)	(14.08)
Bias-corrected	5.055	18.46	-6.603	21.31	-212.8	-23,073	25.25*	16.61**	2.485	-1.622	25.25*
	(11.11)	(40.37)	(32.82)	(29.69)	(243.0)	(55,549)	(14.08)	(7.041)	(5.599)	(16.50)	(14.08)
Robust	5.055	18.46	-6.603	21.31	-212.8	-23,073	25.25	16.61**	2.485	-1.622	25.25
	(14.31)	(45.51)	(39.55)	(33.08)	(268.6)	(64,502)	(17.75)	(8.180)	(6.541)	(18.19)	(17.75)
Observations	1,805	2,216	2,340	1,610	1,805	3,081	1,623	2,945	10,977	15,548	1,623
First-stage instrument t-statistic	6.689	4.658	4.949	8.57	2.168	0.496	9.875	10.428	10.993	8.452	9.875
Bandwidth $\gtrless 0$	0.387	0.266	0.266	0.278	0.278	0.286	0.220	0.578	0.513	0.248	0.220
Effective obs. < 0	184	251	209	189	151	333	233	563	1758	1356	233
Effective obs. ≥ 0	240	250	251	186	178	301	215	598	1772	1494	215

Each column reports regression discontinuity results for a school or stacked sharp sample. Note, School R and the high cut-off sharp samples are identical. 'Bias-corrected' and 'Robust', but not 'Conventional' specification point estimates are bias-corrected for the selected bandwidth following Calonico et al. (2014b). Estimates are non-parametrically estimated using local linear regressions and a triangular kernel. Optimal bandwidths are calculated for each sharp sample based on the common-bandwidth MSE-optimal selector following Calonico et al. (2014b) and Calonico et al. (2016) as described in the text. Observations is the total number of observations available for estimation. The sum of effective observations on either side of the cut-off are those used for estimation. Standard errors are shown in parentheses. Only standard errors for the 'Robust' specification are robust and clustered on Primary FOEI. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 7.11: Numeracy (standardised Year 7 NAPLAN) peer effects on TES

Specifications	A	B	C	D	E	F	G	H	I	J	K
Conventional	11.29	8.945	29.99**	6.535	-1.733	51.02**	-3.148	-33.26	8.031	-4.489	-48.84
	(16.00)	(12.40)	(14.02)	(6.895)	(6.759)	(21.90)	(29.04)	(21.06)	(25.07)	(19.78)	(35.34)
Bias-corrected	16.13	6.384	33.39**	5.242	-3.202	61.15***	-2.962	-40.53*	-2.663	-9.952	-63.39*
	(16.00)	(12.40)	(14.02)	(6.895)	(6.759)	(21.90)	(29.04)	(21.06)	(25.07)	(19.78)	(35.34)
Robust	16.13	6.384	33.39*	5.242	-3.202	61.15**	-2.962	-40.53	-2.663	-9.952	-63.39
	(19.46)	(14.44)	(17.14)	(8.085)	(7.539)	(24.57)	(34.12)	(24.95)	(31.18)	(23.60)	(42.59)
Observations	697	1,234	1,022	1,511	1,533	1,098	1,408	1,728	2,222	1,494	2,711
First-stage instrument t-statistic	4.328	7.084	7.309	9.961	12.613	3.038	2.89	4.145	2.425	4.685	3.307
Bandwidth $\gtrless 0$	0.635	0.560	0.704	0.666	0.679	0.471	0.299	0.319	0.336	0.272	0.386
Effective obs. < 0	133	227	248	303	320	132	116	151	242	138	342
Effective obs. ≥ 0	168	226	251	299	284	154	179	189	209	173	329

Specifications	L	M	N	O	P	Q	R	Low	Mid-low	Mid-high	High
Conventional	9.062	21.19	-6.991	20.67	-102.0	297.6	21.33*	17.74*	1.597	1.072	21.33*
	(15.52)	(28.79)	(43.76)	(30.69)	(78.94)	(257.9)	(11.80)	(9.812)	(5.465)	(18.59)	(11.80)
Bias-corrected	6.852	17.63	-5.446	23.38	-111.4	198.5	23.74**	17.42*	0.133	-0.885	23.74**
	(15.52)	(28.79)	(43.76)	(30.69)	(78.94)	(257.9)	(11.80)	(9.812)	(5.465)	(18.59)	(11.80)
Robust	6.852	17.63	-5.446	23.38	-111.4	198.5	23.74*	17.42	0.133	-0.885	23.74*
	(20.17)	(32.56)	(49.22)	(35.99)	(87.99)	(313.8)	(14.33)	(11.27)	(6.268)	(20.08)	(14.33)
Observations	1,805	2,216	2,340	1,610	1,805	3,081	1,623	2,945	10,977	15,548	1,623
First-stage instrument t-statistic	4.202	10.102	5.508	6.972	3.284	1.463	13.642	9.670	11.816	7.979	13.642
Bandwidth $\gtrless 0$	0.391	0.281	0.276	0.333	0.293	0.306	0.238	0.487	0.439	0.215	0.238
Effective obs. < 0	186	269	218	223	153	372	253	463	1487	1175	253
Effective obs. ≥ 0	240	259	255	212	186	317	220	515	1599	1327	220

Each column reports regression discontinuity results for a school or stacked sharp sample. Note, School R and the high cut-off sharp samples are identical. 'Bias-corrected' and 'Robust', but not 'Conventional' specification point estimates are bias-corrected for the selected bandwidth following Calonico et al. (2014b). Estimates are non-parametrically estimated using local linear regressions and a triangular kernel. Optimal bandwidths are calculated for each sharp sample based on the common-bandwidth MSE-optimal selector following Calonico et al. (2014b) and Calonico et al. (2016) as described in the text. Observations is the total number of observations available for estimation. The sum of effective observations on either side of the cut-off are those used for estimation. Standard errors are shown in parentheses. Only standard errors for the 'Robust' specification are robust and clustered on Primary FOEI. * significant at 10%, ** significant at 5%, *** significant at 1%.

7.7 Alternate specifications and robustness checks

For parsimony and space reasons, I have estimated the alternately specified regressions described below using only the TES as the outcome variable. See Appendix D for the full regression outputs. Overall, changing regression specifications does not generally result in conclusions that differ to those found with the main model (Table 7.1).

7.7.1 Alternate bandwidth selectors

In Section D.2, I show the results of using different bandwidth selectors as implemented in the stata package *rdrobust*. Two classes of bandwidth selectors based on their objective functions are implemented. The first approach minimises the asymptotic mean-square error of the point estimator ('MSE-optimal' selectors) (Calonico et al., 2014a). This approach is more commonly used in the literature, and a common-bandwidth (i.e., bandwidths above and below the cut-off are of equal size) MSE-optimal selector is employed for the main regressions in this paper. The second, newer approach is designed to deliver optimal inference by minimising the coverage error rate of the confidence intervals for RD treatment effects (CER-optimal selectors) (Calonico, Cattaneo, and Farrell, 2015, 2016).

I first examine the results of allowing the MSE-optimal selector to choose different bandwidth sizes above and below the cut-off, which allows for higher sample sizes below the cut-off, where more students are generally located. As expected, there are sizable gains to sample size below the cut-off: increases of 47-156% in the stacked sharp samples. The results in Table D.3 show that significant school effects remain relatively scattered, but that the low and high cut-off stack effects become significant at the 5% and 10% significance level respectively. The point estimates using this selector also tend to be higher compared with the main model.

Implementing the common-bandwidth CER-optimal selector leads to a pattern of higher standard errors as compared to the main model (Table D.4). As in the main model, none of the stacked sharp sample effects are significant at even marginal significance levels.

Allowing the CER-optimal selector to choose different bandwidths above and below the cut-off leads to a pattern of higher statistical significance due to the increase in sample size. Amongst the stacks, only the estimated effect for the high cut-off stack becomes marginally significant (Table D.5).

7.7.2 Parametric and other specifications

I implement a parametric strategy that assumes a functional form for the relationship between the forcing and outcome variables, and which makes use of the whole sample with a uniform kernel (i.e., equal weights) Lee and Lemieux (2010). I use the quadratic polynomial, which is the highest functional form recommended by Gelman and Imbens (2014), who show that higher-order polynomials have undesirable inferential properties. The drawback of this strategy is that incorrectly specified functional forms can generate biases in the treatment effect estimates (Lee and Lemieux, 2010), whereas the strength lies in the efficiency gained by using the entire sample for estimation. Of course, given that the validity of RD designs rests on the local randomisation of subjects close to the cut-off, this strength may be rendered moot. The choice of covariates and clustering is identical to that under non-parametric estimation Table D.6 shows the results of the parametric specification regressions, which possess a general pattern of higher significance in treatment effects as compared to the main non-parametric results. In particular, the low cut-off stack effect becomes significant at the 1% level, and the mid-high and high cut-off stack effects become significant at the 10% level. These effects are economically significant, with the low cut-off stack exhibiting the biggest gains, amounting to increases in TES of 0.27 standard deviations. This increase in statistical significance is expected, given the large increases in sample size without a restrictive bandwidth and kernel. These results might indicate that the lack of statistical significance in the main regressions is a problem of sample size, although the bias arising from violations of the local randomisation assumption is potentially large.

Table D.7 shows the results of the main, non-parametric model with a kernel that is uniform rather than triangular. Using a uniform kernel shrinks the selected bandwidth for the low, mid-high and high cut-off stacks, but increases it for the mid-low cut-off stack. In general, point estimates tend to be higher compared to the results from using a triangular kernel, reflecting the relatively higher weighting of students further from the cut-off. Differences in statistical significance are minimal, with all stack effects remaining insignificant.

Table D.8 shows the results of the main, non-parametric model without covariates, which is immune to the possibility of bias due to the inclusion of covariates which are unequally distributed above and below the cut-off (Calonico et al., 2016). The results, which are based on a different set of bandwidths given the absence of covariates, lead to the same broad conclusions as found using the main model. Again, no stacked sharp sample effects are statistically significant.

Table D.9 shows the results with the alternate treatment indicator of whether or not a student accepted an offer. Note that this indicator takes on the value of 0 if the student was either not offered a place (though technically, there was no offer to accept or decline), or if the student was offered a place but declined. The estimated effects are generally greater in magnitude than those estimated in the main model, which is a result of this treatment indicator better accounting for treatment non-compliance. That is, we would expect treatment effects based on offers to be lower than treatment effects based on acceptances, as in the former case, some 'treated' students do not in fact attend the selective school. Nonetheless, the pattern of significance is largely unchanged compared to the main model, although the low cut-off stack effect becomes significant at a 10% level.

Chapter 8
Discussion and policy implications

Student tracking policies, including academically selective schools specifically, are often contentious topics of debate. In Australia, criticism in the media has centred around the allegedly ethnically segregated, high-stress and strictly academics-focused selective school environment (Broinowski, 2015; Ho, 2016), as well as selective schools acting as barriers to social mobility (Perry, 2015; Smith, 2014; Wilson, Dalton, and Baumann, 2015). Similar arguments are made in other jurisdictions with selective schools. In the U.K., 'grammar' (i.e., academically selective) schools are especially contentious, renewed in 2016 due to the push to lift a long-standing ban on new grammar schools by the Theresa May ministry. The Economist (2016), for example, puts forward similar arguments that grammar schools impede social mobility, and Husbands (2016) argues that attending a grammar school 'destroys self-confidence' and 'produces failure'. Salam (2014) argues that concentrating gifted students in New York City selective schools inherently results in racial and socioeconomic segregation, and that students on the lower end of achievement suffer from high levels of competition. In this chapter, I interpret and discuss my results to evaluate the nature of selective school effects and the validity of some of these common arguments and assertions.

8.1 Do selective schools improve overall student achievement?

The opportunity to attend a selective school represents an opportunity for a student to experience a peer group that is significantly higher achieving compared to counterfactual peer groups. The schools for which we can most clearly and credibly identify selective school effects are those belonging to the low and high cut-off stack stacks. These stacks are where the internal validity of my econometric strategy is strongest, and where counterfactual schools are most well-defined: non-selective government schools for the former, and less-selective schools for the latter.

The absence of significant selective school effects on aggregate and subject-specific assessment scores across schools, including at the low and high cut-off schools, suggests that selective schools do not improve overall student achievement, similar to conclusions reached in other selective school evaluations such as Abdulkadiroğlu et al. (2014). Increasing sample sizes through stacking fails to yield statistically significant results. These results are generally robust to alternate regression specifications, although there is some increased significance for the low and high cut-off stack effects in some cases. Furthermore, there appears to be little systematic evidence for selective school effects that vary consistently across the selectivity spectrum, a finding which contrasts with previous studies of school systems with similarly broad school selectivity, such as Ding and Lehrer (2007) in China and Pop-Eleches and Urquiola (2013) in Romania. The scattered nature of results across selective schools, for example significant and positive School F and Q effects, are then likely indicative of superior education inputs specific to those schools, rather than due to some specifically 'selective-school'-based factor. For example, these schools might possess exceptional teachers or principals, high quality institutions, or the 2007-2009 cohorts were unusually high-achieving by chance.

Given the centrality of peer quality to the selective school treatment, the generally null selective school effects constitute suggestive evidence against significant peer effects. The direct estimations of peer effects, however, suggest that there are statistically significant effects from peer ability in Reading, but not Numeracy, that are apparent for the low cut-off stack only. Peer effects might then only manifest at relatively low achievement levels (yet still high-achieving relative to the general population), or peer effects from incremental improvements in average peer achievement might be too small to be accurately estimated in my samples. The estimate translates to a 0.19 standard deviation increase in TES in response to a 1 standard deviation increase in peer ability, a result which is roughly in the middle of previous peer effect estimates (Sacerdote, 2014).

Care, however, must be taken when interpreting these results. Strictly speaking, I compare students at the margin of each school or stack: students who just made the cut-off, versus students who just missed the cut-off. The apparent paucity of overall selective school effects may reflect the possibility that students who just made the cut-off are more likely to be relatively low-achieving compared to the rest of the selective school. The selective school environment for these students, where they may face a large relative drop in peer rank, higher competition, and so on, might prove to be negative, or not as conducive to academic improvement, as compared to the average or high-achieving students. On the other hand, students who just missed the cut-off, who are otherwise still relatively high-achieving, might then face the opposite scenario at the counterfactual school, as they may be more likely to be the high achievers relative to their peers. This could, for example, boost these students' self-esteem or induce more teacher attention, and such students may actually benefit more from the counterfactual school than the selective school.

One institution that likely magnifies this effect is the common practice of internal (that is, within-school) tracking policies in both comprehensive and selective high schools. These policies commonly take the form of subject area classes (typically mathematics) that are stratified by prior achievement in that subject, or Gifted and Talented classes. Such informal tracking policies are unobserved from an institutional data standpoint and are likely to be quite heterogeneous in application. As an example, Cherrybrook Technology High School, a non-selective school, has both a Gifted and Talented program, into which entry is determined by a school-devised test and other academic measures, as well as stratified classes in some year levels, such as in English and mathematics (Cherrybrook Technology High School 2015; 2016a; 2016b). A student who just missed a selective school cut-off might be more likely to be in, say, the highest-ranking mathematics class at a comprehensive high school, whereas a student who just made the cut-off might be more likely to be in the lowest ranking class at a selective school. The relevant counterfactual environment for a selective student student in the NSW context might then be better characterised as consisting of a comparably high-achieving peer subgroup in a comprehensive high school because of highly salient internal tracking of students. These policies may also explain differences between the findings in this paper and those from others, which further demonstrates the sensitivity of interpretation to context in educational program evaluations. A selective school effect that may be absent or even negative for the marginal student at that selective school, but positive for the average or high-achieving student, might be difficult to detect in straightforward regression discontinuity designs. I further discuss the possibility of heterogeneous treatment effects by individual ability in Section 8.3.1.

8.2 Do selective schools cause higher participation rates in certain courses?

While the TES provides an aggregated 'sufficient statistic' for high school achievement, as it alone determines university entry in the vast majority of cases, I also isolate treatment effects on participation in specific HSC courses. Some courses, such as Chemistry and Mathematics (2 unit), are commonly assumed knowledge for various university degrees. In addition, students may benefit from more advanced coursework even if there is little effect on their overall scores: for example, such coursework may better prepare them for the difficulty levels or style of university courses.

There are a number of reasons to believe that participation in advanced course is affected by selective school attendance, and therefore might be expected to change at selective school cut-offs. First, since selective school cohorts are higher achieving, there is naturally more demand for such courses. Since running a course involves fixed costs and minimum scale requirements, higher demand likely directly translates to higher supply by the school. Second, course participation may be affected by the participation decisions of one's peers. Third, course participation may be affected by the school environment in general: for example, teachers who may be more experienced with talented students may be more supportive of participation or more knowledgeable about the appropriateness of such courses for a given student.

Some tracking programs specifically target specific subjects as part of their policy goals. One such example is Boston's Advanced Work Class, an academically selective class for students in Years 4-6, which has mathematics acceleration as one of its stated policy goals. Cohodes (2015) finds that one of the largest effects of the Advanced Work Class program is the increase in participation in advanced maths courses such as Algebra 1 in middle school and AP Calculus in high school. In a study of selective high schools in the UK, Clark (2010) finds little impact of selective school attendance on test scores, but positive effects on participation in advanced courses. The author notes that U.K. non-selective schools either rarely or never offered some of the advanced courses during the time period considered, whereas advanced courses are generally speaking available to most or all NSW government schools (although such course sizes are likely smaller in comprehensive high schools). Conversely, lower-level English and mathematics courses tend not to be offered at NSW selective schools as mentioned previously. This is especially true for English Advanced, which is the course with the most significant increase in participation across the stacked sharp samples, because of the requirement to take at least 2 units of English.

We can also think of the narrower menu of courses, at least for English and mathematics, as part of the selective school treatment itself. In other words, students who might have otherwise taken English Standard or Mathematics General 2 are 'forced' to take the higher-level course. My results indicate that this effect seems to exist for English Advanced, but not for Mathematics (2 unit). Although it is possible that this possible mismatch in course supply and demand would have negative consequences for achievement, the lack of treatment effects on course scores suggest that this is not the case. In fact, the narrower, more-advanced course menu might be beneficial by giving more students the option to take English or mathematics extension courses, which requires the study of the respective higher-level 2 unit courses. The significant increase in participation in English Extension 1 at the lower end of the selectivity spectrum may be evidence of this. There is, however, little evidence of an increase in participation in Mathematics Extension 1, which likely reflects the aforementioned absence of treatment effects on the participation rates in Mathematics (2 unit). Treatment effects on participation in science subjects are scattered and non-systematic. In general, there is little evidence of systematic causal selective school effects on student course participation decisions, and suggests that access to advanced courses is relatively uniform amongst selective and non-selective schools. This latter point is another example of contextual differences between the NSW and other education systems, which can exhibit significant inequities in access to advanced coursework (such as in the U.K. as mentioned above (Clark, 2010) and in the U.S. (Klugman, 2013)). This equity of access for both selective and non-selective students may be another reason why I find generally null selective school effects on academic achievement.

8.3 Do selective schools impact different types of students in different ways?

8.3.1 High and low achievers

While examining selective effects across the spectrum of selective schools allows a general overview of effects by school-level achievement, doing so fails to account for possible heterogeneous treatment effects for *individuals* of different ability. As mentioned above, if students are internally tracked within both selective and comprehensive high schools, then the relevant counterfactual environment might be relatively more nurturing or self-esteem-boosting than the treatment environment for the marginal student. Another mechanism by which students of different ability respond differently to selective school is the rank-concern theory proposed by Tincani (2015), in which students have more incentive to compete when they face more similar-ability peers due to relatively lower costs in attaining a higher rank, and less incentive when they face higher-ability students. This suggests that the relatively average- or high-achieving student competes and benefits more from selective school environments where the quality of peers is similar to said student and where the variance in student ability is low, whereas the opposite would be true for relatively low-achieving students. To test this rank-concern theory, Tincani (2015) exploits exogenous changes in peer ability from the 2010 Chilean earthquake and finds that increasing the variance of peer ability benefits low-ability students, but harms middle- and high-ability students. In the NSW context, rank is a primary academic concern, as school-grade ranks feed into the mechanism by which school assessment results are scaled and compared to the HSC standardised final exam results. As a result, the rank concern mechanism is a promising candidate for characterising peer effects in this study. To test the existence of heterogeneous treatment effects by individual ability, I use a similar approach to Abdulkadiroğlu et al. (2014) in exploiting the noisiness of the selective score, which allows for some otherwise high- or low-achieving students achieve marginal selective scores by chance. I use Year 7 NAPLAN Numeracy test scores as measures of baseline achievement that are separate from the selective score.

The point estimates of treatment effects on high achievers are positive, which might reflect this mechanism, however none of these estimates are statistically significant at the 5% level. This might indicate that the true treatment effect on high achievers is insignificant, or that the effect is too small to be detected by my analysis. The marginally significant effect for ultra-high achievers (those above the third quartile of in-sharp-sample NAPLAN performance) in the high cut-off stack might be some evidence that these theorised benefits exist for only a specific and exceptional subgroup of students in exceptional school environments. The general statistical insignificance of treatment effects for high achievers is similar to those found by Abdulkadiroğlu et al. (2014). On the other hand, treatment estimates for low achievers are actually positive for the low and mid-low stacks, though negative for the mid-high stack. Of these, only the mid-low stack effect is (marginally) statistically significant. These findings seem to contradict the rank concern mechanism and popular assertions of detrimental effects to the marginal selective school student, at least in terms of academic achievemet, although estimates may suffer from low sample sizes, measurement error, as well as the coarseness of the definitions of high and low achievement.

8.3.2 Male and female students

Some studies also find gains to selective school attendance that are highly gender-specific. From a theoretical standpoint, psychological 'structures of the self' might be more socially interdependent for females and more socially independent for males (Cross and Madson, 1997). As a result, peer achievement effects might be more salient for females than males, which would be beneficial for females if peer effects are positive. On the other hand, experimental evidence presented by Gneezy, Niederle, and Rustichini (2003) suggests that males, but not females, experience a significant increase in performance on an arbitrary work task in response to increased competitiveness, as may be the case in a selective school environment. Jackson (2010) finds that peer achievement effects in Trinidad & Tobago are 38% larger for females than males, a statistically significant difference.

I find that there is a statistically significant selective school effect for females in the high cut-off stack and for males in the low cut-off stack. In fact, the point estimates for females in the other three stacks are negative, although statistically insignificant; for males, the point estimates is positive for all stacks. This might suggest that the environment of increased competitiveness at low cut-off selective schools as compared to comprehensive schools is beneficial to male students and possibly detrimental to female students, but that female students thrive in exceptionally high-achieving environments. This contrasts with other research in Australia by Houng (forthcoming), who finds more robust and significantly higher gains to selective school attendance for females than males.

8.3.3 Selective schools as equalisers of opportunity

School choice in NSW is relatively limited for disadvantaged students, who usually attend the comprehensive school in the local catchment area they happen to reside in. Such students are unlikely to be able to afford to attend private schools, where yearly fees are estimated to cost A$10,174-23,524 on average (Australian Scholarships Group, 2016), and are likely to be limited in their ability to move to districts with better schools. As evidence of the persistence of socioeconomic advantage in education, there is a high correlation ($\rho = 0.81$) between a student's primary school FOEI and high school FOEI. Selective schools, however, expand school choice for gifted and relatively disadvantaged and/or immigrant-background students because they are publicly administered and geography plays no role in the admittance criteria. Conditioning on students who do principally attend a selective high school (of any type), the primary-high school FOEI correlation weakens significantly ($\rho = 0.59$).

This expanded school choice could result in selective school effects that are more pronounced for disadvantaged students. Low socioeconomic status students in selective schools may face reduced conflicts between educational attainment and incentives to conform to social groups, in a manner similar to the 'acting white' phenomenon analysed by Austen-Smith and Fryer (2005). Other theoretical reasons for high-achieving peer effects, such as information spillovers, or teacher behaviour and expectations, are likely to also be amplified for low socioeonomic status students because of the relatively larger changes in school environments for these students.

By dividing the sample by primary-school FOEI, I am able to better differentiate selective treatment effects as compared to either advantaged and disadvantaged counterfactual government schools. The economically and statistically significant treatment effects for disadvantaged students (as measured by primary-school FOEI) for the low cut-off stack is evidence in favour of selective schools as equalisers of opportunity. This result is in line with the economics of education literature, where benefits to selective education are found to be high for disadvantaged students. For example, Cohodes (2015) finds improvements to college enrolments for minority students in a Year 4-6 tracking program in Boston, Massachusetts, and Abdulkadiroğlu et al. (2014) find some evidence for improvements in English score outcomes for minorities in Boston exam schools. Card and Giuliano (2016) find significant achievement gains for black and Hispanic students who attend

Gifted and Talented classes, which the authors attribute to factors such as teacher expectations and reduced negative peer effects. On the other end of socioeconomic advantage, Card and Giuliano (2016) do not find significant gains for non-minority or non-low-income students. Similarly, I do not find significant treatment effects for advantaged students, except at possibly the high cut-off school, suggesting that selective schools do not additionally benefit students who might otherwise enjoy greater school choice or better education inputs.

Dividing the sample by language background in an attempt to proxy for minority status yields little in the way of significant results. While this might suggest that minority students do not benefit more than the average student, which contradicts the findings of previous studies mentioned above, this discrepancy is likely a result of differing geographic contexts. For example, measuring minority status by language background pools together all students from any minority group, such as Asian Australians, a group which is less comparable to the usual U.S. definition of minorities (largely black and Hispanic individuals, and usually excluding Asian Americans). A more comparable minority group might be, for instance, aboriginal Australians, but the very low proportion of aboriginal Australians in selective schools precludes any possible subsample analysis thereof.

The results discussed above raises the question: if disadvantaged students benefit the most from selective schools, to what extent do selective schools serve disadvantaged students, and might this have implications for policy? Dustan (2010) and Atkinson et al. (2006) document low admission rates for low socioeconomic status students in Mexico City and U.K. selective schools respectively, despite finding similar potential benefits for such students. In NSW, selective schools also tend to serve higher socioeconomic status students, with the distribution of students in selective schools sharply skewed towards the higher socioeconomic quartiles (measured by ICSEA). Only 3.7% of students in fully selective schools are in the bottom ICSEA quartile, compared to 24.3% of students in non-selective government schools (Table 5.2). In contrast, 65.93% of students in fully selective schools are in the top ICSEA quartile, compared to 26.6% of students in non-selective government schools.

I argue that there are two main reasons for this overrepresentation of high socioeconomic status students in selective schools. First, selective school applications are voluntary, and significant self-selection on socioeconomic status occurs. Second, amongst those who apply, there exists a strong correlation between the average socioeconomic status of a student's primary school and the selective score achieved.

The correlation between socioeconomic status and selective scores suggests that differences in the quality of education inputs for advantaged and disadvantaged students cause significant achievement differentials over the period of primary, and potentially early childhood, education. This might in itself drive the self-selection in selective school applications, if parents choose to apply on the basis of their perception of their child's ability. An example of how differences in pre-high school education inputs arise in NSW might be the 'Opportunity Class' (OC) program, which are academically selective classrooms embedded in some government primary schools, and for which the admission process is very similar to selective high schools (High Performing Students Unit, 2016a). In a large U.S. school district, Card and Giuliano (2015) find that minority, low-income, and language background other than English students in primary school Gifted and Talented programs (similar to the OC program) were systematically under-represented in a system where parents or teachers nominated students for eligibility testing. If there is similar self-selection on socioeconomic status in OC applications, then the OC program may amplify self-selection in selective school applications and selective score differentials if OC education inputs are superior to those in non-OC classrooms. Policies aimed at improving identification and nurturing of high-achieving and talented students in relatively disadvantaged schools might then reduce disparities in education inputs prior to high school.

On the other hand, the self-selection in applications might be driven by information and/or preference disparities between low and high socioeconomic families. That is, low socioeconomic status families might be less informed about selective schools or the application process, or place less importance on education. This would imply that policies in the form of information campaigns or potentially inexpensive 'nudges' (Thaler and Sunstein, 2008) to improve application rates amongst low socioeconomic status families would improve education outcomes. Another potential policy to reduce self-selection in applications is universal screening, that is, making selective school applications mandatory or opt-out, perhaps at least for families in suburbs close to selective schools. For example, Card and Giuliano (2015) find that switching from the voluntary applications protocol mentioned above to universal screening significantly improved the representation of disadvantaged students in the Gifted and Talented programs.

One caveat implicit in using primary-school FOEI as a measure of socioeconomic status is that it masks heterogeneity in individual-level socioeconomic status within a primary school, unlike common individual-level indicators for disadvantage in U.S. datasets such as free-lunch eligibility. While we might be able to interpret the treatment effects as a comparison of selective schools to advantaged or disadvantaged non-selective schools due to the primary-high school FOEI correlation, the use of average school socioeconomic status might affect the interpretation of treatment effects for disadvantaged or advantaged student' *per se*. For example, the students who receive selective school offers might have a relatively higher socioeconomic status despite the relatively low socioeconomic status of their peers at their primary schools. However, if primary school FOEI is a poor predictor of individual socioconomic status, we might expect that treatment effects for either below- or above-average primary school FOEI students to be relatively similar, which I find is not the case.

Chapter 9
Conclusion

Exploiting fuzzy regression discontinuity designs for three cohorts of applicants to NSW selective schools, I find that receiving an offer to attend a fully selective school does not appear to increase overall student achievement in high-stakes HSC assessments. Commonly espoused arguments that the selective school environment negatively affects students do not seem to be supported by the evidence, at least with respect to academic achievement. I do find evidence for positive, economically and statistically significant selective school effects for disadvantaged students (but little evidence for similar benefits to advantaged students) indicating the potential benefits of expanding school choice for such students. This suggests that the argument that selective schools impede social mobility by unfairly advantaging high socioeconomic status students may be relatively unfounded. However, I find that disadvantaged students are relatively underrepresented in selective schools.

One caveat with my results is that the lack of outcome data for non-government high school students potentially biases my results in a likely upward direction. I partially address this potential issue by formally testing for the presence of this and other kinds of sorting, which do not appear to invalidate the regression discontinuity designs for the samples of most interest. The scope of this paper is also limited to academic achievement and course participation rates in high school, and does not include long-term and/or non-academic selective school effects that may yet exist, for example, on incomes, formation of stronger professional networks, or in fostering a sense of belonging.

My results suggest that measures that improve the representation of low socioeconomic status students in selective schools, for example, through improved identification and support of high-achieving students in disadvantaged primary schools, are likely to improve educational outcomes. These policy implications apply most to NSW and other Australian jurisdictions. We should, however, be cautious in extending the results found here or in other evaluations, as the relatively mixed effects found across the selective school literature likely indicate the sensitivity of education program evaluations to cultural and institutional contexts.

Appendices

Appendix A

Continuity tests for covariates in main model

Continuity tests model the covariate as the dependent variable in the same RD framework as the main model, except with no covariates.

A.1 Gender

Note: single-sex schools are excluded.

Table A.1: Fuzzy regression discontinuity estimates of placebo effects of an offer on Gender (non-parametric)

Specifications	A	B	C	D	E	F	I	K
Conventional	0.326**	0.0804	0.0891	-0.143	-0.00159	-0.0983	0.00350	0.0859
	(0.161)	(0.0864)	(0.126)	(0.101)	(0.124)	(0.166)	(0.120)	(0.0942)
Bias-corrected	0.364**	0.0933	0.105	-0.159	-0.0151	-0.160	0.0141	0.0867
	(0.161)	(0.0864)	(0.126)	(0.101)	(0.124)	(0.166)	(0.120)	(0.0942)
Robust	0.364*	0.0933	0.105	-0.159	-0.0151	-0.160	0.0141	0.0867
	(0.190)	(0.103)	(0.142)	(0.116)	(0.145)	(0.194)	(0.146)	(0.111)
Observations	721	1,280	1,051	1,639	1,588	1,179	2,342	2,804
First-stage instrument t-statistic	9.547	24.525	16.319	13.665	14.372	7.131	14.702	20.158
Bandwidth $\gtrless 0$	0.549	0.608	0.497	0.469	0.359	0.532	0.269	0.291
Effective obs. < 0	120	254	180	233	179	173	194	255
Effective obs. ≥ 0	155	246	194	256	211	171	175	291

Specifications	Q	R	Low	Mid-low	Mid-high	High
Conventional	-0.0311	-0.000866	0.170*	-0.0109	0.0518	-0.000187
	(0.0918)	(0.0875)	(0.0882)	(0.0427)	(0.0487)	(0.0949)
Bias-corrected	-0.0368	0.0247	0.193**	-0.0131	0.0520	0.0292
	(0.0918)	(0.0875)	(0.0882)	(0.0427)	(0.0487)	(0.0949)
Robust	-0.0368	0.0247	0.193*	-0.0131	0.0520	0.0292
	(0.104)	(0.108)	(0.101)	(0.0479)	(0.0553)	(0.115)
Observations	3,195	1,658	3,045	11,493	16,067	1,658
First-stage instrument t-statistic	18.664	25.348	21.129	40.9	40.273	26.301
Bandwidth $\gtrless 0$	0.241	0.245	0.390	0.489	0.221	0.257
Effective obs. < 0	278	261	390	1694	1218	275
Effective obs. ≥ 0	258	222	433	1738	1357	228

Each column reports regression discontinuity results for a school or stacked sharp sample. Note, School R and the high cut-off sharp samples are identical. 'Bias-corrected' and 'Robust', but not 'Conventional' specification point estimates are bias-corrected for the selected bandwidth following Calonico et al. (2014b). Estimates are non-parametrically estimated using local linear regressions and a triangular kernel. Optimal bandwidths are calculated for each sharp sample based on the common-bandwidth MSE-optimal selector following Calonico et al. (2014b) and Calonico et al. (2016) as described in the text. Observations is the total number of observations available for estimation. The sum of effective observations on either side of the cut-off are those used for estimation. Standard errors are shown in parentheses. Only standard errors for the 'Robust' specification are robust and clustered on Primary FOEI. * significant at 10%, ** significant at 5%, *** significant at 1%.

A.2 Language background other than English

Table A.2: Fuzzy regression discontinuity estimates of placebo effects of an offer on language background other than English (non-parametric)

Specifications	A	B	C	D	E	F	G	H	I	J	K
Conventional	-0.0155	0.0398	0.0270	-0.0915	0.138	-0.0688	-0.0967	0.0726	-0.173	-0.0882	-0.0721
	(0.122)	(0.0985)	(0.0567)	(0.142)	(0.134)	(0.158)	(0.0643)	(0.0910)	(0.106)	(0.0856)	(0.0661)
Bias-corrected	-0.0205	0.0399	0.00878	-0.174	0.148	-0.0951	-0.174***	0.121	-0.217**	-0.0970	-0.0856
	(0.122)	(0.0985)	(0.0567)	(0.142)	(0.134)	(0.158)	(0.0643)	(0.0910)	(0.106)	(0.0856)	(0.0661)
Robust	-0.0205	0.0399	0.00878	-0.174	0.148	-0.0951	-0.174**	0.121	-0.217*	-0.0970	-0.0856
	(0.141)	(0.115)	(0.0626)	(0.159)	(0.143)	(0.183)	(0.0775)	(0.105)	(0.125)	(0.0971)	(0.0745)
Observations	721	1,280	1,052	1,639	1,588	1,179	1,455	1,774	2,342	1,533	2,804
First-stage instrument t-statistic	10.083	24.925	15.017	13.415	16.251	7.625	9.803	12.178	13.987	16.632	21.198
Bandwidth $\gtrless 0$	0.604	0.670	0.494	0.456	0.406	0.564	0.442	0.440	0.217	0.418	0.312
Effective obs. < 0	133	292	180	225	204	185	180	219	157	222	274
Effective obs. ≥ 0	164	267	194	251	233	180	251	242	150	252	303

Specifications	L	M	N	O	P	Q	R	Low	Mid-low	Mid-high	High
Conventional	0.104	0.00790	-0.0641	-0.0619	-0.0998	0.000520	-0.0542**	-0.00565	-0.0321	-0.0378	-0.0543*
	(0.0929)	(0.0518)	(0.0452)	(0.0527)	(0.0799)	(0.0365)	(0.0253)	(0.0627)	(0.0463)	(0.0253)	(0.0284)
Bias-corrected	0.115	-0.00446	-0.0702	-0.0750	-0.123	-0.00544	-0.0503**	-0.0348	-0.0467	-0.0407	-0.0502*
	(0.0929)	(0.0518)	(0.0452)	(0.0527)	(0.0799)	(0.0365)	(0.0253)	(0.0627)	(0.0463)	(0.0253)	(0.0284)
Robust	0.115	-0.00446	-0.0702	-0.0750	-0.123	-0.00544	-0.0503	-0.0348	-0.0467	-0.0407	-0.0502
	(0.108)	(0.0597)	(0.0575)	(0.0616)	(0.0981)	(0.0411)	(0.0316)	(0.0754)	(0.0514)	(0.0290)	(0.0350)
Observations	1,857	2,286	2,416	1,661	1,868	3,195	1,658	3,046	11,493	16,067	1,658
First-stage instrument t-statistic	12.573	35.397	41.067	24.952	18.012	14.605	23.827	21.062	31.543	46.223	24.065
Bandwidth $\gtrless 0$	0.293	0.496	0.797	0.835	0.452	0.187	0.241	0.422	0.351	0.274	0.255
Effective obs. < 0	136	457	575	456	229	202	259	416	1176	1504	274
Effective obs. ≥ 0	191	355	362	333	278	199	221	458	1356	1594	226

Each column reports regression discontinuity results for a school or stacked sharp sample. Note, School R and the high cut-off sharp samples are identical. 'Bias-corrected' and 'Robust', but not 'Conventional' specification point estimates are bias-corrected for the selected bandwidth following Calonico et al. (2014b). Estimates are non-parametrically estimated using local linear regressions and a triangular kernel. Optimal bandwidths are calculated for each sharp sample based on the common-bandwidth MSE-optimal selector following Calonico et al. (2014b) and Calonico et al. (2016) as described in the text. Observations is the total number of observations available for estimation. The sum of effective observations on either side of the cut-off are those used for estimation. Standard errors are shown in parentheses. Only standard errors for the 'Robust' specification are robust and clustered on Primary FOEI. * significant at 10%, ** significant at 5%, *** significant at 1%.

A.3 Primary school FOEI

Table A.3: Fuzzy regression discontinuity estimates of placebo effects of an offer on primary FOEI (non-parametric)

Specifications	A	B	C	D	E	F	G	H	I	J	K
Conventional	16.29	7.188	-3.806	-8.023	5.083	13.75	18.65**	4.699	4.605	0.517	10.86
	(11.02)	(10.71)	(14.49)	(14.73)	(7.749)	(9.729)	(9.397)	(11.39)	(11.04)	(6.602)	(13.05)
Bias-corrected	21.07*	5.556	-5.059	-11.18	5.672	16.59*	22.55**	3.770	3.656	-0.526	13.74
	(11.02)	(10.71)	(14.49)	(14.73)	(7.749)	(9.729)	(9.397)	(11.39)	(11.04)	(6.602)	(13.05)
Robust	21.07*	5.556	-5.059	-11.18	5.672	16.59	22.55**	3.770	3.656	-0.526	13.74
	(11.98)	(12.20)	(15.65)	(15.87)	(8.608)	(10.55)	(10.64)	(13.00)	(12.61)	(7.678)	(14.53)
Observations	721	1,280	1,052	1,639	1,588	1,179	1,455	1,774	2,342	1,533	2,804
First-stage instrument t-statistic	12.249	28.668	18.41	11.764	14.669	6.718	11.012	15.829	17.232	18.368	19.067
Bandwidth $\gtreqless 0$	0.750	0.838	0.570	0.392	0.364	0.545	0.511	0.646	0.475	0.487	0.277
Effective obs. < 0	179	351	212	190	179	178	210	311	357	251	241
Effective obs. ≥ 0	192	318	218	215	212	174	265	304	257	283	272

Specifications	L	M	N	O	P	Q	R	Low	Mid-low	Mid-high	High
Conventional	4.651	-4.141	-0.671	2.426	8.456	1.348	-2.447	7.694	4.679	0.945	-6.463
	(7.919)	(12.05)	(9.683)	(7.959)	(13.35)	(12.85)	(16.76)	(6.869)	(3.881)	(3.249)	(8.754)
Bias-corrected	5.583	-3.216	1.242	3.401	11.96	1.542	-9.187	7.376	4.624	1.973	-9.960
	(7.919)	(12.05)	(9.683)	(7.959)	(13.35)	(12.85)	(16.76)	(6.869)	(3.881)	(3.249)	(8.754)
Robust	5.583	-3.216	1.242	3.401	11.96	1.542	-9.187	7.376	4.624	1.973	-9.960
	(8.995)	(12.79)	(10.08)	(9.258)	(14.89)	(13.91)	(18.97)	(7.963)	(4.341)	(3.791)	(10.38)
Observations	1,857	2,286	2,416	1,661	1,868	3,195	1,658	3,044	11,493	16,067	1,658
First-stage instrument t-statistic	13.71	35.228	42.95	17.238	14.003	21.312	28.583	20.895	33.278	55.058	28.789
Bandwidth $\gtreqless 0$	0.326	0.501	0.889	0.517	0.289	0.295	0.400	0.399	0.378	0.370	0.279
Effective obs. < 0	158	462	637	324	154	351	401	394	1285	2039	294
Effective obs. ≥ 0	210	358	366	278	183	314	278	440	1442	1948	239

Each column reports regression discontinuity results for a school or stacked sharp sample. Note, School R and the high cut-off sharp samples are identical. 'Bias-corrected' and 'Robust', but not 'Conventional' specification point estimates are bias-corrected for the selected bandwidth following Calonico et al. (2014b). Estimates are non-parametrically estimated using local linear regressions and a triangular kernel. Optimal bandwidths are calculated for each sharp sample based on the common-bandwidth MSE-optimal selector following Calonico et al. (2014b) and Calonico et al. (2016) as described in the text. Observations is the total number of observations available for estimation. The sum of effective observations on either side of the cut-off are those used for estimation. Standard errors are shown in parentheses. Only standard errors for the 'Robust' specification are robust and clustered on Primary FOEI. * significant at 10%, ** significant at 5%, *** significant at 1%.

A.4 Year 7 NAPLAN Reading

Table A.4: Fuzzy regression discontinuity estimates of placebo effects of an offer on individual Year 7 NAPLAN Reading scores (non-parametric)

Specifications	A	B	C	D	E	F	G	H	I	J	K
Conventional	-4.931	3.798	-25.17**	-7.630	-2.656	10.32	4.601	6.898	-6.245	-15.28	1.350
	(12.97)	(11.10)	(12.26)	(18.35)	(11.76)	(23.41)	(11.81)	(10.28)	(10.75)	(9.970)	(9.697)
Bias-corrected	-3.134	6.583	-27.89**	-6.741	-2.797	9.683	5.227	7.936	-8.472	-18.04*	4.040
	(12.97)	(11.10)	(12.26)	(18.35)	(11.76)	(23.41)	(11.81)	(10.28)	(10.75)	(9.970)	(9.697)
Robust	-3.134	6.583	-27.89**	-6.741	-2.797	9.683	5.227	7.936	-8.472	-18.04	4.040
	(14.52)	(13.01)	(13.88)	(20.68)	(12.62)	(26.94)	(14.20)	(12.01)	(12.95)	(11.84)	(11.45)
Observations	699	1,238	1,025	1,513	1,540	1,100	1,411	1,733	2,226	1,500	2,717
First-stage instrument t-statistic	9.809	22.802	14.436	11.113	15.874	6.673	10.637	13.502	15.757	18.223	20.103
Bandwidth $\gtrless 0$	0.660	0.568	0.439	0.392	0.400	0.491	0.483	0.517	0.333	0.491	0.299
Effective obs. < 0	140	231	151	177	196	140	195	263	240	252	259
Effective obs. ≥ 0	174	231	167	214	228	159	261	266	209	281	298

Specifications	L	M	N	O	P	Q	R	Low	Mid-low	Mid-high	High
Conventional	-10.67	3.062	-3.089	5.552	15.35	13.77	12.31	-3.284	-4.745	-1.838	11.63
	(13.57)	(5.533)	(5.869)	(7.031)	(9.718)	(8.547)	(7.852)	(8.523)	(4.962)	(4.535)	(8.741)
Bias-corrected	-13.14	0.260	-5.560	0.787	22.10**	13.30	16.30**	-1.404	-4.929	-2.586	15.30*
	(13.57)	(5.533)	(5.869)	(7.031)	(9.718)	(8.547)	(7.852)	(8.523)	(4.962)	(4.535)	(8.741)
Robust	-13.14	0.260	-5.560	0.787	22.10*	13.30	16.30*	-1.404	-4.929	-2.586	15.30
	(15.46)	(6.299)	(7.182)	(8.233)	(11.40)	(10.01)	(9.151)	(9.908)	(5.500)	(5.145)	(10.34)
Observations	1,809	2,220	2,344	1,614	1,809	3,090	1,628	2,956	11,006	15,583	1,628
First-stage instrument t-statistic	11.769	38.807	42.951	19.927	16.235	22.166	25.779	20.658	35.184	39.362	27.36
Bandwidth $\gtrless 0$	0.277	0.585	0.876	0.584	0.374	0.294	0.249	0.400	0.408	0.215	0.266
Effective obs. < 0	124	516	615	355	202	350	265	384	1369	1176	284
Effective obs. ≥ 0	182	387	364	291	229	313	221	432	1521	1327	232

Each column reports regression discontinuity results for a school or stacked sharp sample. Note, School R and the high cut-off sharp samples are identical. 'Bias-corrected' and 'Robust', but not 'Conventional' specification point estimates are bias-corrected for the selected bandwidth following Calonico et al. (2014b). Estimates are non-parametrically estimated using local linear regressions and a triangular kernel. Optimal bandwidths are calculated for each sharp sample based on the common-bandwidth MSE-optimal selector following Calonico et al. (2014b) and Calonico et al. (2016) as described in the text. Observations is the total number of observations available for estimation. The sum of effective observations on either side of the cut-off are those used for estimation. Standard errors are shown in parentheses. Only standard errors for the 'Robust' specification are robust and clustered on Primary FOEI. * significant at 10%, ** significant at 5%, *** significant at 1%.

A.5 Year 7 NAPLAN Numeracy

Table A.5: Fuzzy regression discontinuity estimates of placebo effects of an offer on individual Year 7 NAPLAN Numeracy scores (non-parametric)

Specifications	A	B	C	D	E	F	G	H	I	J	K
Conventional	-21.57*	-8.494	6.646	-4.904	-10.93	22.14*	17.71*	-3.552	-11.64	-0.339	-4.945
	(13.04)	(8.738)	(12.89)	(12.50)	(9.596)	(11.77)	(9.561)	(14.61)	(11.23)	(10.21)	(10.55)
Bias-corrected	-26.37**	-9.738	10.18	-5.595	-13.81	24.29**	18.09*	-2.906	-10.06	2.087	-7.939
	(13.04)	(8.738)	(12.89)	(12.50)	(9.596)	(11.77)	(9.561)	(14.61)	(11.23)	(10.21)	(10.55)
Robust	-26.37*	-9.738	10.18	-5.595	-13.81	24.29*	18.09	-2.906	-10.06	2.087	-7.939
	(14.70)	(9.996)	(14.71)	(14.24)	(10.84)	(13.18)	(11.16)	(17.74)	(13.60)	(12.14)	(12.29)
Observations	699	1,237	1,028	1,512	1,538	1,098	1,408	1,729	2,223	1,496	2,714
First-stage instrument t-statistic	10.881	23.318	15.048	8.964	15.428	6.313	12.567	10.772	14.904	14.992	20.922
Bandwidth $\gtrless 0$	0.649	0.613	0.456	0.304	0.388	0.461	0.609	0.389	0.277	0.370	0.315
Effective obs. < 0	138	250	158	132	187	127	237	188	197	200	274
Effective obs. ≥ 0	170	244	175	164	225	152	283	220	180	230	305

Specifications	L	M	N	O	P	Q	R	Low	Mid-low	Mid-high	High
Conventional	0.756	1.823	0.149	-7.782	2.862	-7.605	-18.46	-4.851	-1.840	-5.724	-18.94
	(11.41)	(7.453)	(6.604)	(9.512)	(14.67)	(14.27)	(12.33)	(6.847)	(6.110)	(4.622)	(12.85)
Bias-corrected	-0.218	4.542	2.157	-15.41	4.383	-9.732	-22.41*	-4.130	-1.741	-7.158	-22.97*
	(11.41)	(7.453)	(6.604)	(9.512)	(14.67)	(14.27)	(12.33)	(6.847)	(6.110)	(4.622)	(12.85)
Robust	-0.218	4.542	2.157	-15.41	4.383	-9.732	-22.41	-4.130	-1.741	-7.158	-22.97
	(13.39)	(9.146)	(7.953)	(11.07)	(16.78)	(16.09)	(15.05)	(7.904)	(6.803)	(5.255)	(15.57)
Observations	1,807	2,220	2,341	1,611	1,807	3,086	1,626	2,957	10,987	15,566	1,626
First-stage instrument t-statistic	13.785	37.034	42.004	18.379	13.079	17.082	23.236	22.144	25.972	44.256	22.651
Bandwidth $\gtrless 0$	0.319	0.539	0.837	0.525	0.261	0.216	0.216	0.435	0.279	0.259	0.209
Effective obs. < 0	147	487	582	325	145	246	231	413	902	1417	225
Effective obs. ≥ 0	206	369	363	280	170	228	216	467	1105	1536	212

Each column reports regression discontinuity results for a school or stacked sharp sample. Note, School R and the high cut-off sharp samples are identical. 'Bias-corrected' and 'Robust', but not 'Conventional' specification point estimates are bias-corrected for the selected bandwidth following Calonico et al. (2014b). Estimates are non-parametrically estimated using local linear regressions and a triangular kernel. Optimal bandwidths are calculated for each sharp sample based on the common-bandwidth MSE-optimal selector following Calonico et al. (2014b) and Calonico et al. (2016) as described in the text. Observations is the total number of observations available for estimation. The sum of effective observations on either side of the cut-off are those used for estimation. Standard errors are shown in parentheses. Only standard errors for the 'Robust' specification are robust and clustered on Primary FOEI. * significant at 10%, ** significant at 5%, *** significant at 1%.

A.6 Cohort dummies

Table A.6: Fuzzy regression discontinuity estimates of placebo effects of an offer on the 2008 test year dummy (non-parametric)

Specifications	A	B	C	D	E	F	G	H	I	J	K
Conventional	-0.0508	0.0116	0.0591	0.276**	0.171*	-0.162	0.164	-0.0709	-0.118	0.0809	0.000599
	(0.129)	(0.0957)	(0.137)	(0.130)	(0.0977)	(0.199)	(0.141)	(0.115)	(0.139)	(0.102)	(0.0963)
Bias-corrected	-0.0332	0.00160	0.0723	0.313**	0.183*	-0.196	0.208	-0.0359	-0.161	0.0811	-0.0228
	(0.129)	(0.0957)	(0.137)	(0.130)	(0.0977)	(0.199)	(0.141)	(0.115)	(0.139)	(0.102)	(0.0963)
Robust	-0.0332	0.00160	0.0723	0.313**	0.183*	-0.196	0.208	-0.0359	-0.161	0.0811	-0.0228
	(0.154)	(0.116)	(0.159)	(0.146)	(0.110)	(0.229)	(0.166)	(0.133)	(0.164)	(0.125)	(0.110)
Observations	721	1,280	1,052	1,639	1,588	1,179	1,455	1,774	2,342	1,533	2,804
First-stage instrument t-statistic	11.213	27.112	17.15	11.59	16.415	6.052	8.33	13.409	14.576	18.333	23.25
Bandwidth $\geqslant 0$	0.691	0.713	0.530	0.373	0.401	0.444	0.336	0.500	0.258	0.486	0.347
Effective obs. < 0	157	302	195	178	199	130	131	250	185	250	299
Effective obs. ≥ 0	180	279	207	208	232	146	198	264	171	283	321

Specifications	L	M	N	O	P	Q	R	Low	Mid-low	Mid-high	High
Conventional	0.0464	0.0247	-0.0184	-0.00154	0.0967	-0.0849	0.00266	0.00414	0.0842	0.00413	0.00937
	(0.150)	(0.0773)	(0.0549)	(0.0714)	(0.0952)	(0.125)	(0.0846)	(0.0744)	(0.0522)	(0.0411)	(0.0980)
Bias-corrected	0.0577	0.00115	-0.0586	0.0305	0.156	-0.0934	-0.0214	0.00926	0.0930*	0.00589	-0.0171
	(0.150)	(0.0773)	(0.0549)	(0.0714)	(0.0952)	(0.125)	(0.0846)	(0.0744)	(0.0522)	(0.0411)	(0.0980)
Robust	0.0577	0.00115	-0.0586	0.0305	0.156	-0.0934	-0.0214	0.00926	0.0930	0.00589	-0.0171
	(0.169)	(0.0961)	(0.0636)	(0.0852)	(0.109)	(0.139)	(0.105)	(0.0854)	(0.0583)	(0.0467)	(0.119)
Observations	1,857	2,286	2,416	1,661	1,868	3,195	1,658	3,046	11,493	16,067	1,658
First-stage instrument t-statistic	10.503	32.781	50.374	22.389	17.651	16.092	23.39	25.401	29.671	46.159	24.597
Bandwidth $\geqslant 0$	0.251	0.426	1.246	0.655	0.421	0.203	0.223	0.505	0.327	0.271	0.243
Effective obs. < 0	118	398	856	386	219	229	235	499	1080	1490	260
Effective obs. ≥ 0	168	331	379	309	259	216	217	539	1278	1579	221

Each column reports regression discontinuity results for a school or stacked sharp sample. Note, School R and the high cut-off sharp samples are identical. 'Bias-corrected' and 'Robust', but not 'Conventional' specification point estimates are bias-corrected for the selected bandwidth following Calonico et al. (2014b). Estimates are non-parametrically estimated using local linear regressions and a triangular kernel. Optimal bandwidths are calculated for each sharp sample based on the common-bandwidth MSE-optimal selector following Calonico et al. (2014b) and Calonico et al. (2016) as described in the text. Observations is the total number of observations available for estimation. The sum of effective observations on either side of the cut-off are those used for estimation. Standard errors are shown in parentheses. Only standard errors for the 'Robust' specification are robust and clustered on Primary FOEI. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table A.7: Fuzzy regression discontinuity estimates of placebo effects of an offer on the 2009 test year dummy (non-parametric)

Specifications	A	B	C	D	E	F	G	H	I	J	K
Conventional	0.106	0.0346	-0.177	-0.0872	-0.0155	0.0792	0.0414	0.0751	0.109	0.0692	-0.0456
	(0.125)	(0.104)	(0.117)	(0.129)	(0.106)	(0.215)	(0.137)	(0.107)	(0.122)	(0.103)	(0.0876)
Bias-corrected	0.0997	0.0583	-0.222*	-0.119	-0.00665	0.0740	0.0808	0.0681	0.120	0.154	-0.0415
	(0.125)	(0.104)	(0.117)	(0.129)	(0.106)	(0.215)	(0.137)	(0.107)	(0.122)	(0.103)	(0.0876)
Robust	0.0997	0.0583	-0.222	-0.119	-0.00665	0.0740	0.0808	0.0681	0.120	0.154	-0.0415
	(0.147)	(0.124)	(0.140)	(0.150)	(0.121)	(0.243)	(0.160)	(0.134)	(0.149)	(0.130)	(0.102)
Observations	721	1,280	1,052	1,639	1,588	1,179	1,455	1,774	2,342	1,533	2,804
First-stage instrument t-statistic	11.073	26.182	20.704	10.166	17.217	6.184	10.287	12.809	15.327	19.51	22.207
Bandwidth ⩾ 0	0.696	0.684	0.669	0.336	0.427	0.438	0.439	0.507	0.314	0.530	0.337
Effective obs. < 0	161	294	247	154	209	129	180	258	230	275	297
Effective obs. ≥ 0	182	273	247	187	237	145	251	266	205	303	318

Specifications	L	M	N	O	P	Q	R	Low	Mid-low	Mid-high	High
Conventional	0.0242	-0.0554	-0.0704	0.000638	0.0374	0.0966	-0.00918	-0.0180	0.0354	-0.0323	-0.00473
	(0.143)	(0.0647)	(0.0575)	(0.0916)	(0.115)	(0.110)	(0.0868)	(0.0796)	(0.0525)	(0.0443)	(0.0816)
Bias-corrected	0.0536	-0.0712	-0.0990*	-0.0193	0.0338	0.118	-0.00536	-0.0146	0.0505	-0.0354	-0.000797
	(0.143)	(0.0647)	(0.0575)	(0.0916)	(0.115)	(0.110)	(0.0868)	(0.0796)	(0.0525)	(0.0443)	(0.0816)
Robust	0.0536	-0.0712	-0.0990	-0.0193	0.0338	0.118	-0.00536	-0.0146	0.0505	-0.0354	-0.000797
	(0.160)	(0.0816)	(0.0716)	(0.110)	(0.133)	(0.125)	(0.109)	(0.0931)	(0.0580)	(0.0511)	(0.101)
Observations	1,857	2,286	2,416	1,661	1,868	3,195	1,658	3,046	11,493	16,067	1,658
First-stage instrument t-statistic	10.956	37.788	43.887	17.778	13.836	17.43	27.908	22.53	30.541	42.71	28.74
Bandwidth ⩾ 0	0.256	0.558	0.900	0.494	0.281	0.220	0.266	0.446	0.334	0.247	0.290
Effective obs. < 0	119	501	644	314	153	251	285	434	1108	1356	306
Effective obs. ≥ 0	171	379	367	260	180	233	233	486	1303	1492	243

Each column reports regression discontinuity results for a school or stacked sharp sample. Note, School R and the high cut-off sharp samples are identical. 'Bias-corrected' and 'Robust', but not 'Conventional' specification point estimates are bias-corrected for the selected bandwidth following Calonico et al. (2014b). Estimates are non-parametrically estimated using local linear regressions and a triangular kernel. Optimal bandwidths are calculated for each sharp sample based on the common-bandwidth MSE-optimal selector following Calonico et al. (2014b) and Calonico et al. (2016) as described in the text. Observations is the total number of observations available for estimation. The sum of effective observations on either side of the cut-off are those used for estimation. Standard errors are shown in parentheses. Only standard errors for the 'Robust' specification are robust and clustered on Primary FOEI. * significant at 10%, ** significant at 5%, *** significant at 1%.

Appendix B

Density discontinuity tests for all outcome variables and subsamples

Table B.1: Forcing variable manipulation testing based on density discontinuity following Cattaneo et al. (2016b) in analysis samples for all outcome variables

School/stack	Tertiary entrance score	Year 9 NAPLAN Reading score	Year 9 NAPLAN Numeracy score	Mathematics (2 unit) score	Mathematics Extension 1 score	English Advanced score	English Extension 1 score	Chemistry score	Physics score	Biology score	HSC course participation variables
A	.085	.079	.072	-.069	.134	.299**	.666**	.248	.270	.316*	.081
	(.123)	(.108)	(.112)	(.196)	(.230)	(.124)	(.324)	(.156)	(.168)	(.191)	(.122)
B	.013	-.054	-.073	-.221	.360	-.099	.349	-.176	.086	-.016	.014
	(.086)	(.080)	(.078)	(.192)	(.227)	(.128)	(.466)	(.198)	(.141)	(.158)	(.086)
C	.031	-.027	0	.138	.263*	-.017	-.060	-.122	-.043	.143	.032
	(.110)	(.104)	(.104)	(.179)	(.153)	(.148)	(.520)	(.182)	(.136)	(.158)	(.110)
D	-.062	-.028	-.013	-.040	-.090	-.024	-.420	-.286	-.187	.029	-.061
	(.098)	(.097)	(.098)	(.159)	(.242)	(.141)	(.298)	(.187)	(.201)	(.167)	(.098)
E	.170	.153	.154	.259	.474*	.274*	.449	.274	.198	.199	.167
	(.115)	(.113)	(.112)	(.170)	(.249)	(.157)	(.362)	(.233)	(.215)	(.216)	(.115)
F	.111	.154	.169*	.251*	.281	.185	.356	.341*	.255	-.090	.107
	(.097)	(.095)	(.095)	(.142)	(.240)	(.121)	(.248)	(.203)	(.191)	(.157)	(.096)
G	.152	.176*	.170*	.181	.143	.244	.760*	-.018	-.003	-.232	.153
	(.100)	(.096)	(.096)	(.113)	(.238)	(.178)	(.433)	(.212)	(.211)	(.209)	(.101)
H	.078	.089	.089	.213	.380	.084	.758**	-.053	.340	.084	.078
	(.099)	(.097)	(.095)	(.142)	(.296)	(.126)	(.297)	(.151)	(.265)	(.121)	(.099)
I	.137	.130	.125	.102	.307*	.106	.169	.146	.143	.155	.129
	(.089)	(.090)	(.087)	(.118)	(.173)	(.110)	(.269)	(.126)	(.149)	(.154)	(.089)
J	-.137	-.183	-.185	-.047	-.178	-.197	-.552	-.286	-.279	.444	-.131
	(.136)	(.127)	(.127)	(.178)	(.216)	(.181)	(.526)	(.240)	(.194)	(.321)	(.135)
K	.168**	.154**	.150**	.228**	.450*	.325***	.100	.080	.067	.275***	.169**
	(.067)	(.061)	(.061)	(.100)	(.231)	(.081)	(.286)	(.205)	(.138)	(.094)	(.067)
L	-.101	-.082	-.087	-.084	-.269	-.110	.138	-.001	-.417	-.058	-.100
	(.118)	(.109)	(.110)	(.154)	(.251)	(.142)	(.363)	(.198)	(.310)	(.163)	(.118)
M	.060	.096	.095	.024	.104	.073	.253	.058	.033	.025	.058
	(.129)	(.118)	(.119)	(.203)	(.231)	(.182)	(.343)	(.213)	(.204)	(.212)	(.129)
N	.058	.075	.066	.130	-.062	.166	.391	.053	.504	.140	.058
	(.118)	(.113)	(.113)	(.143)	(.238)	(.142)	(.287)	(.205)	(.370)	(.123)	(.118)
O	.020	.044	.044	.082	-.034	-.001	-.090	.150	-.018	.052	.018
	(.118)	(.111)	(.112)	(.181)	(.209)	(.149)	(.421)	(.147)	(.144)	(.250)	(.117)
P	-.045	-.025	-.027	.075	.038	-.049	.005	-.258	.069	.069	-.045
	(.110)	(.105)	(.105)	(.148)	(.257)	(.136)	(.295)	(.233)	(.304)	(.147)	(.110)
Q	-.134	-.136	-.136	-.106	-.194	-.169	-.349	-.008	-.031	.042	-.134
	(.112)	(.109)	(.110)	(.133)	(.224)	(.132)	(.317)	(.193)	(.228)	(.128)	(.112)
R	.311	.263	.258	.297	.444	.410	.416	.683*	.424*	.362	.310
	(.206)	(.193)	(.194)	(.253)	(.342)	(.254)	(.541)	(.351)	(.234)	(.235)	(.206)
Low	.019	-.030	-.059	.081	.158	-.022	.427**	-.043	.038	.186*	.018
	(.070)	(.064)	(.063)	(.133)	(.148)	(.096)	(.210)	(.126)	(.105)	(.104)	(.070)
Mid-low	.074*	.077*	.077*	.159***	.154*	.160***	.314***	-.014	.043	.120***	.073*
	(.041)	(.040)	(.039)	(.051)	(.092)	(.039)	(.120)	(.058)	(.052)	(.043)	(.041)
Mid-high	.098***	.095***	.096***	.094*	.216***	.137***	.156	.066	.059	.133***	.099***
	(.035)	(.033)	(.033)	(.050)	(.060)	(.044)	(.116)	(.058)	(.068)	(.051)	(.035)
High	.311	.263	.258	.297	.444	.410	.416	.683*	.424*	.362	.310
	(.206)	(.193)	(.194)	(.253)	(.342)	(.254)	(.541)	(.351)	(.234)	(.235)	(.206)

Cattaneo et al. (2016b) test estimates are differences in local, order 2 polynomial density estimators using a triangular kernel, in a bandwidth selected by a common MSE-optimal bandwdth selector. Estimates are for the unrestricted CJM test as described in the text. Jackknifed standard errors in parentheses. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table B.2: Forcing variable manipulation testing based on density discontinuity following Cattaneo et al. (2016b) in subsamples

Subsample	Stacked sharp sample			
	Low	Mid-low	Mid-high	High
High achievers (above median)	.104	.248***	.172***	.472
	(.136)	(.093)	(.055)	(.300)
High achievers (above third quartile)	.041	.329**	.136	.338
	(.156)	(.132)	(.090)	(.352)
Low achievers (below median)	.018	.077	-.007	
	(.076)	(.048)	(.020)	
Language background other than English	.249*	.069	.149***	.330
	(.134)	(.054)	(.037)	(.241)
Language background is English	-.021	.080*	.026	
	(.081)	(.043)	(.062)	
Male	-.047	.118**	.108**	.196
	(.093)	(.050)	(.054)	(.246)
Female	.055	.025	.125***	.305
	(.094)	(.055)	(.039)	(.304)
From advantaged primary school (above FOEI state average)	.121	.084*	.109***	.392*
	(.081)	(.044)	(.039)	(.235)
From disadvantaged primary school (below FOEI state average)	-.281**	.118	.052	-.098
	(.134)	(.095)	(.080)	(.357)
From advantaged primary school (above FOEI sample average)	.227*	.127***	.041	.258
	(.130)	(.044)	(.046)	(.236)
From disadvantaged primary school (below FOEI sample average)	-.091	.028	.150***	.131
	(.075)	(.056)	(.054)	(.310)

Cattaneo et al. (2016b) test estimates are differences in local, order 2 polynomial density estimators using a triangular kernel, in a bandwidth selected by a common MSE-optimal bandwdth selector. Estimates are for the unrestricted CJM test as described in the text. Jackknifed standard errors in parentheses. * significant at 10%, ** significant at 5%, *** significant at 1%.

Appendix C

Main non-parametric regression results and graphs

C.1 TES

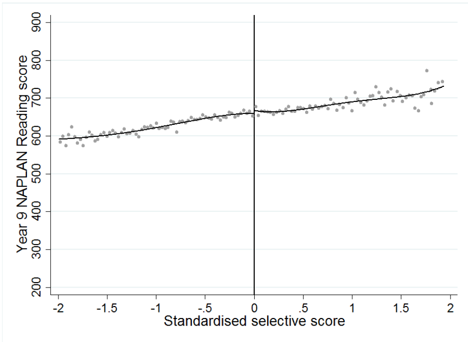
Table C.1: Fuzzy regression discontinuity estimates of effects of an offer on TES

Specifications	A	B	C	D	E	F	G	H	I	J	K
Conventional	6.680	-7.798	40.73**	3.639	-7.832	47.53**	-3.260	-14.23	-11.65	2.817	-16.12
	(14.17)	(14.83)	(19.91)	(13.63)	(9.083)	(19.72)	(16.62)	(10.68)	(13.58)	(14.93)	(10.16)
Bias-corrected	9.728	-14.57	43.39**	2.491	-11.76	52.69***	-6.221	-19.09*	-16.99	1.218	-20.37**
	(14.17)	(14.83)	(19.91)	(13.63)	(9.083)	(19.72)	(16.62)	(10.68)	(13.58)	(14.93)	(10.16)
Robust	9.728	-14.57	43.39*	2.491	-11.76	52.69**	-6.221	-19.09	-16.99	1.218	-20.37*
	(16.73)	(17.15)	(23.52)	(15.51)	(10.30)	(22.35)	(19.62)	(12.83)	(15.68)	(17.97)	(12.28)
Observations	697	1,234	1,022	1,511	1,533	1,098	1,408	1,728	2,222	1,494	2,711
First-stage instrument t-statistic	12.877	19.969	15.518	11.374	18.339	6.368	7.891	11.99	14.767	16.031	21.773
Bandwidth $\gtrless 0$	0.861	0.380	0.495	0.392	0.464	0.444	0.338	0.428	0.196	0.397	0.305
Effective obs. < 0	190	153	174	177	220	121	131	211	139	209	268
Effective obs. ≥ 0	207	166	190	214	251	146	199	236	137	239	299

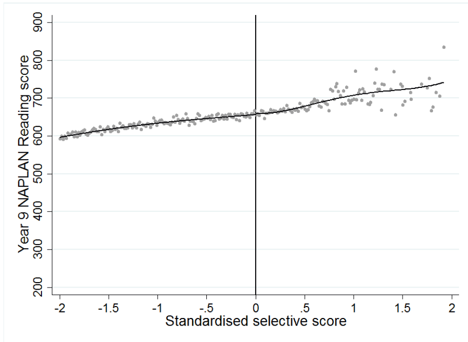
Specifications	L	M	N	O	P	Q	R	Low	Mid-low	Mid-high	High
Conventional	7.047	9.327	-1.694	5.600	-8.297	19.86**	12.73*	19.04*	2.278	0.283	12.73*
	(8.162)	(6.719)	(8.625)	(9.782)	(8.846)	(8.231)	(7.324)	(11.12)	(4.577)	(4.672)	(7.324)
Bias-corrected	7.454	9.114	-1.959	8.565	-13.65	22.12***	13.45*	16.89	0.379	-0.378	13.45*
	(8.162)	(6.719)	(8.625)	(9.782)	(8.846)	(8.231)	(7.324)	(11.12)	(4.577)	(4.672)	(7.324)
Robust	7.454	9.114	-1.959	8.565	-13.65	22.12**	13.45	16.89	0.379	-0.378	13.45
	(9.638)	(7.948)	(10.91)	(12.26)	(10.60)	(9.266)	(9.105)	(13.34)	(5.500)	(5.264)	(9.105)
Observations	1,805	2,216	2,340	1,610	1,805	3,081	1,623	2,945	10,977	15,548	1,623
First-stage instrument t-statistic	21.293	36.536	24.512	16.251	17.465	18.571	23.062	23.872	34.993	39.649	23.062
Bandwidth $\gtrless 0$	0.471	0.517	0.318	0.413	0.427	0.226	0.221	0.470	0.468	0.220	0.221
Effective obs. < 0	243	474	244	271	221	264	234	441	1580	1200	234
Effective obs. ≥ 0	247	363	268	239	262	241	215	499	1665	1350	215

Each column reports regression discontinuity results for a school or stacked sharp sample. Note, School R and the high cut-off sharp samples are identical. 'Bias-corrected' and 'Robust', but not 'Conventional' specification point estimates are bias-corrected for the selected bandwidth following Calonico et al. (2014b). Estimates are non-parametrically estimated using local linear regressions and a triangular kernel. Optimal bandwidths are calculated for each sharp sample based on the common-bandwidth MSE-optimal selector following Calonico et al. (2014b) and Calonico et al. (2016) as described in the text. Observations is the total number of observations available for estimation. The sum of effective observations on either side of the cut-off are those used for estimation. Standard errors are shown in parentheses. Only standard errors for the 'Robust' specification are robust and clustered on Primary FOEI. * significant at 10%, ** significant at 5%, *** significant at 1%.

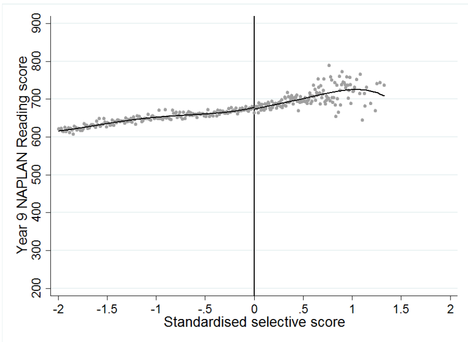
C.2 Year 9 NAPLAN Reading and Numeracy



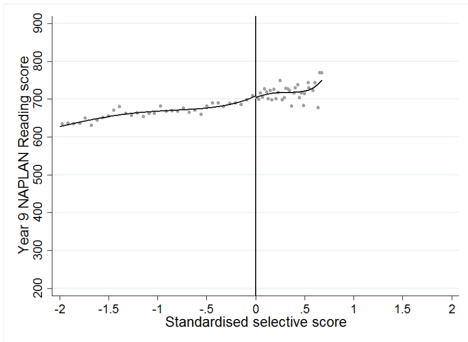
(a) Low cut-off



(b) Mid-low cut-off

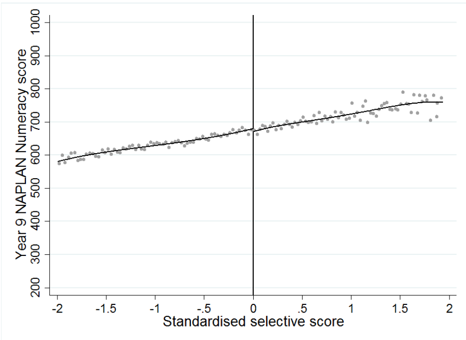


(c) Mid-high cut-off

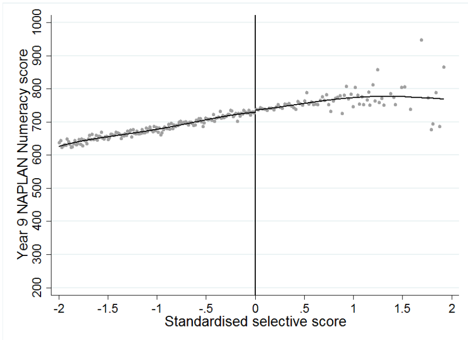


(d) High cut-off

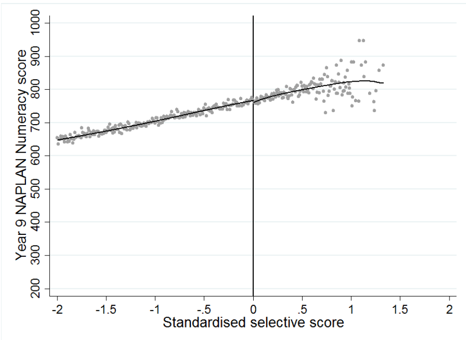
Figure C.1: Individual Year 9 NAPLAN Reading scores in stacked sharp samples



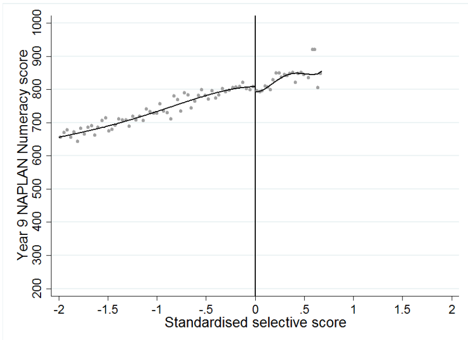
(a) Low cut-off



(b) Mid-low cut-off



(c) Mid-high cut-off



(d) High cut-off

Figure C.2: Individual Year 9 NAPLAN Numeracy scores in stacked sharp samples

Table C.2: Fuzzy regression discontinuity estimates of effects of an offer on Year 9 NAPLAN Reading scores

Specifications	A	B	C	D	E	F	G	H	I	J	K
Conventional	-8.401	-6.658	13.32*	-10.07	-15.77*	4.326	13.10	-1.112	6.294	-12.12	0.546
	(10.97)	(6.740)	(7.558)	(10.48)	(8.112)	(10.57)	(10.93)	(8.703)	(9.824)	(10.68)	(6.173)
Bias-corrected	-7.263	-6.954	18.27**	-10.62	-17.74**	7.376	15.86	-3.812	3.182	-16.57	-1.430
	(10.97)	(6.740)	(7.558)	(10.48)	(8.112)	(10.57)	(10.93)	(8.703)	(9.824)	(10.68)	(6.173)
Robust	-7.263	-6.954	18.27**	-10.62	-17.74*	7.376	15.86	-3.812	3.182	-16.57	-1.430
	(12.73)	(7.965)	(9.056)	(12.28)	(9.220)	(11.88)	(12.74)	(10.51)	(11.60)	(12.84)	(7.703)
Observations	760	1,401	1,193	1,643	1,630	1,177	1,494	1,832	2,340	1,594	2,878
First-stage instrument t-statistic	10.76	26.446	22.76	12.775	21.026	8.367	7.811	13.318	15.592	14.013	24.483
Bandwidth $\gtrless 0$	0.650	0.623	0.669	0.389	0.499	0.560	0.336	0.477	0.294	0.316	0.353
Effective obs. < 0	148	289	277	190	264	185	137	239	210	167	309
Effective obs. ≥ 0	180	249	270	218	266	183	200	263	189	201	323

Specifications	L	M	N	O	P	Q	R	Low	Mid-low	Mid-high	High
Conventional	-3.443	-5.798	-6.219	-6.875	-0.665	-8.472	-11.86	2.537	-2.243	-4.854	-11.86
	(11.87)	(6.327)	(7.234)	(7.503)	(9.462)	(9.781)	(8.832)	(5.413)	(3.869)	(3.979)	(8.832)
Bias-corrected	-2.144	-4.864	-6.555	-9.696	1.653	-9.393	-16.34*	4.237	-2.709	-4.590	-16.34*
	(11.87)	(6.327)	(7.234)	(7.503)	(9.462)	(9.781)	(8.832)	(5.413)	(3.869)	(3.979)	(8.832)
Robust	-2.144	-4.864	-6.555	-9.696	1.653	-9.393	-16.34	4.237	-2.709	-4.590	-16.34
	(13.23)	(7.822)	(8.577)	(9.242)	(10.97)	(10.95)	(10.71)	(6.144)	(4.307)	(4.528)	(10.71)
Observations	1,952	2,338	2,493	1,703	1,933	3,246	1,687	3,346	11,692	16,523	1,687
First-stage instrument t-statistic	10.105	32.46	28.071	20.082	17.01	17.757	24.802	27.992	34.804	40.116	24.802
Bandwidth $\gtrless 0$	0.212	0.417	0.390	0.540	0.404	0.211	0.210	0.493	0.394	0.218	0.210
Effective obs. < 0	104	399	312	337	219	241	228	534	1379	1216	228
Effective obs. ≥ 0	155	330	294	282	256	229	212	552	1502	1355	212

Each column reports regression discontinuity results for a school or stacked sharp sample. Note, School R and the high cut-off sharp samples are identical. 'Bias-corrected' and 'Robust', but not 'Conventional' specification point estimates are bias-corrected for the selected bandwidth following Calonico et al. (2014b). Estimates are non-parametrically estimated using local linear regressions and a triangular kernel. Optimal bandwidths are calculated for each sharp sample based on the common-bandwidth MSE-optimal selector following Calonico et al. (2014b) and Calonico et al. (2016) as described in the text. Observations is the total number of observations available for estimation. The sum of effective observations on either side of the cut-off are those used for estimation. Standard errors are shown in parentheses. Only standard errors for the 'Robust' specification are robust and clustered on Primary FOEI. * significant at 10%, ** significant at 5%, *** significant at 1%.

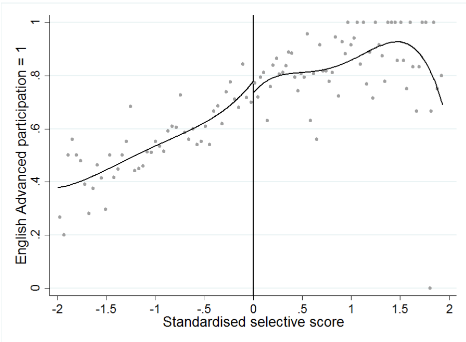
Table C.3: Fuzzy regression discontinuity estimates of effects of an offer on Year 9 NAPLAN Numeracy scores

Specifications	A	B	C	D	E	F	G	H	I	J	K
Conventional	-1.175	0.0380	-1.736	10.24	5.552	-5.180	0.184	-7.000	32.21***	12.67	-18.56**
	(9.009)	(6.595)	(6.361)	(9.899)	(9.654)	(14.27)	(12.65)	(10.31)	(11.61)	(9.234)	(8.033)
Bias-corrected	1.266	-0.0730	-3.669	9.384	6.237	-5.687	4.235	-9.032	36.88***	17.02*	-21.63***
	(9.009)	(6.595)	(6.361)	(9.899)	(9.654)	(14.27)	(12.65)	(10.31)	(11.61)	(9.234)	(8.033)
Robust	1.266	-0.0730	-3.669	9.384	6.237	-5.687	4.235	-9.032	36.88***	17.02	-21.63**
	(10.73)	(7.686)	(7.902)	(11.19)	(11.08)	(16.35)	(15.02)	(11.98)	(13.44)	(11.54)	(9.317)
Observations	758	1,389	1,175	1,637	1,622	1,168	1,487	1,822	2,334	1,589	2,876
First-stage instrument t-statistic	12.744	24.171	34.99	11.452	17.799	6.291	9.093	13.675	15.575	19.01	24.344
Bandwidth $\gtrless 0$	0.732	0.586	1.150	0.343	0.420	0.418	0.415	0.485	0.262	0.483	0.345
Effective obs. < 0	174	273	451	160	220	115	175	245	187	255	302
Effective obs. ≥ 0	197	234	371	194	237	142	241	263	171	279	319

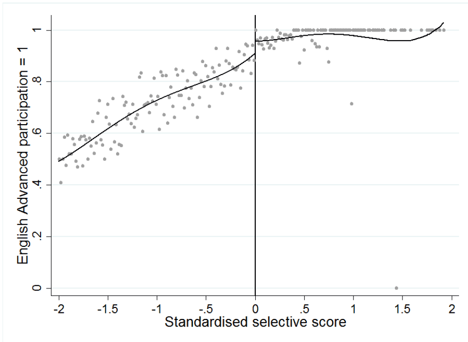
Specifications	L	M	N	O	P	Q	R	Low	Mid-low	Mid-high	High
Conventional	5.772	6.566	-12.16	8.888	-3.203	6.162	-10.97	-2.566	10.37**	-4.093	-10.97
	(7.076)	(6.707)	(8.919)	(10.18)	(7.647)	(7.717)	(9.843)	(5.589)	(5.106)	(4.186)	(9.843)
Bias-corrected	2.755	10.44	-14.02	7.314	-5.132	7.634	-6.836	-2.662	11.69**	-4.701	-6.836
	(7.076)	(6.707)	(8.919)	(10.18)	(7.647)	(7.717)	(9.843)	(5.589)	(5.106)	(4.186)	(9.843)
Robust	2.755	10.44	-14.02	7.314	-5.132	7.634	-6.836	-2.662	11.69**	-4.701	-6.836
	(8.340)	(8.206)	(10.93)	(12.07)	(9.089)	(8.731)	(11.88)	(6.430)	(5.652)	(4.736)	(11.88)
Observations	1,941	2,332	2,484	1,693	1,921	3,238	1,683	3,314	11,641	16,465	1,683
First-stage instrument t-statistic	20.573	33.04	27.172	18.773	17.469	17.956	23.548	27.935	28.594	39.837	23.548
Bandwidth $\gtrless 0$	0.432	0.425	0.371	0.495	0.424	0.214	0.215	0.508	0.306	0.216	0.215
Effective obs. < 0	229	404	294	315	229	245	230	548	1029	1200	230
Effective obs. ≥ 0	248	333	288	261	262	231	215	558	1213	1344	215

Each column reports regression discontinuity results for a school or stacked sharp sample. Note, School R and the high cut-off sharp samples are identical. 'Bias-corrected' and 'Robust', but not 'Conventional' specification point estimates are bias-corrected for the selected bandwidth following Calonico et al. (2014b). Estimates are non-parametrically estimated using local linear regressions and a triangular kernel. Optimal bandwidths are calculated for each sharp sample based on the common-bandwidth MSE-optimal selector following Calonico et al. (2014b) and Calonico et al. (2016) as described in the text. Observations is the total number of observations available for estimation. The sum of effective observations on either side of the cut-off are those used for estimation. Standard errors are shown in parentheses. Only standard errors for the 'Robust' specification are robust and clustered on Primary FOEI. * significant at 10%, ** significant at 5%, *** significant at 1%.

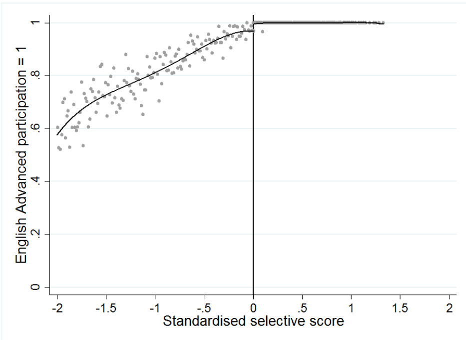
C.3 HSC English participation and achievement



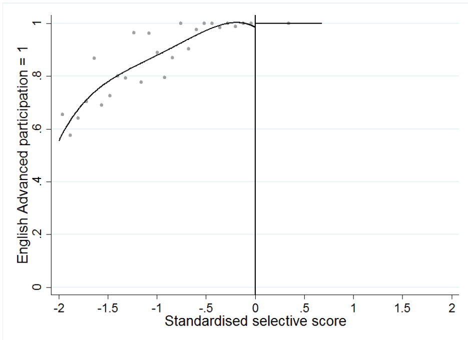
(a) Low cut-off



(b) Mid-low cut-off



(c) Mid-high cut-off



(d) High cut-off

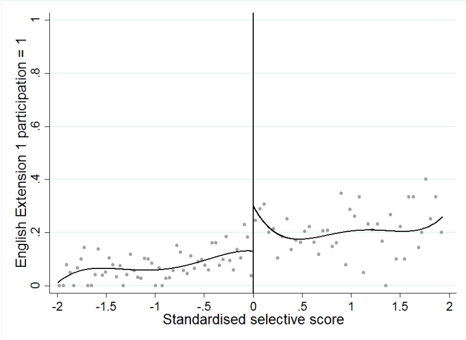
Figure C.3: English Advanced participation in stacked sharp samples

Table C.4: Fuzzy regression discontinuity estimates of effects of an offer on English Advanced participation

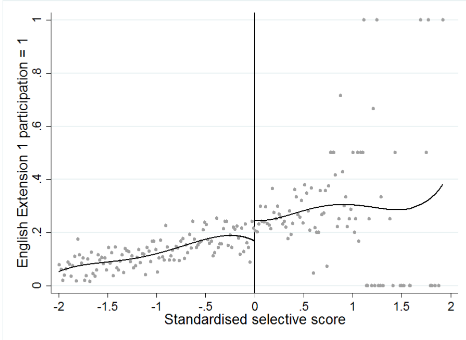
Specifications	A	B	C	D	E	F	G	H	I	J	K
Conventional	0.0914	-0.280**	0.0595	0.0609	0.0792	0.291***	0.0312	-0.0203	0.0492	0.112**	0.0956***
	(0.107)	(0.114)	(0.119)	(0.0658)	(0.0632)	(0.104)	(0.0902)	(0.0545)	(0.0401)	(0.0458)	(0.0199)
Bias-corrected	0.0724	-0.318***	0.0829	0.0546	0.0767	0.337***	0.0521	-0.0132	0.0497	0.132***	0.104***
	(0.107)	(0.114)	(0.119)	(0.0658)	(0.0632)	(0.104)	(0.0902)	(0.0545)	(0.0401)	(0.0458)	(0.0199)
Robust	0.0724	-0.318**	0.0829	0.0546	0.0767	0.337***	0.0521	-0.0132	0.0497	0.132**	0.104***
	(0.123)	(0.130)	(0.137)	(0.0750)	(0.0709)	(0.125)	(0.105)	(0.0649)	(0.0482)	(0.0556)	(0.0233)
Observations	700	1,236	1,023	1,512	1,535	1,100	1,409	1,731	2,224	1,497	2,714
First-stage instrument t-statistic	11.795	19.065	14.708	12.652	15.253	8.945	10.005	12.74	15.822	20.097	52.725
Bandwidth $\gtrless 0$	0.742	0.341	0.437	0.440	0.373	0.720	0.458	0.467	0.267	0.527	1.320
Effective obs. < 0	164	142	148	205	183	228	182	230	192	273	1076
Effective obs. ≥ 0	188	151	164	240	216	224	252	248	172	298	378

Specifications	L	M	N	O	P	Q	R	Low	Mid-low	Mid-high	High
Conventional	0.0137	-0.00488	0.00470	-0.0124	-0.00753	0.00212	-0.00250	-0.0648	0.0646**	0.0202*	-0.00250
	(0.0268)	(0.0108)	(0.0267)	(0.00859)	(0.00637)	(0.00957)	(0.00298)	(0.0694)	(0.0256)	(0.0105)	(0.00298)
Bias-corrected	0.0107	-0.00707	0.00410	-0.00579	-0.00713	0.0137	-0.00311	-0.0861	0.0662***	0.0210**	-0.00311
	(0.0268)	(0.0108)	(0.0267)	(0.00859)	(0.00637)	(0.00957)	(0.00298)	(0.0694)	(0.0256)	(0.0105)	(0.00298)
Robust	0.0107	-0.00707	0.00410	-0.00579	-0.00713	0.0137	-0.00311	-0.0861	0.0662**	0.0210*	-0.00311
	(0.0302)	(0.0127)	(0.0327)	(0.0116)	(0.00942)	(0.0117)	(0.00371)	(0.0797)	(0.0282)	(0.0121)	(0.00371)
Observations	1,808	2,221	2,343	1,615	1,807	3,082	1,624	2,951	10,991	15,570	1,624
First-stage instrument t-statistic	13.855	45.732	32.951	24.665	24.756	39.293	31.698	24.45	35.159	45.974	31.698
Bandwidth $\gtrless 0$	0.308	0.752	0.531	0.766	0.740	0.674	0.275	0.451	0.419	0.277	0.275
Effective obs. < 0	144	623	394	426	371	758	293	424	1415	1510	293
Effective obs. ≥ 0	196	431	331	325	340	430	237	477	1543	1600	237

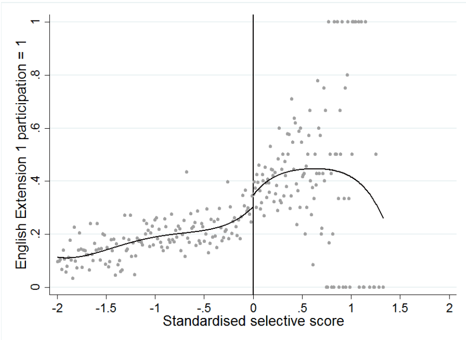
Each column reports regression discontinuity results for a school or stacked sharp sample. Note, School R and the high cut-off sharp samples are identical. 'Bias-corrected' and 'Robust', but not 'Conventional' specification point estimates are bias-corrected for the selected bandwidth following Calonico et al. (2014b). Estimates are non-parametrically estimated using local linear regressions and a triangular kernel. Optimal bandwidths are calculated for each sharp sample based on the common-bandwidth MSE-optimal selector following Calonico et al. (2014b) and Calonico et al. (2016) as described in the text. Observations is the total number of observations available for estimation. The sum of effective observations on either side of the cut-off are those used for estimation. Standard errors are shown in parentheses. Only standard errors for the 'Robust' specification are robust and clustered on Primary FOEI. * significant at 10%, ** significant at 5%, *** significant at 1%.



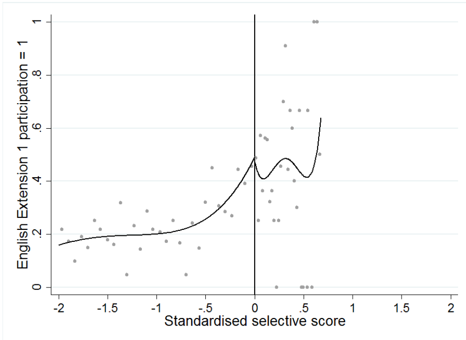
(a) Low cut-off



(b) Mid-low cut-off

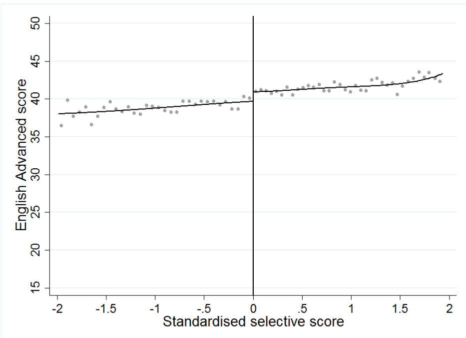


(c) Mid-high cut-off

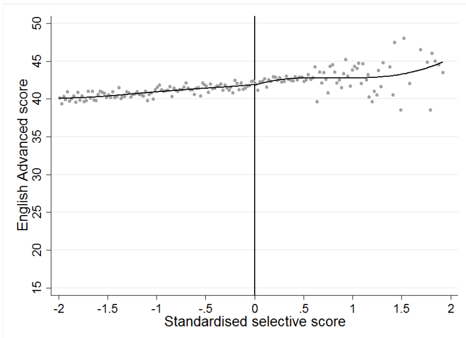


(d) High cut-off

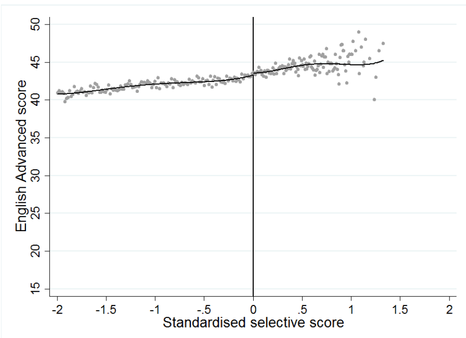
Figure C.4: English Extension 1 participation in stacked sharp samples



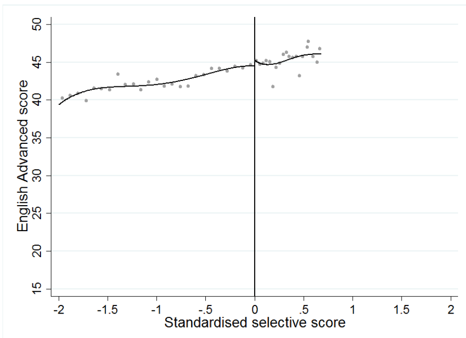
(a) Low cut-off



(b) Mid-low cut-off



(c) Mid-high cut-off



(d) High cut-off

Figure C.5: HSC English Advanced score in stacked sharp samples

Table C.5: Fuzzy regression discontinuity estimates of effects of an offer on English Extension 1 participation

Specifications	A	B	C	D	E	F	G	H	I	J	K
Conventional	0.132	0.136*	0.145*	-0.0536	0.0571	0.117	0.0221	0.127	-0.123	0.0743	0.00661
	(0.111)	(0.0794)	(0.0778)	(0.0962)	(0.105)	(0.153)	(0.0717)	(0.0824)	(0.107)	(0.0794)	(0.0646)
Bias-corrected	0.152	0.137*	0.169**	-0.0625	0.0626	0.141	0.00685	0.180**	-0.165	0.0426	0.0142
	(0.111)	(0.0794)	(0.0778)	(0.0962)	(0.105)	(0.153)	(0.0717)	(0.0824)	(0.107)	(0.0794)	(0.0646)
Robust	0.152	0.137	0.169*	-0.0625	0.0626	0.141	0.00685	0.180*	-0.165	0.0426	0.0142
	(0.131)	(0.0949)	(0.0897)	(0.116)	(0.122)	(0.173)	(0.0830)	(0.102)	(0.129)	(0.0933)	(0.0779)
Observations	700	1,236	1,023	1,512	1,535	1,100	1,409	1,731	2,224	1,497	2,714
First-stage instrument t-statistic	10.191	22.466	15.812	10.356	15.44	6.688	12.009	14.603	15.286	16.34	23.46
Bandwidth $\gtrless 0$	0.658	0.551	0.485	0.407	0.382	0.460	0.556	0.600	0.226	0.387	0.338
Effective obs. < 0	140	222	171	186	185	127	220	290	161	207	294
Effective obs. ≥ 0	173	223	184	220	222	152	268	292	155	234	317

Specifications	L	M	N	O	P	Q	R	Low	Mid-low	Mid-high	High
Conventional	0.256**	0.0149	0.110*	0.174*	0.0725	-0.0932	-0.0993	0.155***	0.0729**	0.0305	-0.0993
	(0.111)	(0.0529)	(0.0659)	(0.0957)	(0.0883)	(0.0865)	(0.0897)	(0.0562)	(0.0364)	(0.0387)	(0.0897)
Bias-corrected	0.271**	0.00684	0.0626	0.188**	0.0480	-0.102	-0.119	0.179***	0.0721**	0.0297	-0.119
	(0.111)	(0.0529)	(0.0659)	(0.0957)	(0.0883)	(0.0865)	(0.0897)	(0.0562)	(0.0364)	(0.0387)	(0.0897)
Robust	0.271**	0.00684	0.0626	0.188	0.0480	-0.102	-0.119	0.179***	0.0721*	0.0297	-0.119
	(0.128)	(0.0640)	(0.0746)	(0.121)	(0.104)	(0.0965)	(0.109)	(0.0668)	(0.0423)	(0.0438)	(0.109)
Observations	1,808	2,221	2,343	1,615	1,807	3,082	1,624	2,951	10,991	15,570	1,624
First-stage instrument t-statistic	12.032	35.073	36.45	16.636	17.31	18.024	26.549	25.096	33.372	41.009	26.549
Bandwidth $\gtrless 0$	0.270	0.477	0.569	0.436	0.415	0.217	0.255	0.520	0.422	0.232	0.255
Effective obs. < 0	122	439	414	286	214	247	274	501	1425	1266	274
Effective obs. ≥ 0	177	348	340	246	256	229	225	544	1552	1408	225

Each column reports regression discontinuity results for a school or stacked sharp sample. Note, School R and the high cut-off sharp samples are identical. 'Bias-corrected' and 'Robust', but not 'Conventional' specification point estimates are bias-corrected for the selected bandwidth following Calonico et al. (2014b). Estimates are non-parametrically estimated using local linear regressions and a triangular kernel. Optimal bandwidths are calculated for each sharp sample based on the common-bandwidth MSE-optimal selector following Calonico et al. (2014b) and Calonico et al. (2016) as described in the text. Observations is the total number of observations available for estimation. The sum of effective observations on either side of the cut-off are those used for estimation. Standard errors are shown in parentheses. Only standard errors for the 'Robust' specification are robust and clustered on Primary FOEI. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table C.6: Fuzzy regression discontinuity estimates of effects of an offer on English Advanced scores

Specifications	A	B	C	D	E	F	G	H	I	J	K
Conventional	1.469	1.552**	0.479	0.427	-0.210	0.675	0.483	-0.775	-1.097	0.0854	-0.690
	(1.154)	(0.758)	(1.062)	(1.179)	(0.603)	(1.200)	(0.970)	(0.590)	(0.691)	(0.774)	(0.596)
Bias-corrected	1.804	1.392*	0.597	0.250	-0.241	0.764	0.172	-1.106*	-1.237*	-0.286	-0.783
	(1.154)	(0.758)	(1.062)	(1.179)	(0.603)	(1.200)	(0.970)	(0.590)	(0.691)	(0.774)	(0.596)
Robust	1.804	1.392	0.597	0.250	-0.241	0.764	0.172	-1.106	-1.237	-0.286	-0.783
	(1.374)	(0.901)	(1.281)	(1.337)	(0.695)	(1.386)	(1.153)	(0.702)	(0.820)	(0.916)	(0.690)
Observations	476	774	675	1,019	1,095	849	858	1,331	1,564	1,109	1,909
First-stage instrument t-statistic	6.306	20.35	30.791	8.340	18.646	5.520	7.453	12.313	15.959	14.743	16.775
Bandwidth $\gtrless 0$	0.641	0.649	0.633	0.321	0.563	0.502	0.368	0.448	0.276	0.377	0.249
Effective obs. < 0	81	181	151	114	224	118	122	207	177	172	183
Effective obs. ≥ 0	149	193	179	166	266	161	202	225	176	228	256

Specifications	L	M	N	O	P	Q	R	Low	Mid-low	Mid-high	High
Conventional	0.445	0.917**	0.405	0.574	-0.916**	0.910*	0.00130	0.898	-0.362	0.237	0.00130
	(0.611)	(0.451)	(0.508)	(0.534)	(0.458)	(0.472)	(0.577)	(0.643)	(0.347)	(0.268)	(0.577)
Bias-corrected	0.492	0.845*	0.452	0.888*	-1.232***	0.997**	-0.225	0.893	-0.494	0.244	-0.225
	(0.611)	(0.451)	(0.508)	(0.534)	(0.458)	(0.472)	(0.577)	(0.643)	(0.347)	(0.268)	(0.577)
Robust	0.492	0.845	0.452	0.888	-1.232**	0.997*	-0.225	0.893	-0.494	0.244	-0.225
	(0.715)	(0.555)	(0.630)	(0.674)	(0.539)	(0.546)	(0.665)	(0.740)	(0.387)	(0.302)	(0.665)
Observations	1,466	1,620	1,853	1,245	1,473	2,450	1,368	1,919	7,808	11,996	1,368
First-stage instrument t-statistic	13.566	28.316	23.761	16.501	17.312	21.809	17.777	20.972	27.855	36.25	17.777
Bandwidth $\gtrless 0$	0.307	0.334	0.330	0.427	0.422	0.281	0.144	0.531	0.351	0.201	0.144
Effective obs. < 0	139	304	238	272	212	328	148	346	990	1065	148
Effective obs. ≥ 0	195	283	273	245	259	296	163	445	1293	1267	163

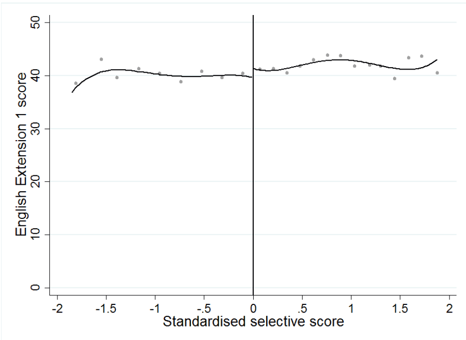
Each column reports regression discontinuity results for a school or stacked sharp sample. Note, School R and the high cut-off sharp samples are identical. 'Bias-corrected' and 'Robust', but not 'Conventional' specification point estimates are bias-corrected for the selected bandwidth following Calonico et al. (2014b). Estimates are non-parametrically estimated using local linear regressions and a triangular kernel. Optimal bandwidths are calculated for each sharp sample based on the common-bandwidth MSE-optimal selector following Calonico et al. (2014b) and Calonico et al. (2016) as described in the text. Observations is the total number of observations available for estimation. The sum of effective observations on either side of the cut-off are those used for estimation. Standard errors are shown in parentheses. Only standard errors for the 'Robust' specification are robust and clustered on Primary FOEI. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table C.7: Fuzzy regression discontinuity estimates of effects of an offer on English Extension 1 scores

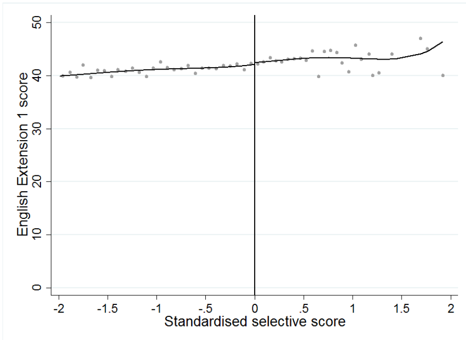
Specifications	A	B	D	E	F	G	H	I	J	K
Conventional	3.641	0.929	1.586	-0.222	0.996	0.970	-2.393**	0.418	-1.564	0.246
	(5.598)	(1.687)	(1.258)	(1.765)	(3.381)	(2.569)	(1.195)	(1.597)	(3.163)	(1.184)
Bias-corrected	1.658	0.743	2.183*	-0.0836	0.685	1.098	-2.924**	0.205	-2.695	-0.00375
	(5.598)	(1.687)	(1.258)	(1.765)	(3.381)	(2.569)	(1.195)	(1.597)	(3.163)	(1.184)
Robust	1.658	0.743	2.183	-0.0836	0.685	1.098	-2.924*	0.205	-2.695	-0.00375
	(6.657)	(1.967)	(1.449)	(2.148)	(3.677)	(2.913)	(1.493)	(1.824)	(3.723)	(1.494)
Observations	123	131	164	234	203	106	328	307	205	367
First-stage instrument t-statistic	1.605	18.42	4.995	4.79	1.093	1.491	7.235	7.66	2.57	20.995
Bandwidth $\gtrless 0$	0.292	0.492	0.587	0.352	0.576	0.324	0.314	0.234	0.252	0.438
Effective obs. < 0	10	23	37	37	43	18	40	28	15	63
Effective obs. ≥ 0	25	43	48	60	60	29	60	40	42	69

Specifications	L	M	N	O	P	Q	R	Low	Mid-low	Mid-high	High
Conventional	-0.344	1.516	0.321	-0.934	-0.522	1.839	0.559	1.870	-0.0497	0.213	0.559
	(1.031)	(1.459)	(0.886)	(1.472)	(0.718)	(1.228)	(0.676)	(1.629)	(0.799)	(0.615)	(0.676)
Bias-corrected	-0.386	1.453	0.248	-0.351	-0.682	1.978	0.772	2.337	-0.167	0.198	0.772
	(1.031)	(1.459)	(0.886)	(1.472)	(0.718)	(1.228)	(0.676)	(1.629)	(0.799)	(0.615)	(0.676)
Robust	-0.386	1.453	0.248	-0.351	-0.682	1.978	0.772	2.337	-0.167	0.198	0.772
	(1.263)	(1.791)	(1.059)	(1.728)	(0.846)	(1.455)	(0.822)	(2.008)	(0.902)	(0.705)	(0.822)
Observations	423	236	561	280	574	587	441	367	1,544	3,021	441
First-stage instrument t-statistic	6.843	N/A	10.737	5.718	13.802	10.355	14.45	5.844	8.218	22.097	14.45
Bandwidth $\gtrless 0$	0.410	0.402	0.292	0.390	0.404	0.242	0.200	0.479	0.353	0.239	0.200
Effective obs. < 0	60	43	76	48	102	90	94	53	210	351	94
Effective obs. ≥ 0	130	55	141	85	189	89	89	108	338	554	89

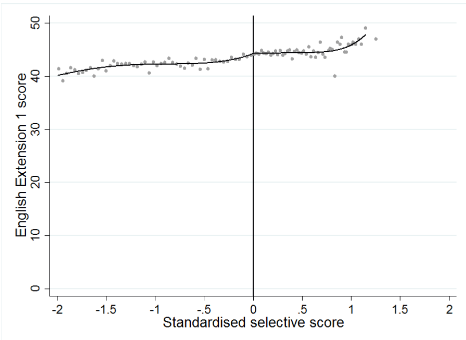
Each column reports regression discontinuity results for a school or stacked sharp sample. Note, School R and the high cut-off sharp samples are identical. 'Bias-corrected' and 'Robust', but not 'Conventional' specification point estimates are bias-corrected for the selected bandwidth following Calonico et al. (2014b). Estimates are non-parametrically estimated using local linear regressions and a triangular kernel. Optimal bandwidths are calculated for each sharp sample based on the common-bandwidth MSE-optimal selector following Calonico et al. (2014b) and Calonico et al. (2016) as described in the text. Observations is the total number of observations available for estimation. The sum of effective observations on either side of the cut-off are those used for estimation. Standard errors are shown in parentheses. Only standard errors for the 'Robust' specification are robust and clustered on Primary FOEI. * significant at 10%, ** significant at 5%, *** significant at 1%.



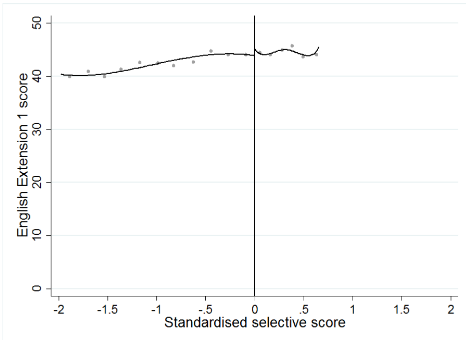
(a) Low cut-off



(b) Mid-low cut-off



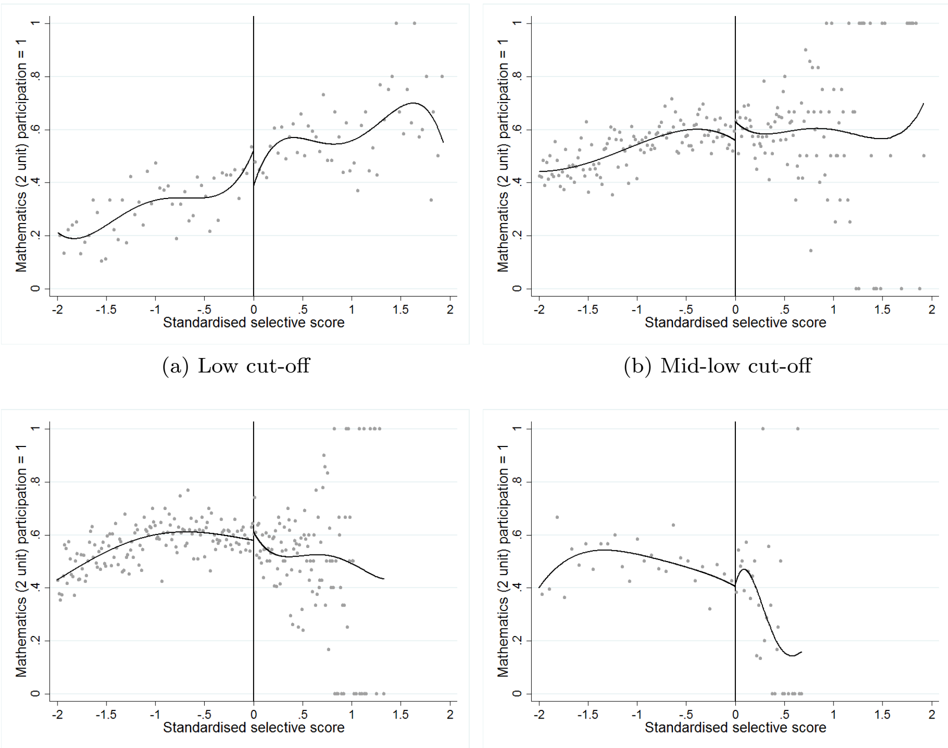
(c) Mid-high cut-off



(d) High cut-off

Figure C.6: HSC English Extension 1 score in stacked sharp samples

C.4 HSC Mathematics participation and achievement



(a) Low cut-off

(b) Mid-low cut-off

(c) Mid-high cut-off

(d) High cut-off

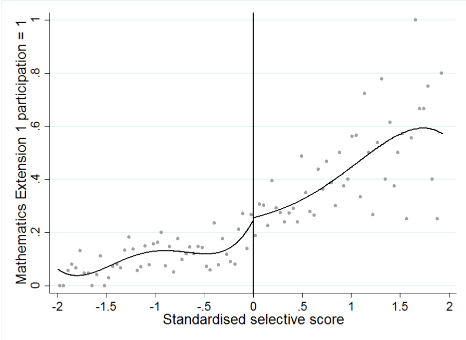
Figure C.7: Mathematics (2 Unit) participation in stacked sharp samples

Table C.8: Fuzzy regression discontinuity estimates of effects of an offer on Mathematics (2 unit) participation

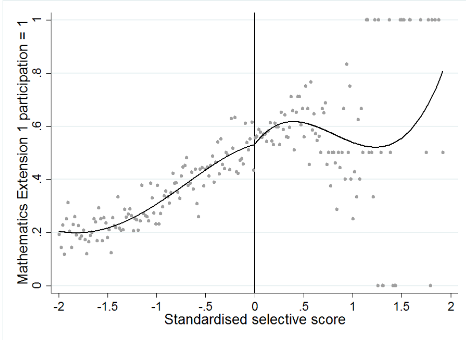
Specifications	A	B	C	D	E	F	G	H	I	J	K
Conventional	0.0796	-0.174	0.0861	0.110	0.0445	-0.0527	0.253	0.143	0.0585	0.194*	0.0999
	(0.128)	(0.129)	(0.102)	(0.128)	(0.124)	(0.146)	(0.156)	(0.112)	(0.128)	(0.0994)	(0.0989)
Bias-corrected	0.0724	-0.206	0.118	0.135	0.0796	-0.0830	0.351**	0.188*	0.0670	0.235**	0.0890
	(0.128)	(0.129)	(0.102)	(0.128)	(0.124)	(0.146)	(0.156)	(0.112)	(0.128)	(0.0994)	(0.0989)
Robust	0.0724	-0.206	0.118	0.135	0.0796	-0.0830	0.351*	0.188	0.0670	0.235**	0.0890
	(0.156)	(0.151)	(0.121)	(0.143)	(0.141)	(0.170)	(0.183)	(0.132)	(0.155)	(0.119)	(0.116)
Observations	700	1,236	1,023	1,512	1,535	1,100	1,409	1,731	2,224	1,497	2,714
First-stage instrument t-statistic	12.293	20.783	17.177	9.786	14.935	6.706	8.385	13.593	15.472	19.099	22.852
Bandwidth ⩾ 0	0.829	0.431	0.540	0.326	0.368	0.506	0.335	0.532	0.258	0.512	0.305
Effective obs. < 0	183	172	191	140	181	149	129	271	185	267	268
Effective obs. ≥ 0	204	178	207	178	210	165	198	272	171	292	299

Specifications	L	M	N	O	P	Q	R	Low	Mid-low	Mid-high	High
Conventional	-0.150	0.0362	-0.109	0.139	0.113	-0.0929	-0.00162	-0.0405	0.0840	0.0243	-0.00162
	(0.146)	(0.0826)	(0.0874)	(0.0920)	(0.0909)	(0.0901)	(0.0760)	(0.0854)	(0.0540)	(0.0463)	(0.0760)
Bias-corrected	-0.113	0.0352	-0.119	0.186**	0.144	-0.0774	-0.0160	-0.0548	0.108**	0.0315	-0.0160
	(0.146)	(0.0826)	(0.0874)	(0.0920)	(0.0909)	(0.0901)	(0.0760)	(0.0854)	(0.0540)	(0.0463)	(0.0760)
Robust	-0.113	0.0352	-0.119	0.186*	0.144	-0.0774	-0.0160	-0.0548	0.108*	0.0315	-0.0160
	(0.165)	(0.104)	(0.105)	(0.108)	(0.107)	(0.104)	(0.0929)	(0.0991)	(0.0607)	(0.0520)	(0.0929)
Observations	1,808	2,221	2,343	1,615	1,807	3,082	1,624	2,951	10,991	15,570	1,624
First-stage instrument t-statistic	11.577	32.209	26.19	16.712	17.066	22.981	28.541	23.182	28.274	40.382	28.541
Bandwidth ⩾ 0	0.265	0.413	0.371	0.434	0.423	0.301	0.271	0.416	0.320	0.226	0.271
Effective obs. < 0	120	390	287	281	219	364	288	399	1039	1239	288
Effective obs. ≥ 0	174	323	287	246	260	315	236	444	1239	1387	236

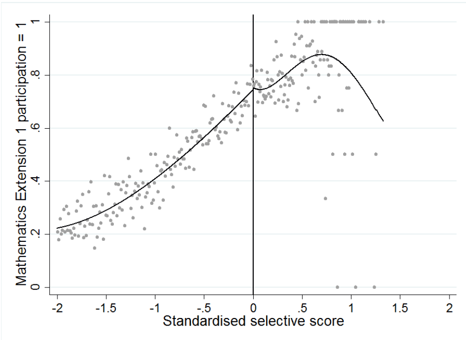
Each column reports regression discontinuity results for a school or stacked sharp sample. Note, School R and the high cut-off sharp samples are identical. 'Bias-corrected' and 'Robust', but not 'Conventional' specification point estimates are bias-corrected for the selected bandwidth following Calonico et al. (2014b). Estimates are non-parametrically estimated using local linear regressions and a triangular kernel. Optimal bandwidths are calculated for each sharp sample based on the common-bandwidth MSE-optimal selector following Calonico et al. (2014b) and Calonico et al. (2016) as described in the text. Observations is the total number of observations available for estimation. The sum of effective observations on either side of the cut-off are those used for estimation. Standard errors are shown in parentheses. Only standard errors for the 'Robust' specification are robust and clustered on Primary FOEI. * significant at 10%, ** significant at 5%, *** significant at 1%.



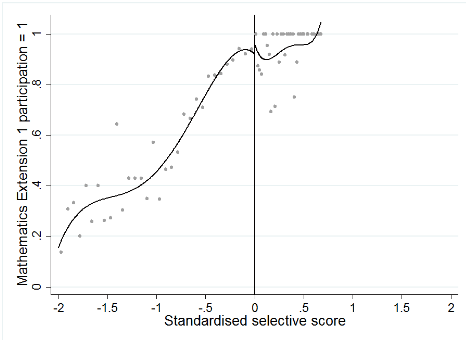
(a) Low cut-off



(b) Mid-low cut-off

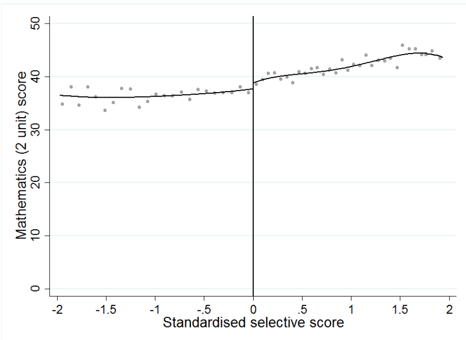


(c) Mid-high cut-off

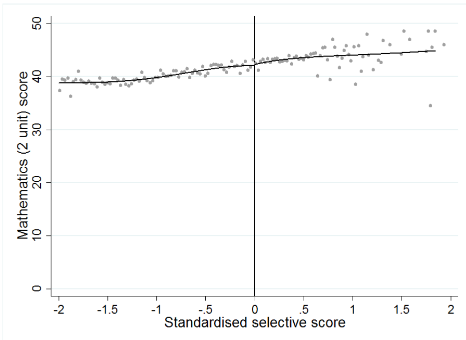


(d) High cut-off

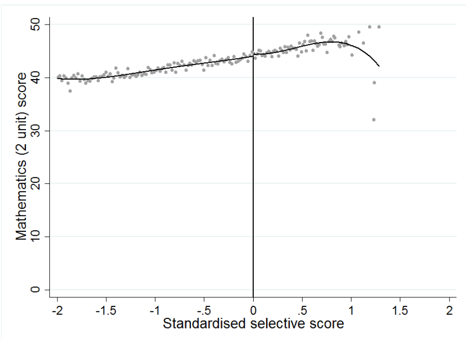
Figure C.8: Mathematics Extension 1 participation in stacked sharp samples



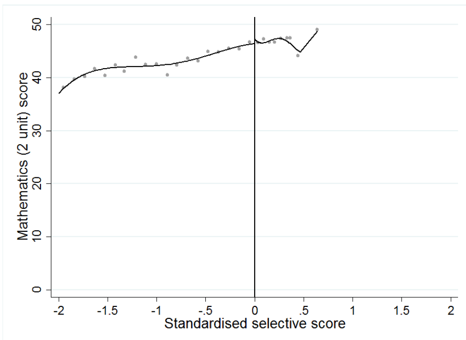
(a) Low cut-off



(b) Mid-low cut-off



(c) Mid-high cut-off



(d) High cut-off

Figure C.9: HSC Mathematics (2 unit) score in stacked sharp samples

Table C.9: Fuzzy regression discontinuity estimates of effects of an offer on Mathematics Extension 1 participation

Specifications	A	B	C	D	E	F	G	H	I	J	K
Conventional	-0.0387	0.135**	0.114	0.130	0.0299	0.270**	-0.00953	-0.175	0.123	0.0708	-0.0142
	(0.110)	(0.0678)	(0.116)	(0.109)	(0.0993)	(0.129)	(0.131)	(0.123)	(0.113)	(0.0988)	(0.0884)
Bias-corrected	-0.0710	0.138**	0.113	0.180*	0.0211	0.305**	0.0367	-0.146	0.104	0.102	-0.00344
	(0.110)	(0.0678)	(0.116)	(0.109)	(0.0993)	(0.129)	(0.131)	(0.123)	(0.113)	(0.0988)	(0.0884)
Robust	-0.0710	0.138*	0.113	0.180	0.0211	0.305**	0.0367	-0.146	0.104	0.102	-0.00344
	(0.130)	(0.0808)	(0.134)	(0.122)	(0.114)	(0.152)	(0.149)	(0.145)	(0.137)	(0.122)	(0.106)
Observations	700	1,236	1,023	1,512	1,535	1,100	1,409	1,731	2,224	1,497	2,714
First-stage instrument t-statistic	9.545	26.261	14.534	10.468	18.977	7.38	7.953	11.439	15.616	19.523	21.693
Bandwidth $\gtrless 0$	0.586	0.675	0.433	0.354	0.472	0.543	0.300	0.403	0.264	0.544	0.322
Effective obs. < 0	120	284	148	155	225	164	117	199	190	280	278
Effective obs. ≥ 0	159	263	163	196	252	174	180	224	172	303	310

Specifications	L	M	N	O	P	Q	R	Low	Mid-low	Mid-high	High
Conventional	-0.0969	0.0848*	0.000833	-0.0404	0.0508	0.0400	-0.0275	0.0420	0.0452	-0.00216	-0.0275
	(0.115)	(0.0507)	(0.0859)	(0.0664)	(0.0789)	(0.0674)	(0.0436)	(0.0662)	(0.0519)	(0.0366)	(0.0436)
Bias-corrected	-0.127	0.112**	-0.0260	-0.0177	0.0763	0.0218	-0.0162	0.0306	0.0568	-0.00617	-0.0162
	(0.115)	(0.0507)	(0.0859)	(0.0664)	(0.0789)	(0.0674)	(0.0436)	(0.0662)	(0.0519)	(0.0366)	(0.0436)
Robust	-0.127	0.112*	-0.0260	-0.0177	0.0763	0.0218	-0.0162	0.0306	0.0568	-0.00617	-0.0162
	(0.129)	(0.0600)	(0.103)	(0.0812)	(0.0941)	(0.0813)	(0.0522)	(0.0752)	(0.0574)	(0.0410)	(0.0522)
Observations	1,808	2,221	2,343	1,615	1,807	3,082	1,624	2,951	10,991	15,570	1,624
First-stage instrument t-statistic	12.757	33.695	25.211	17.017	17.547	21.761	27.5	23.078	28.631	40.565	27.5
Bandwidth $\gtrless 0$	0.283	0.425	0.327	0.442	0.422	0.320	0.283	0.400	0.318	0.228	0.283
Effective obs. < 0	126	399	247	287	218	385	297	382	1036	1242	297
Effective obs. ≥ 0	186	331	273	248	259	324	238	431	1234	1392	238

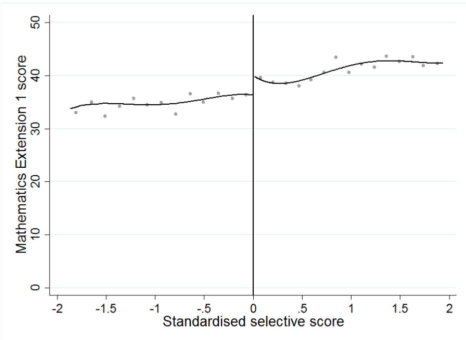
Each column reports regression discontinuity results for a school or stacked sharp sample. Note, School R and the high cut-off sharp samples are identical. 'Bias-corrected' and 'Robust', but not 'Conventional' specification point estimates are bias-corrected for the selected bandwidth following Calonico et al. (2014b). Estimates are non-parametrically estimated using local linear regressions and a triangular kernel. Optimal bandwidths are calculated for each sharp sample based on the common-bandwidth MSE-optimal selector following Calonico et al. (2014b) and Calonico et al. (2016) as described in the text. Observations is the total number of observations available for estimation. The sum of effective observations on either side of the cut-off are those used for estimation. Standard errors are shown in parentheses. Only standard errors for the 'Robust' specification are robust and clustered on Primary FOEI. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table C.10: Fuzzy regression discontinuity estimates of effects of an offer on Mathematics (2 unit) scores

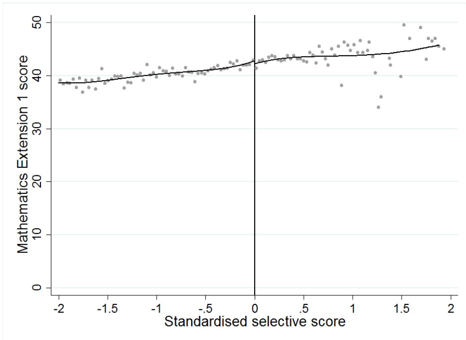
Specifications	A	B	C	D	E	F	G	H	I	J	K
Conventional	3.069	-0.0334	2.637	-0.214	-1.209	1.609	-0.952	0.388	-1.006	1.457	-1.116
	(2.365)	(1.967)	(1.873)	(1.539)	(0.963)	(1.793)	(1.700)	(1.217)	(1.621)	(1.298)	(0.938)
Bias-corrected	3.367	-0.655	3.018	-0.200	-1.566	2.321	-0.726	0.378	-1.437	1.735	-1.346
	(2.365)	(1.967)	(1.873)	(1.539)	(0.963)	(1.793)	(1.700)	(1.217)	(1.621)	(1.298)	(0.938)
Robust	3.367	-0.655	3.018	-0.200	-1.566	2.321	-0.726	0.378	-1.437	1.735	-1.346
	(2.755)	(2.240)	(2.312)	(1.800)	(1.116)	(2.486)	(2.107)	(1.527)	(1.881)	(1.428)	(1.138)
Observations	311	437	461	713	745	539	683	871	1,233	908	1,432
First-stage instrument t-statistic	8.4	15.272	N/A	8.029	24.229	4.756	6.503	8.923	8.27	26.501	21.45
Bandwidth $\gtrless 0$	0.568	0.507	0.600	0.378	0.510	0.762	0.431	0.535	0.189	0.370	0.377
Effective obs. < 0	51	70	94	99	142	134	91	160	83	119	176
Effective obs. ≥ 0	87	96	123	134	137	143	120	157	84	160	214

Specifications	L	M	N	O	P	Q	R	Low	Mid-low	Mid-high	High
Conventional	0.0325	0.432	-0.345	1.888**	2.304	1.748*	0.178	1.569	-0.215	0.122	0.178
	(0.892)	(0.747)	(0.603)	(0.853)	(1.916)	(0.987)	(0.640)	(1.140)	(0.650)	(0.493)	(0.640)
Bias-corrected	0.0896	0.261	-0.388	2.569***	2.293	1.966**	0.143	1.518	-0.241	0.0839	0.143
	(0.892)	(0.747)	(0.603)	(0.853)	(1.916)	(0.987)	(0.640)	(1.140)	(0.650)	(0.493)	(0.640)
Robust	0.0896	0.261	-0.388	2.569**	2.293	1.966*	0.143	1.518	-0.241	0.0839	0.143
	(1.080)	(0.896)	(0.774)	(1.035)	(2.032)	(1.096)	(0.790)	(1.383)	(0.735)	(0.565)	(0.790)
Observations	978	1,182	1,251	874	891	1,613	735	1,206	5,679	8,207	735
First-stage instrument t-statistic	601.685	18.46	24.862	12.189	45.838	10.399	18.126	23.49	22.706	37.948	18.126
Bandwidth $\gtrless 0$	0.303	0.309	0.708	0.456	0.179	0.203	0.242	0.642	0.350	0.243	0.242
Effective obs. < 0	99	173	312	182	52	132	112	244	673	779	112
Effective obs. ≥ 0	93	166	203	152	59	95	98	347	790	809	98

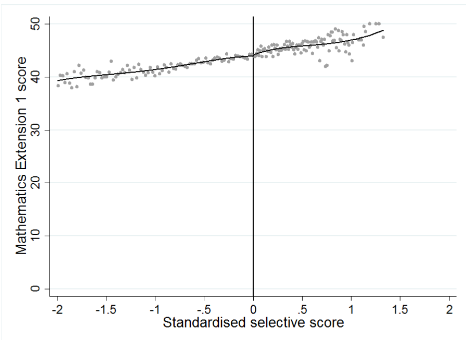
Each column reports regression discontinuity results for a school or stacked sharp sample. Note, School R and the high cut-off sharp samples are identical. 'Bias-corrected' and 'Robust', but not 'Conventional' specification point estimates are bias-corrected for the selected bandwidth following Calonico et al. (2014b). Estimates are non-parametrically estimated using local linear regressions and a triangular kernel. Optimal bandwidths are calculated for each sharp sample based on the common-bandwidth MSE-optimal selector following Calonico et al. (2014b) and Calonico et al. (2016) as described in the text. Observations is the total number of observations available for estimation. The sum of effective observations on either side of the cut-off are those used for estimation. Standard errors are shown in parentheses. Only standard errors for the 'Robust' specification are robust and clustered on Primary FOEI. * significant at 10%, ** significant at 5%, *** significant at 1%.



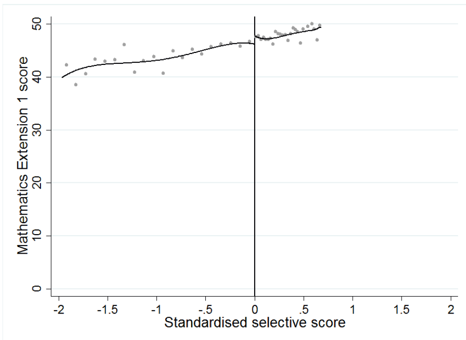
(a) Low cut-off



(b) Mid-low cut-off



(c) Mid-high cut-off



(d) High cut-off

Figure C.10: HSC Mathematics Extension 1 score in stacked sharp samples

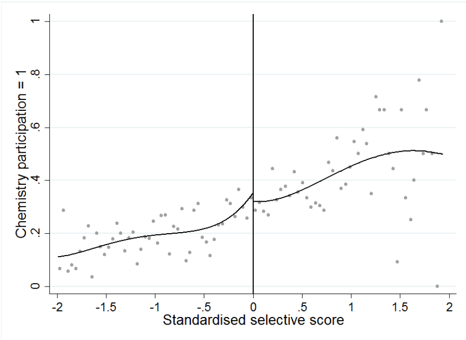
Table C.11: Fuzzy regression discontinuity estimates of effects of an offer on Mathematics Extension 1 scores

Specifications	A	B	C	D	E	F	G	H	I	J	K
Conventional	7.183***	-3.136	2.758	2.768*	-2.291	6.114	-0.578	-4.181**	-1.490	1.484	-2.636**
	(2.662)	(2.252)	(3.056)	(1.626)	(1.538)	(3.835)	(1.792)	(1.948)	(1.145)	(1.569)	(1.325)
Bias-corrected	8.257***	-3.613	3.225	2.905*	-2.756*	7.521**	-1.586	-5.738***	-1.539	2.067	-3.033**
	(2.662)	(2.252)	(3.056)	(1.626)	(1.538)	(3.835)	(1.792)	(1.948)	(1.145)	(1.569)	(1.325)
Robust	8.257***	-3.613	3.225	2.905	-2.756	7.521	-1.586	-5.738**	-1.539	2.067	-3.033*
	(3.011)	(2.622)	(3.458)	(1.833)	(1.826)	(4.764)	(2.137)	(2.310)	(1.395)	(1.824)	(1.651)
Observations	154	247	229	416	466	280	606	547	836	725	1,067
First-stage instrument t-statistic	N/A	N/A	N/A	19.314	11.474	2.556	7.096	5.032	12.584	11.31	19.237
Bandwidth $\gtrless 0$	0.497	0.412	0.504	0.516	0.621	0.463	0.409	0.285	0.278	0.288	0.316
Effective obs. < 0	22	19	27	85	119	32	108	79	109	91	176
Effective obs. ≥ 0	35	52	56	130	163	57	171	88	121	126	199

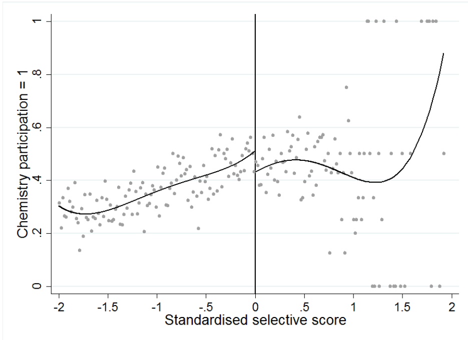
Specifications	L	M	N	O	P	Q	R	Low	Mid-low	Mid-high	High
Conventional	-0.00446	1.238	-0.369	1.150	-0.297	2.257*	1.019**	2.423	-0.263	0.140	1.019**
	(0.832)	(0.956)	(0.861)	(0.879)	(0.960)	(1.211)	(0.421)	(1.871)	(0.552)	(0.499)	(0.421)
Bias-corrected	0.0881	0.934	-0.481	0.627	-0.739	2.472**	1.034**	3.005	-0.373	0.0257	1.034**
	(0.832)	(0.956)	(0.861)	(0.879)	(0.960)	(1.211)	(0.421)	(1.871)	(0.552)	(0.499)	(0.421)
Robust	0.0881	0.934	-0.481	0.627	-0.739	2.472*	1.034**	3.005	-0.373	0.0257	1.034**
	(0.943)	(1.088)	(1.091)	(1.159)	(1.153)	(1.362)	(0.515)	(2.152)	(0.625)	(0.577)	(0.515)
Observations	639	1,176	922	839	736	1,465	995	628	3,867	6,827	995
First-stage instrument t-statistic	21.813	25.779	19.678	14.408	16.49	12.029	25.501	N/A	29.796	38.957	25.501
Bandwidth $\gtrless 0$	0.419	0.256	0.370	0.471	0.342	0.183	0.233	0.523	0.469	0.263	0.233
Effective obs. < 0	101	183	170	220	120	155	229	75	793	1007	229
Effective obs. ≥ 0	154	201	204	207	173	173	199	159	992	1179	199

Each column reports regression discontinuity results for a school or stacked sharp sample. Note, School R and the high cut-off sharp samples are identical. 'Bias-corrected' and 'Robust', but not 'Conventional' specification point estimates are bias-corrected for the selected bandwidth following Calonico et al. (2014b). Estimates are non-parametrically estimated using local linear regressions and a triangular kernel. Optimal bandwidths are calculated for each sharp sample based on the common-bandwidth MSE-optimal selector following Calonico et al. (2014b) and Calonico et al. (2016) as described in the text. Observations is the total number of observations available for estimation. The sum of effective observations on either side of the cut-off are those used for estimation. Standard errors are shown in parentheses. Only standard errors for the 'Robust' specification are robust and clustered on Primary FOEI. * significant at 10%, ** significant at 5%, *** significant at 1%.

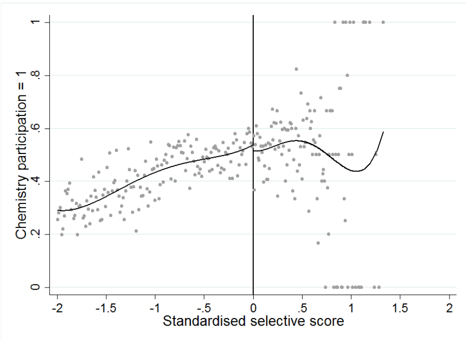
C.5 HSC Science participation and achievement



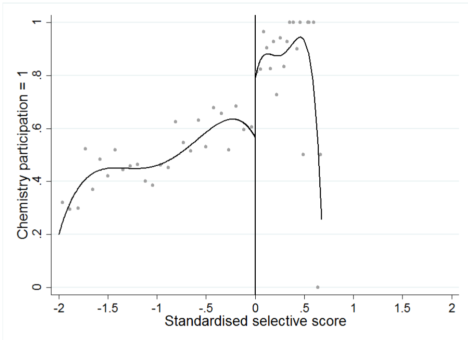
(a) Low cut-off



(b) Mid-low cut-off



(c) Mid-high cut-off



(d) High cut-off

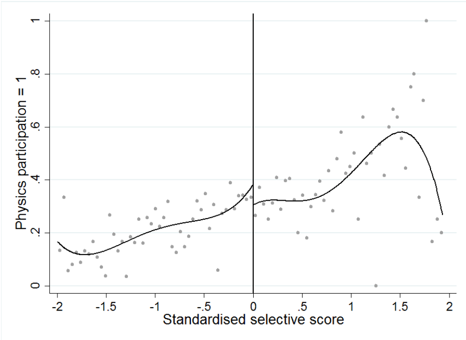
Figure C.11: Chemistry participation in stacked sharp samples

Table C.12: Fuzzy regression discontinuity estimates of effects of an offer on Chemistry participation

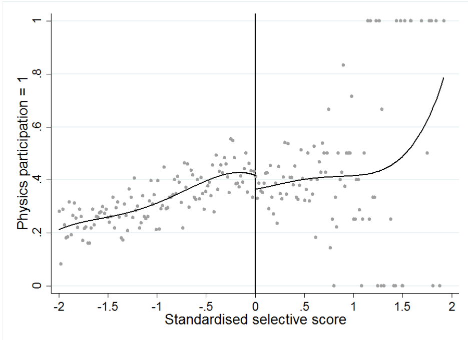
Specifications	A	B	C	D	E	F	G	H	I	J	K
Conventional	0.162	-0.0737	0.0801	-0.200	-0.0861	-0.0678	-0.160	-0.267**	0.368**	-0.0594	-0.133
	(0.117)	(0.0789)	(0.133)	(0.130)	(0.104)	(0.153)	(0.134)	(0.134)	(0.150)	(0.126)	(0.0962)
Bias-corrected	0.200*	-0.0897	0.0865	-0.229*	-0.130	-0.0691	-0.166	-0.250*	0.420***	-0.0704	-0.183*
	(0.117)	(0.0789)	(0.133)	(0.130)	(0.104)	(0.153)	(0.134)	(0.134)	(0.150)	(0.126)	(0.0962)
Robust	0.200	-0.0897	0.0865	-0.229	-0.130	-0.0691	-0.166	-0.250	0.420**	-0.0704	-0.183
	(0.137)	(0.0928)	(0.157)	(0.145)	(0.121)	(0.174)	(0.153)	(0.158)	(0.174)	(0.150)	(0.114)
Observations	700	1,236	1,023	1,512	1,535	1,100	1,409	1,731	2,224	1,497	2,714
First-stage instrument t-statistic	10.229	28.12	14.822	8.98	16.535	6.917	9.742	13.326	14.717	15.227	20.188
Bandwidth $\gtrless 0$	0.620	0.707	0.447	0.282	0.410	0.476	0.391	0.496	0.193	0.358	0.276
Effective obs. < 0	129	292	152	128	203	137	157	247	138	186	238
Effective obs. ≥ 0	166	272	169	155	229	155	230	259	137	222	270

Specifications	L	M	N	O	P	Q	R	Low	Mid-low	Mid-high	High
Conventional	0.00882	-0.114	-0.0360	0.0331	-0.0712	0.190*	0.279***	0.00871	-0.0737	-0.0316	0.279***
	(0.134)	(0.0890)	(0.0905)	(0.0921)	(0.108)	(0.109)	(0.0875)	(0.0765)	(0.0587)	(0.0446)	(0.0875)
Bias-corrected	0.0278	-0.117	-0.00953	0.0509	-0.110	0.210*	0.277***	0.00913	-0.0727	-0.0330	0.277***
	(0.134)	(0.0890)	(0.0905)	(0.0921)	(0.108)	(0.109)	(0.0875)	(0.0765)	(0.0587)	(0.0446)	(0.0875)
Robust	0.0278	-0.117	-0.00953	0.0509	-0.110	0.210*	0.277**	0.00913	-0.0727	-0.0330	0.277**
	(0.156)	(0.105)	(0.107)	(0.113)	(0.129)	(0.123)	(0.112)	(0.0901)	(0.0651)	(0.0505)	(0.112)
Observations	1,808	2,221	2,343	1,615	1,807	3,082	1,624	2,951	10,991	15,570	1,624
First-stage instrument t-statistic	14.061	29.282	25.621	18.393	16.737	16.457	20.763	22.938	28.337	43.292	20.763
Bandwidth $\gtrless 0$	0.335	0.309	0.336	0.501	0.413	0.195	0.190	0.424	0.315	0.251	0.190
Effective obs. < 0	158	298	252	315	213	216	200	404	1024	1381	200
Effective obs. ≥ 0	214	272	276	265	256	209	204	450	1221	1502	204

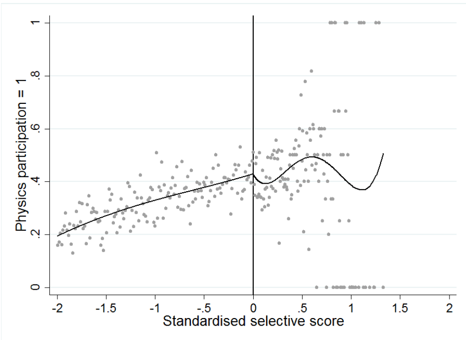
Each column reports regression discontinuity results for a school or stacked sharp sample. Note, School R and the high cut-off sharp samples are identical. 'Bias-corrected' and 'Robust', but not 'Conventional' specification point estimates are bias-corrected for the selected bandwidth following Calonico et al. (2014b). Estimates are non-parametrically estimated using local linear regressions and a triangular kernel. Optimal bandwidths are calculated for each sharp sample based on the common-bandwidth MSE-optimal selector following Calonico et al. (2014b) and Calonico et al. (2016) as described in the text. Observations is the total number of observations available for estimation. The sum of effective observations on either side of the cut-off are those used for estimation. Standard errors are shown in parentheses. Only standard errors for the 'Robust' specification are robust and clustered on Primary FOEI. * significant at 10%, ** significant at 5%, *** significant at 1%.



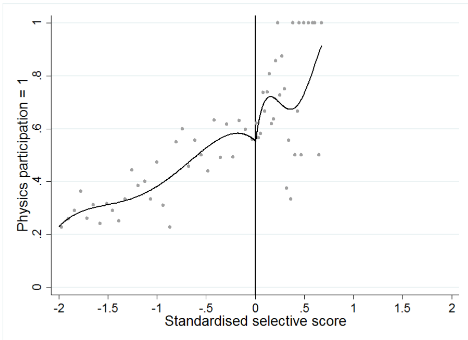
(a) Low cut-off



(b) Mid-low cut-off

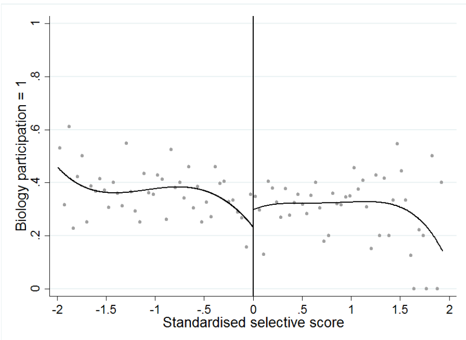


(c) Mid-high cut-off

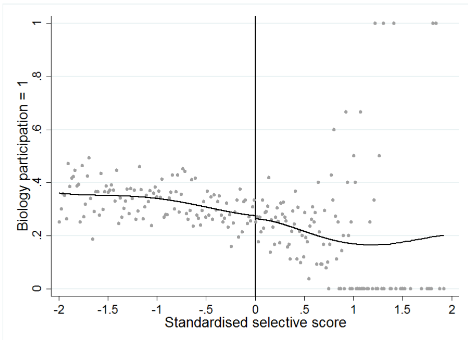


(d) High cut-off

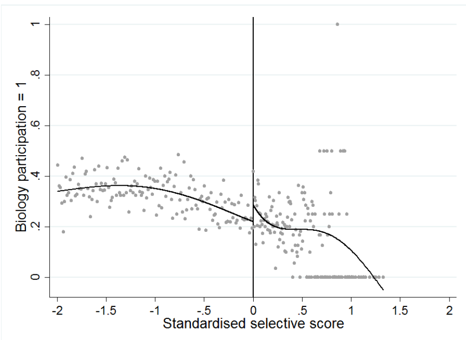
Figure C.12: Physics participation in stacked sharp samples



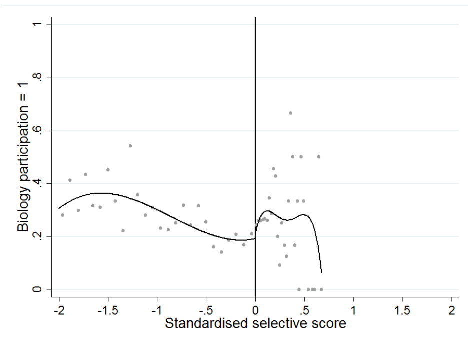
(a) Low cut-off



(b) Mid-low cut-off



(c) Mid-high cut-off



(d) High cut-off

Figure C.13: Biology participation in stacked sharp samples

Table C.13: Fuzzy regression discontinuity estimates of effects of an offer on Physics participation

Specifications	A	B	C	D	E	F	G	H	I	J	K
Conventional	0.0732	0.0605	-0.0624	-0.0589	-0.0628	0.175	-0.0438	-0.00742	0.178	-0.0589	-0.0915
	(0.122)	(0.0779)	(0.103)	(0.120)	(0.0816)	(0.143)	(0.148)	(0.0936)	(0.121)	(0.102)	(0.0834)
Bias-corrected	0.0790	0.0831	-0.0575	-0.0357	-0.0589	0.203	-0.00149	-0.0167	0.210*	-0.00811	-0.119
	(0.122)	(0.0779)	(0.103)	(0.120)	(0.0816)	(0.143)	(0.148)	(0.0936)	(0.121)	(0.102)	(0.0834)
Robust	0.0790	0.0831	-0.0575	-0.0357	-0.0589	0.203	-0.00149	-0.0167	0.210	-0.00811	-0.119
	(0.137)	(0.0898)	(0.118)	(0.134)	(0.0941)	(0.167)	(0.164)	(0.114)	(0.145)	(0.123)	(0.0986)
Observations	700	1,236	1,023	1,512	1,535	1,100	1,409	1,731	2,224	1,497	2,714
First-stage instrument t-statistic	9.484	29.628	14.941	11.223	17.888	6.058	8.813	12.738	15.584	17.92	23.506
Bandwidth $\gtrless 0$	0.567	0.781	0.450	0.384	0.441	0.415	0.337	0.494	0.256	0.479	0.333
Effective obs. < 0	117	323	153	171	215	104	130	246	183	247	289
Effective obs. ≥ 0	156	291	169	213	240	137	200	259	170	272	315

Specifications	L	M	N	O	P	Q	R	Low	Mid-low	Mid-high	High
Conventional	-0.0270	-0.0989	0.0554	-0.0445	-0.0371	0.237**	0.0834	0.00580	-0.0226	0.0290	0.0834
	(0.0884)	(0.0763)	(0.0697)	(0.0862)	(0.0874)	(0.107)	(0.0986)	(0.0721)	(0.0521)	(0.0435)	(0.0986)
Bias-corrected	-0.0280	-0.0649	0.0807	-0.0116	-0.0405	0.240**	0.124	0.00172	-0.00878	0.0350	0.124
	(0.0884)	(0.0763)	(0.0697)	(0.0862)	(0.0874)	(0.107)	(0.0986)	(0.0721)	(0.0521)	(0.0435)	(0.0986)
Robust	-0.0280	-0.0649	0.0807	-0.0116	-0.0405	0.240**	0.124	0.00172	-0.00878	0.0350	0.124
	(0.101)	(0.0945)	(0.0833)	(0.103)	(0.101)	(0.120)	(0.118)	(0.0837)	(0.0577)	(0.0489)	(0.118)
Observations	1,808	2,221	2,343	1,615	1,807	3,082	1,624	2,951	10,991	15,570	1,624
First-stage instrument t-statistic	14.477	32.867	25.097	22.155	16.961	17.143	23.183	23.429	29.245	39.588	23.183
Bandwidth $\gtrless 0$	0.326	0.436	0.328	0.640	0.392	0.210	0.227	0.428	0.331	0.219	0.227
Effective obs. < 0	155	405	247	375	203	239	244	407	1076	1195	244
Effective obs. ≥ 0	210	335	273	307	241	224	218	453	1277	1347	218

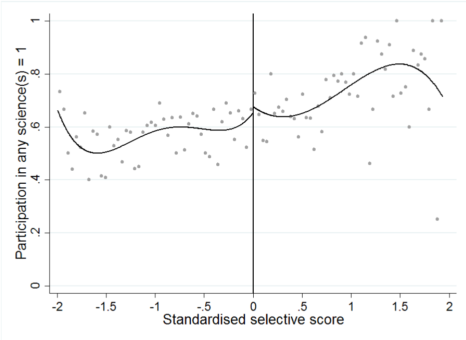
Each column reports regression discontinuity results for a school or stacked sharp sample. Note, School R and the high cut-off sharp samples are identical. 'Bias-corrected' and 'Robust', but not 'Conventional' specification point estimates are bias-corrected for the selected bandwidth following Calonico et al. (2014b). Estimates are non-parametrically estimated using local linear regressions and a triangular kernel. Optimal bandwidths are calculated for each sharp sample based on the common-bandwidth MSE-optimal selector following Calonico et al. (2014b) and Calonico et al. (2016) as described in the text. Observations is the total number of observations available for estimation. The sum of effective observations on either side of the cut-off are those used for estimation. Standard errors are shown in parentheses. Only standard errors for the 'Robust' specification are robust and clustered on Primary FOEI. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table C.14: Fuzzy regression discontinuity estimates of effects of an offer on Biology participation

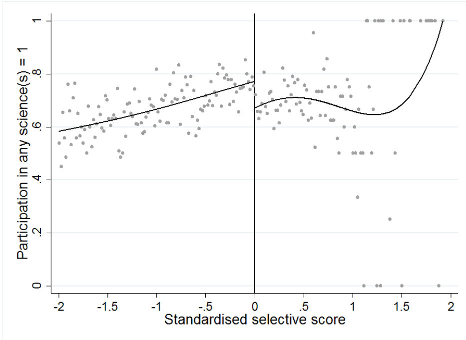
Specifications	A	B	C	D	E	F	G	H	I	J	K
Conventional	0.105	-0.0189	0.128	0.0441	0.0624	-0.446***	-0.256*	-0.175*	0.0530	0.107	0.250***
	(0.149)	(0.103)	(0.116)	(0.119)	(0.101)	(0.159)	(0.136)	(0.102)	(0.0905)	(0.0871)	(0.0795)
Bias-corrected	0.149	-0.0218	0.144	0.0538	0.0578	-0.511***	-0.315**	-0.113	0.0263	0.103	0.304***
	(0.149)	(0.103)	(0.116)	(0.119)	(0.101)	(0.159)	(0.136)	(0.102)	(0.0905)	(0.0871)	(0.0795)
Robust	0.149	-0.0218	0.144	0.0538	0.0578	-0.511***	-0.315**	-0.113	0.0263	0.103	0.304***
	(0.175)	(0.119)	(0.136)	(0.132)	(0.114)	(0.181)	(0.152)	(0.124)	(0.109)	(0.105)	(0.0950)
Observations	700	1,236	1,023	1,512	1,535	1,100	1,409	1,731	2,224	1,497	2,714
First-stage instrument t-statistic	9.567	21.895	15.168	11.028	18.811	6.245	8.750	13.914	16.446	15.703	23.536
Bandwidth $\gtrless 0$	0.560	0.534	0.488	0.384	0.468	0.435	0.340	0.550	0.318	0.383	0.366
Effective obs. < 0	116	216	173	171	224	119	133	273	234	205	318
Effective obs. ≥ 0	155	215	185	213	251	144	202	277	204	234	325

Specifications	L	M	N	O	P	Q	R	Low	Mid-low	Mid-high	High
Conventional	-0.0552	-0.0310	0.0151	0.0836	0.131	-0.0275	0.00882	0.0525	-0.0542	0.0775*	0.00882
	(0.0899)	(0.0606)	(0.0725)	(0.0768)	(0.101)	(0.0999)	(0.0650)	(0.0691)	(0.0443)	(0.0400)	(0.0650)
Bias-corrected	-0.0399	-0.0311	0.0312	0.149*	0.157	-0.0392	-0.0245	0.0533	-0.0640	0.0906**	-0.0245
	(0.0899)	(0.0606)	(0.0725)	(0.0768)	(0.101)	(0.0999)	(0.0650)	(0.0691)	(0.0443)	(0.0400)	(0.0650)
Robust	-0.0399	-0.0311	0.0312	0.149	0.157	-0.0392	-0.0245	0.0533	-0.0640	0.0906**	-0.0245
	(0.109)	(0.0719)	(0.0878)	(0.0949)	(0.118)	(0.112)	(0.0783)	(0.0805)	(0.0491)	(0.0448)	(0.0783)
Observations	1,808	2,221	2,343	1,615	1,807	3,082	1,624	2,951	10,991	15,570	1,624
First-stage instrument t-statistic	20.507	32.735	27.258	16.812	16.45	16.296	23.546	28.811	32.096	40.681	23.546
Bandwidth $\gtrless 0$	0.470	0.406	0.367	0.438	0.366	0.197	0.231	0.583	0.372	0.228	0.231
Effective obs. < 0	242	381	285	286	198	217	247	566	1246	1246	247
Effective obs. ≥ 0	246	319	286	247	224	211	218	604	1408	1394	218

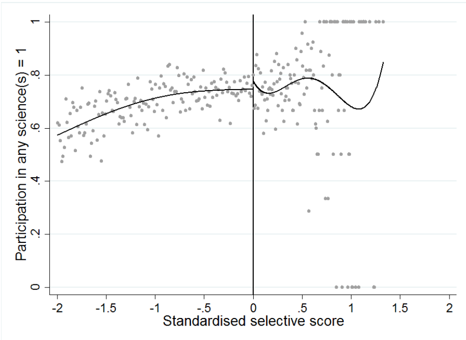
Each column reports regression discontinuity results for a school or stacked sharp sample. Note, School R and the high cut-off sharp samples are identical. ‘Bias-corrected’ and ‘Robust’, but not ‘Conventional’ specification point estimates are bias-corrected for the selected bandwidth following Calonico et al. (2014b). Estimates are non-parametrically estimated using local linear regressions and a triangular kernel. Optimal bandwidths are calculated for each sharp sample based on the common-bandwidth MSE-optimal selector following Calonico et al. (2014b) and Calonico et al. (2016) as described in the text. Observations is the total number of observations available for estimation. The sum of effective observations on either side of the cut-off are those used for estimation. Standard errors are shown in parentheses. Only standard errors for the ‘Robust’ specification are robust and clustered on Primary FOEI. * significant at 10%, ** significant at 5%, *** significant at 1%.



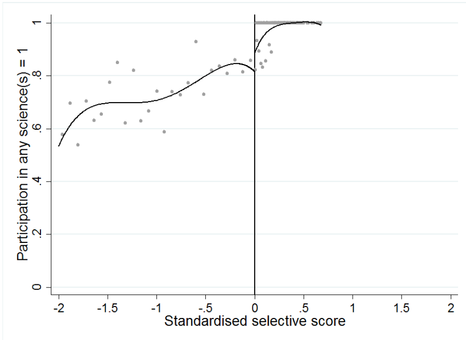
(a) Low cut-off



(b) Mid-low cut-off

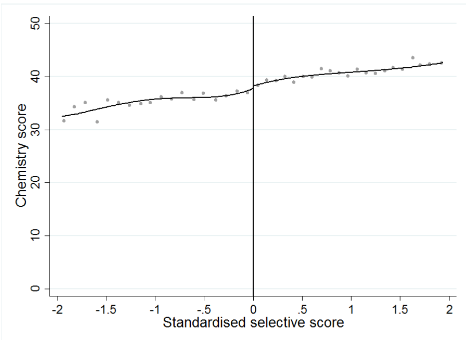


(c) Mid-high cut-off

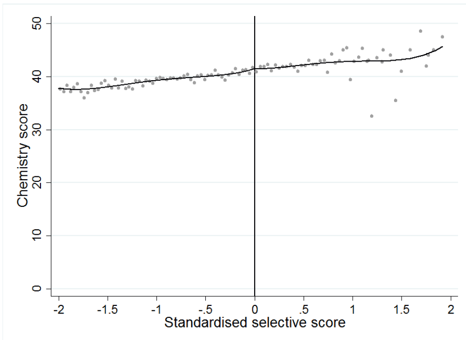


(d) High cut-off

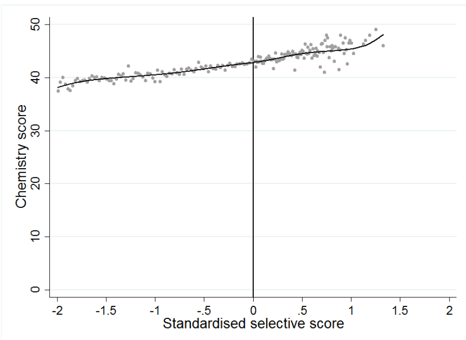
Figure C.14: Any science subject(s) participation in stacked sharp samples



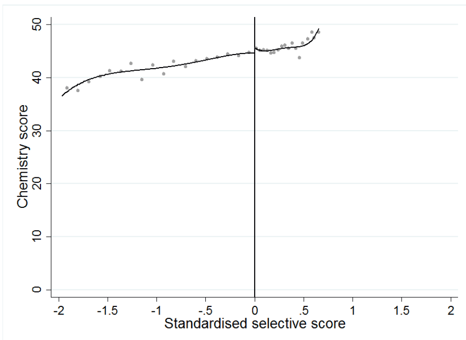
(a) Low cut-off



(b) Mid-low cut-off



(c) Mid-high cut-off



(d) High cut-off

Figure C.15: HSC Chemistry score in stacked sharp samples

Table C.15: Fuzzy regression discontinuity estimates of effects of an offer on participation in at least one science course

Specifications	A	B	C	D	E	F	G	H	I	J	K
Conventional	0.0915	0.0453	0.101	-0.137	-0.168**	-0.234*	-0.117	-0.322***	0.359***	0.0652	0.0406
	(0.134)	(0.120)	(0.0945)	(0.0861)	(0.0849)	(0.137)	(0.116)	(0.111)	(0.119)	(0.0765)	(0.0668)
Bias-corrected	0.142	0.0214	0.117	-0.159*	-0.203**	-0.257*	-0.0648	-0.282**	0.397***	0.114	0.0454
	(0.134)	(0.120)	(0.0945)	(0.0861)	(0.0849)	(0.137)	(0.116)	(0.111)	(0.119)	(0.0765)	(0.0668)
Robust	0.142	0.0214	0.117	-0.159*	-0.203*	-0.257	-0.0648	-0.282**	0.397***	0.114	0.0454
	(0.153)	(0.141)	(0.110)	(0.0959)	(0.105)	(0.164)	(0.137)	(0.132)	(0.135)	(0.0927)	(0.0795)
Observations	700	1,236	1,023	1,512	1,535	1,100	1,409	1,731	2,224	1,497	2,714
First-stage instrument t-statistic	9.644	21.926	16.786	15.4	21.413	7.541	8.99	13.268	14.502	19.478	23.813
Bandwidth $\gtrless 0$	0.580	0.497	0.507	0.549	0.547	0.586	0.374	0.522	0.178	0.505	0.332
Effective obs. < 0	119	196	177	253	264	181	152	266	122	263	288
Effective obs. ≥ 0	158	199	194	286	272	189	222	266	126	288	315

Specifications	L	M	N	O	P	Q	R	Low	Mid-low	Mid-high	High
Conventional	-0.0303	-0.0303	-0.0104	0.139	-0.0130	0.0925	0.103	0.0905	-0.104**	0.0268	0.103
	(0.109)	(0.0653)	(0.0804)	(0.0904)	(0.110)	(0.0760)	(0.0635)	(0.0743)	(0.0460)	(0.0356)	(0.0635)
Bias-corrected	-0.00690	-0.000911	-0.00388	0.212**	-0.0488	0.0947	0.115*	0.0947	-0.0973**	0.0357	0.115*
	(0.109)	(0.0653)	(0.0804)	(0.0904)	(0.110)	(0.0760)	(0.0635)	(0.0743)	(0.0460)	(0.0356)	(0.0635)
Robust	-0.00690	-0.000911	-0.00388	0.212*	-0.0488	0.0947	0.115	0.0947	-0.0973*	0.0357	0.115
	(0.137)	(0.0775)	(0.0956)	(0.109)	(0.133)	(0.0899)	(0.0760)	(0.0867)	(0.0541)	(0.0406)	(0.0760)
Observations	1,808	2,221	2,343	1,615	1,807	3,082	1,624	2,951	10,991	15,570	1,624
First-stage instrument t-statistic	15.638	32.277	28.892	15.74	17.383	19.998	27.959	23.767	30.18	46.68	27.959
Bandwidth $\gtrless 0$	0.393	0.379	0.406	0.384	0.453	0.261	0.289	0.447	0.383	0.283	0.289
Effective obs. < 0	189	354	310	255	228	304	306	419	1277	1540	306
Effective obs. ≥ 0	241	305	299	231	278	280	242	476	1450	1627	242

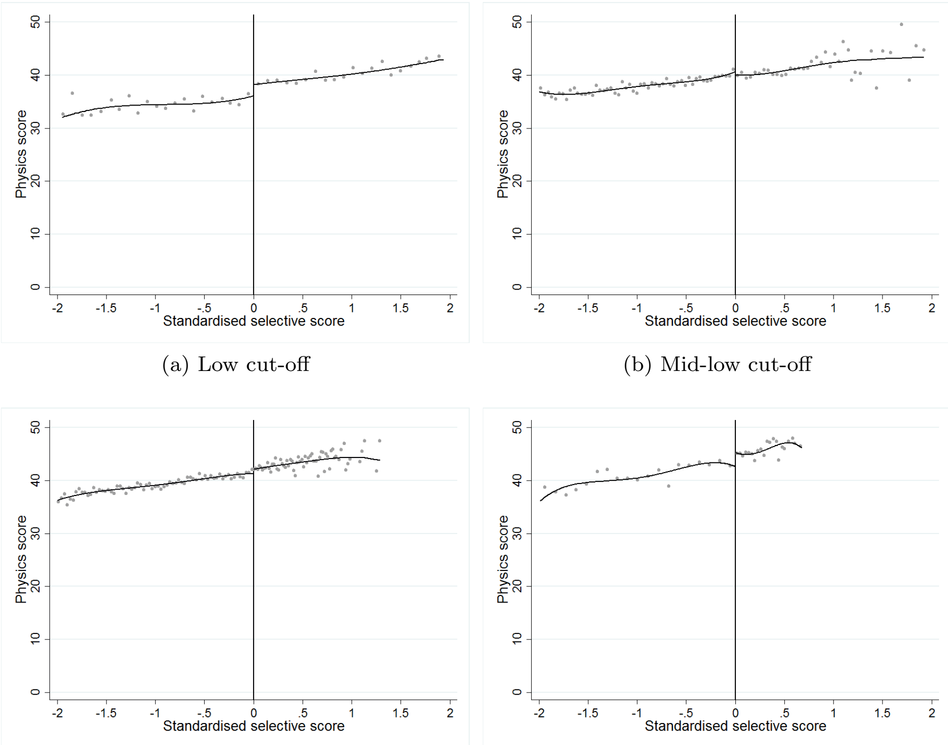
Each column reports regression discontinuity results for a school or stacked sharp sample. Note, School R and the high cut-off sharp samples are identical. 'Bias-corrected' and 'Robust', but not 'Conventional' specification point estimates are bias-corrected for the selected bandwidth following Calonico et al. (2014b). Estimates are non-parametrically estimated using local linear regressions and a triangular kernel. Optimal bandwidths are calculated for each sharp sample based on the common-bandwidth MSE-optimal selector following Calonico et al. (2014b) and Calonico et al. (2016) as described in the text. Observations is the total number of observations available for estimation. The sum of effective observations on either side of the cut-off are those used for estimation. Standard errors are shown in parentheses. Only standard errors for the 'Robust' specification are robust and clustered on Primary FOEI. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table C.16: Fuzzy regression discontinuity estimates of effects of an offer on Chemistry scores

Specifications	A	B	C	D	E	F	G	H	I	J	K
Conventional	1.905	-3.814**	2.689	-0.737	-0.342	1.427	1.406	-2.806**	0.915	0.480	-1.330
	(1.857)	(1.581)	(1.913)	(1.467)	(0.803)	(2.449)	(2.119)	(1.246)	(1.185)	(1.407)	(0.844)
Bias-corrected	1.895	-4.486***	2.690	-1.185	-0.342	1.865	1.635	-3.250***	0.808	0.716	-1.753**
	(1.857)	(1.581)	(1.913)	(1.467)	(0.803)	(2.449)	(2.119)	(1.246)	(1.185)	(1.407)	(0.844)
Robust	1.895	-4.486**	2.690	-1.185	-0.342	1.865	1.635	-3.250**	0.808	0.716	-1.753*
	(2.217)	(1.792)	(2.374)	(1.666)	(0.925)	(2.674)	(2.496)	(1.477)	(1.479)	(1.635)	(1.020)
Observations	226	314	279	529	479	306	534	555	957	623	1,063
First-stage instrument t-statistic	5.034	22.748	N/A	5.993	N/A	2.505	4.05	4.961	17.837	7.998	16.44
Bandwidth $\gtrless 0$	0.727	0.402	0.629	0.363	0.884	0.487	0.369	0.376	0.261	0.348	0.304
Effective obs. < 0	46	46	52	69	153	43	76	91	103	83	140
Effective obs. ≥ 0	70	56	77	101	121	62	98	66	108	108	174

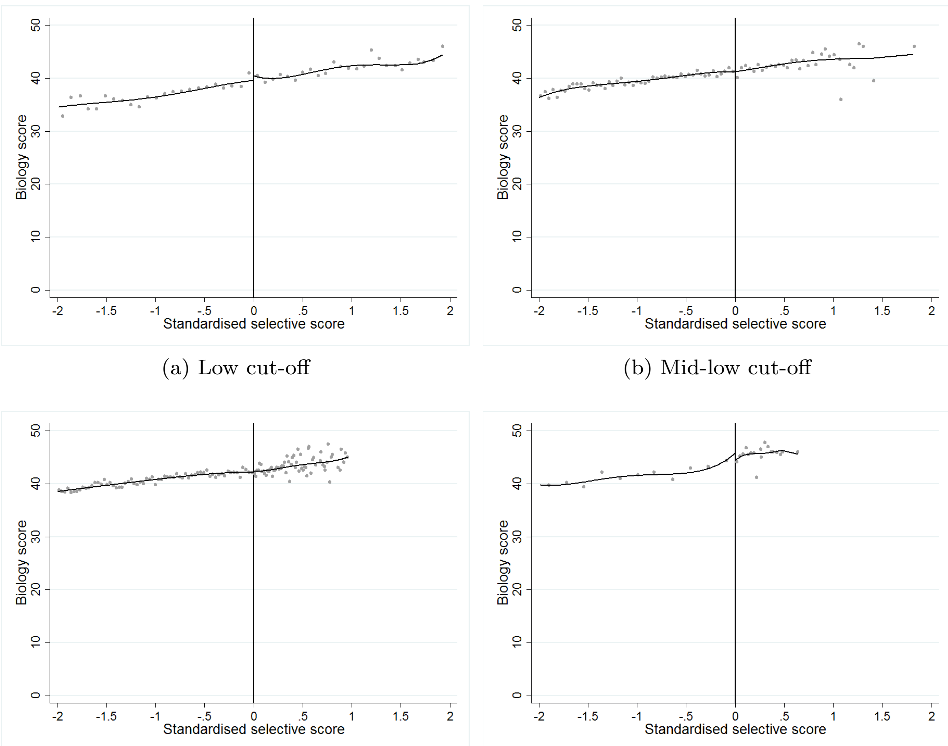
Specifications	L	M	N	O	P	Q	R	Low	Mid-low	Mid-high	High
Conventional	-0.927	0.813	-1.481*	-0.613	0.206	0.913	0.505	0.177	0.150	-0.279	0.505
	(0.832)	(0.543)	(0.793)	(0.846)	(0.727)	(0.756)	(0.601)	(1.161)	(0.480)	(0.393)	(0.601)
Bias-corrected	-1.153	0.807	-1.347*	-0.616	-0.0592	1.037	0.596	-0.154	0.115	-0.337	0.596
	(0.832)	(0.543)	(0.793)	(0.846)	(0.727)	(0.756)	(0.601)	(1.161)	(0.480)	(0.393)	(0.601)
Robust	-1.153	0.807	-1.347	-0.616	-0.0592	1.037	0.596	-0.154	0.115	-0.337	0.596
	(1.011)	(0.677)	(0.968)	(0.962)	(0.890)	(0.832)	(0.733)	(1.328)	(0.543)	(0.439)	(0.733)
Observations	658	990	855	647	665	1,376	883	816	3,976	6,244	883
First-stage instrument t-statistic	14.375	31.236	22.916	16.174	15.647	11.778	18.045	13.489	25.736	35.773	18.045
Bandwidth $\gtrless 0$	0.375	0.554	0.519	0.408	0.460	0.234	0.260	0.519	0.478	0.251	0.260
Effective obs. < 0	83	271	166	136	102	151	173	130	751	707	173
Effective obs. ≥ 0	109	198	157	111	163	158	196	191	795	791	196

Each column reports regression discontinuity results for a school or stacked sharp sample. Note, School R and the high cut-off sharp samples are identical. 'Bias-corrected' and 'Robust', but not 'Conventional' specification point estimates are bias-corrected for the selected bandwidth following Calonico et al. (2014b). Estimates are non-parametrically estimated using local linear regressions and a triangular kernel. Optimal bandwidths are calculated for each sharp sample based on the common-bandwidth MSE-optimal selector following Calonico et al. (2014b) and Calonico et al. (2016) as described in the text. Observations is the total number of observations available for estimation. The sum of effective observations on either side of the cut-off are those used for estimation. Standard errors are shown in parentheses. Only standard errors for the 'Robust' specification are robust and clustered on Primary FOEI. * significant at 10%, ** significant at 5%, *** significant at 1%.



(a) Low cut-off
(b) Mid-low cut-off
(c) Mid-high cut-off
(d) High cut-off

Figure C.16: HSC Physics score in stacked sharp samples



(a) Low cut-off
(b) Mid-low cut-off
(c) Mid-high cut-off
(d) High cut-off

Figure C.17: HSC Biology score in stacked sharp samples

Table C.17: Fuzzy regression discontinuity estimates of effects of an offer on Physics scores

Specifications	A	B	C	D	E	F	G	H	I	J	K
Conventional	0.252	2.842*	3.648**	1.292	-4.242***	0.896	-0.176	-1.322	2.126*	-3.383***	-2.097*
	(2.262)	(1.593)	(1.765)	(0.990)	(1.220)	(1.814)	(1.707)	(1.292)	(1.151)	(1.215)	(1.089)
Bias-corrected	0.471	2.425	3.726**	1.207	-5.053***	1.353	-0.760	-1.734	2.291**	-3.453***	-2.493**
	(2.262)	(1.593)	(1.765)	(0.990)	(1.220)	(1.814)	(1.707)	(1.292)	(1.151)	(1.215)	(1.089)
Robust	0.471	2.425	3.726*	1.207	-5.053***	1.353	-0.760	-1.734	2.291*	-3.453**	-2.493*
	(2.771)	(1.819)	(2.067)	(1.152)	(1.366)	(2.357)	(2.078)	(1.540)	(1.344)	(1.359)	(1.355)
Observations	182	347	302	446	422	251	657	220	756	733	882
First-stage instrument t-statistic	46.808	11.426	12.096	N/A	10.43	3.404	5.849	7.495	16.278	10.154	16.134
Bandwidth $\gtrless 0$	0.672	0.542	0.614	0.534	0.342	0.846	0.330	0.589	0.300	0.425	0.311
Effective obs. < 0	38	60	69	102	72	70	80	51	97	136	115
Effective obs. ≥ 0	55	76	67	111	69	67	105	45	75	133	134

Specifications	L	M	N	O	P	Q	R	Low	Mid-low	Mid-high	High
Conventional	0.0199	2.829***	-2.759*	0.321	0.0121	1.624*	2.745**	2.765**	-0.757	0.469	2.745**
	(1.964)	(0.667)	(1.521)	(0.843)	(1.475)	(0.857)	(1.124)	(1.116)	(0.487)	(0.425)	(1.124)
Bias-corrected	0.432	2.691***	-3.594**	-0.0977	0.347	1.808**	2.825**	2.608**	-0.809*	0.297	2.825**
	(1.964)	(0.667)	(1.521)	(0.843)	(1.475)	(0.857)	(1.124)	(1.116)	(0.487)	(0.425)	(1.124)
Robust	0.432	2.691***	-3.594*	-0.0977	0.347	1.808*	2.825**	2.608**	-0.809	0.297	2.825**
	(2.495)	(0.804)	(1.871)	(1.048)	(1.709)	(0.983)	(1.393)	(1.300)	(0.558)	(0.491)	(1.393)
Observations	266	1,090	330	793	289	1,197	748	829	3,478	4,838	748
First-stage instrument t-statistic	5.275	48.391	59.65	17.702	5.814	11.931	10.868	21.332	27.333	34.416	10.868
Bandwidth $\gtrless 0$	0.420	0.306	0.267	0.810	0.517	0.306	0.201	0.604	0.482	0.351	0.201
Effective obs. < 0	43	178	43	267	46	175	127	174	697	785	127
Effective obs. ≥ 0	44	141	49	191	85	203	139	197	658	767	139

Each column reports regression discontinuity results for a school or stacked sharp sample. Note, School R and the high cut-off sharp samples are identical. 'Bias-corrected' and 'Robust', but not 'Conventional' specification point estimates are bias-corrected for the selected bandwidth following Calonico et al. (2014b). Estimates are non-parametrically estimated using local linear regressions and a triangular kernel. Optimal bandwidths are calculated for each sharp sample based on the common-bandwidth MSE-optimal selector following Calonico et al. (2014b) and Calonico et al. (2016) as described in the text. Observations is the total number of observations available for estimation. The sum of effective observations on either side of the cut-off are those used for estimation. Standard errors are shown in parentheses. Only standard errors for the 'Robust' specification are robust and clustered on Primary FOEI. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table C.18: Fuzzy regression discontinuity estimates of effects of an offer on Biology scores

Specifications	A	B	C	D	E	F	G	H	I	J	K
Conventional	1.932	-3.894***	1.346	-2.491**	-3.355*	3.455*	1.797	-1.630	-2.369*	-0.751	-2.916**
	(1.509)	(1.388)	(1.088)	(1.058)	(1.856)	(1.859)	(1.706)	(1.179)	(1.315)	(1.745)	(1.145)
Bias-corrected	2.220	-4.406***	1.580	-2.539**	-4.509**	4.433**	2.120	-1.908	-2.859**	-0.661	-3.214***
	(1.509)	(1.388)	(1.088)	(1.058)	(1.856)	(1.859)	(1.706)	(1.179)	(1.315)	(1.745)	(1.145)
Robust	2.220	-4.406***	1.580	-2.539**	-4.509*	4.433**	2.120	-1.908	-2.859*	-0.661	-3.214**
	(1.735)	(1.568)	(1.305)	(1.259)	(2.393)	(2.168)	(2.035)	(1.482)	(1.496)	(2.127)	(1.379)
Observations	231	381	405	481	499	329	366	657	771	338	905
First-stage instrument t-statistic	3.825	N/A	N/A	7.229	3.968	2.685	11.835	28.926	9.445	3.616	5.817
Bandwidth $\gtrless 0$	0.506	0.491	0.755	0.637	0.244	0.572	0.351	0.328	0.151	0.421	0.357
Effective obs. < 0	34	63	92	89	37	49	26	66	27	37	77
Effective obs. ≥ 0	42	52	110	77	54	53	29	36	41	56	104

Specifications	L	M	N	O	P	Q	R	Low	Mid-low	Mid-high	High
Conventional	1.146	-0.631	2.437	1.731	-1.135	0.263	-1.120	-0.350	-0.555	0.181	-1.120
	(1.165)	(1.135)	(1.511)	(1.286)	(1.087)	(1.102)	(1.100)	(0.852)	(0.694)	(0.564)	(1.100)
Bias-corrected	1.315	-1.151	2.918*	2.190*	-1.483	0.669	-1.828*	-0.449	-0.773	0.261	-1.828*
	(1.165)	(1.135)	(1.511)	(1.286)	(1.087)	(1.102)	(1.100)	(0.852)	(0.694)	(0.564)	(1.100)
Robust	1.315	-1.151	2.918*	2.190	-1.483	0.669	-1.828	-0.449	-0.773	0.261	-1.828
	(1.348)	(1.361)	(1.736)	(1.514)	(1.309)	(1.313)	(1.394)	(1.013)	(0.808)	(0.624)	(1.394)
Observations	713	493	873	357	636	928	425	1,014	3,434	4,899	425
First-stage instrument t-statistic	7.01	8.824	15.204	6.46	100.328	6.852	N/A	17.243	13.949	16.079	N/A
Bandwidth $\gtrless 0$	0.289	0.250	0.259	0.482	0.269	0.347	0.194	0.636	0.329	0.210	0.194
Effective obs. < 0	48	48	56	57	36	108	38	199	298	271	38
Effective obs. ≥ 0	53	41	54	41	41	65	58	206	329	315	58

Each column reports regression discontinuity results for a school or stacked sharp sample. Note, School R and the high cut-off sharp samples are identical. 'Bias-corrected' and 'Robust', but not 'Conventional' specification point estimates are bias-corrected for the selected bandwidth following Calonico et al. (2014b). Estimates are non-parametrically estimated using local linear regressions and a triangular kernel. Optimal bandwidths are calculated for each sharp sample based on the common-bandwidth MSE-optimal selector following Calonico et al. (2014b) and Calonico et al. (2016) as described in the text. Observations is the total number of observations available for estimation. The sum of effective observations on either side of the cut-off are those used for estimation. Standard errors are shown in parentheses. Only standard errors for the 'Robust' specification are robust and clustered on Primary FOEI. * significant at 10%, ** significant at 5%, *** significant at 1%.

Appendix D

Alternative regression specifications

D.1 Peer effects estimation with Year 5 NAPLAN as baseline peer achievement

Table D.1: Reading (standardised Year 5 NAPLAN) peer effects on TES

Specifications	A	B	C	D	E	F	G	H	I	J	K
Conventional	-8.986	28.10	83.96**	-7.894	0.00925	52.73	-53.07	-36.25	-2.919	10.60	-8.435
	(15.89)	(18.87)	(35.45)	(18.85)	(14.67)	(52.00)	(56.94)	(45.40)	(114.6)	(30.10)	(60.25)
Bias-corrected	-13.27	27.52	112.1***	-12.86	-2.406	70.55	-16.66	-46.32	-25.57	15.13	-40.38
	(15.89)	(18.87)	(35.45)	(18.85)	(14.67)	(52.00)	(56.94)	(45.40)	(114.6)	(30.10)	(60.25)
Robust	-13.27	27.52	112.1***	-12.86	-2.406	70.55	-16.66	-46.32	-25.57	15.13	-40.38
	(18.26)	(22.20)	(41.26)	(24.15)	(18.16)	(62.86)	(63.08)	(53.45)	(127.5)	(36.47)	(73.58)
Observations	224	388	308	436	490	391	458	576	691	486	834
First-stage instrument t-statistic	5.658	4.08	2.139	3.789	6.383	1.787	1.854	1.694	0.555	3.158	1.168
Bandwidth $\gtrless 0$	0.701	0.588	0.718	0.490	0.388	0.621	0.188	0.264	0.198	0.405	0.372
Effective obs. < 0	42	79	75	70	74	68	22	42	42	72	107
Effective obs. ≥ 0	62	76	71	77	79	75	34	60	49	77	107

Specifications	L	M	N	O	P	Q	R	Low	Mid-low	Mid-high	High
Conventional	-17.14	91.31**	-69.34	25.17	-174.6	8.919	37.78**	33.76**	-0.0316	0.216	37.78**
	(27.45)	(42.89)	(75.62)	(40.31)	(263.1)	(61.65)	(18.74)	(13.31)	(10.95)	(22.69)	(18.74)
Bias-corrected	-13.37	97.82**	-58.80	19.19	-217.3	7.813	37.20**	37.35***	-2.183	1.541	37.20**
	(27.45)	(42.89)	(75.62)	(40.31)	(263.1)	(61.65)	(18.74)	(13.31)	(10.95)	(22.69)	(18.74)
Robust	-13.37	97.82**	-58.80	19.19	-217.3	7.813	37.20*	37.35**	-2.183	1.541	37.20*
	(32.84)	(49.47)	(98.46)	(44.36)	(314.2)	(71.23)	(22.52)	(14.98)	(13.07)	(24.45)	(22.52)
Observations	586	699	741	522	597	987	512	917	3,522	4,960	512
First-stage instrument t-statistic	2.317	6.614	3.062	9.686	2.37	3.352	23.995	7.321	6.196	5.84	23.995
Bandwidth $\gtrless 0$	0.244	0.303	0.346	0.272	0.281	0.386	0.294	0.584	0.447	0.259	0.294
Effective obs. < 0	38	93	88	66	54	150	114	178	537	477	114
Effective obs. ≥ 0	56	82	91	69	68	136	71	190	542	515	71

Each column reports regression discontinuity results for a school or stacked sharp sample. Note, School R and the high cut-off sharp samples are identical. 'Bias-corrected' and 'Robust', but not 'Conventional' specification point estimates are bias-corrected for the selected bandwidth following Calonico et al. (2014b). Estimates are non-parametrically estimated using local linear regressions and a triangular kernel. Optimal bandwidths are calculated for each sharp sample based on the common-bandwidth MSE-optimal selector following Calonico et al. (2014b) and Calonico et al. (2016) as described in the text. Observations is the total number of observations available for estimation. The sum of effective observations on either side of the cut-off are those used for estimation. Standard errors are shown in parentheses. Only standard errors for the 'Robust' specification are robust and clustered on Primary FOEI. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table D.2: Numeracy (standardised Year 5 NAPLAN) peer effects on TES

Specifications	A	B	C	D	E	F	G	H	I	J	K
Conventional	-13.93	37.26	107.9**	-8.815	-1.069	44.77	-115.0	-30.56	4.239	10.41	-1.233
	(23.29)	(22.76)	(48.04)	(20.51)	(13.19)	(42.54)	(108.8)	(32.34)	(135.0)	(26.18)	(85.02)
Bias-corrected	-19.91	38.79*	137.9***	-13.92	-1.461	54.76	-98.85	-30.35	-23.02	13.49	-45.35
	(23.29)	(22.76)	(48.04)	(20.51)	(13.19)	(42.54)	(108.8)	(32.34)	(135.0)	(26.18)	(85.02)
Robust	-19.91	38.79	137.9**	-13.92	-1.461	54.76	-98.85	-30.35	-23.02	13.49	-45.35
	(26.50)	(26.77)	(54.45)	(25.39)	(15.39)	(50.52)	(123.2)	(38.14)	(153.5)	(31.27)	(104.1)
Observations	224	388	308	436	490	391	458	576	691	486	834
First-stage instrument t-statistic	5.466	4.334	2.125	3.691	7.089	2.335	1.222	3.378	0.277	3.308	.65
Bandwidth $\gtrless 0$	0.675	0.667	0.677	0.473	0.356	0.662	0.275	0.380	0.206	0.372	0.399
Effective obs. < 0	40	92	74	69	67	73	36	63	43	64	117
Effective obs. ≥ 0	62	84	67	72	75	76	60	70	51	72	109

Specifications	L	M	N	O	P	Q	R	Low	Mid-low	Mid-high	High
Conventional	-15.15	62.82**	-102.5	27.91	-121.7	8.782	57.72**	42.71***	-0.0645	-0.485	57.72**
	(55.25)	(28.38)	(115.2)	(33.69)	(161.9)	(59.83)	(27.58)	(15.31)	(9.978)	(21.40)	(27.58)
Bias-corrected	-1.722	72.61**	-85.97	17.64	-120.5	3.795	58.93**	47.74***	-2.000	0.712	58.93**
	(55.25)	(28.38)	(115.2)	(33.69)	(161.9)	(59.83)	(27.58)	(15.31)	(9.978)	(21.40)	(27.58)
Robust	-1.722	72.61**	-85.97	17.64	-120.5	3.795	58.93*	47.74***	-2.000	0.712	58.93*
	(65.32)	(34.54)	(148.3)	(37.88)	(181.2)	(66.77)	(33.36)	(18.41)	(11.88)	(22.90)	(33.36)
Observations	586	699	741	522	597	987	512	917	3,522	4,960	512
First-stage instrument t-statistic	0.818	10.853	2.493	10.949	2.753	2.801	25.651	7.385	6.61	5.756	25.651
Bandwidth $\gtrless 0$	0.201	0.322	0.337	0.310	0.254	0.364	0.317	0.705	0.446	0.249	0.317
Effective obs. < 0	28	99	82	71	49	143	127	212	537	453	127
Effective obs. ≥ 0	45	87	91	73	63	129	76	214	540	491	76

Each column reports regression discontinuity results for a school or stacked sharp sample. Note, School R and the high cut-off sharp samples are identical. 'Bias-corrected' and 'Robust', but not 'Conventional' specification point estimates are bias-corrected for the selected bandwidth following Calonico et al. (2014b). Estimates are non-parametrically estimated using local linear regressions and a triangular kernel. Optimal bandwidths are calculated for each sharp sample based on the common-bandwidth MSE-optimal selector following Calonico et al. (2014b) and Calonico et al. (2016) as described in the text. Observations is the total number of observations available for estimation. The sum of effective observations on either side of the cut-off are those used for estimation. Standard errors are shown in parentheses. Only standard errors for the 'Robust' specification are robust and clustered on Primary FOEI. * significant at 10%, ** significant at 5%, *** significant at 1%.

D.2 Alternate bandwidth selectors

D.2.1 Separate mean-square error (MSE-optimal) selectors

Table D.3: Fuzzy regression discontinuity estimates of effects of an offer on TES (separate MSE-optimal bandwidth selectors)

Specifications	A	B	C	D	E	F	G	H	I	J	K
Conventional	6.168	9.880	39.90**	11.79	0.103	32.92**	-9.157	-15.03*	-1.798	-6.298	-9.757
	(14.23)	(13.37)	(16.84)	(11.13)	(8.086)	(15.11)	(11.87)	(8.871)	(11.60)	(14.17)	(7.893)
Bias-corrected	8.698	6.058	45.55***	12.70	-2.574	42.35***	-8.090	-21.08**	-8.623	-6.052	-13.32*
	(14.23)	(13.37)	(16.84)	(11.13)	(8.086)	(15.11)	(11.87)	(8.871)	(11.60)	(14.17)	(7.893)
Robust	8.698	6.058	45.55**	12.70	-2.574	42.35**	-8.090	-21.08**	-8.623	-6.052	-13.32
	(16.91)	(15.83)	(20.32)	(12.98)	(9.210)	(18.47)	(14.03)	(10.24)	(13.49)	(16.82)	(9.765)
Observations	697	1,234	1,022	1,511	1,533	1,098	1,408	1,728	2,222	1,494	2,711
First-stage instrument t-statistic	10.038	24.717	27.756	11.066	16.963	8.463	7.619	8.345	25.849	12.097	23.321
Bandwidth < 0	0.820	0.676	0.730	0.775	0.992	0.826	0.769	0.856	0.779	0.861	1.011
Bandwidth ≥ 0	0.920	0.435	0.512	0.468	0.542	0.464	0.335	0.327	0.199	0.282	0.316
Effective obs. < 0	183	284	257	346	497	265	309	413	579	434	820
Effective obs. ≥ 0	213	179	197	253	268	152	197	193	138	177	306

Specifications	L	M	N	O	P	Q	R	Low	Mid-low	Mid-high	High
Conventional	7.263	9.151	3.740	13.33	-0.000274	16.35**	11.76**	23.03**	2.657	3.558	11.76**
	(7.230)	(6.625)	(5.757)	(10.08)	(7.585)	(6.785)	(5.661)	(9.542)	(4.084)	(3.679)	(5.661)
Bias-corrected	10.27	9.928	3.747	14.64	-4.975	18.89***	11.71**	24.11**	1.007	3.690	11.71**
	(7.230)	(6.625)	(5.757)	(10.08)	(7.585)	(6.785)	(5.661)	(9.542)	(4.084)	(3.679)	(5.661)
Robust	10.27	9.928	3.747	14.64	-4.975	18.89**	11.71*	24.11**	1.007	3.690	11.71*
	(8.458)	(7.785)	(7.014)	(11.95)	(9.393)	(7.748)	(6.276)	(11.29)	(4.847)	(4.237)	(6.276)
Observations	1,805	2,216	2,340	1,610	1,805	3,081	1,623	2,945	10,977	15,548	1,623
First-stage instrument t-statistic	10.421	29.893	28.396	13.879	17.769	15.884	80.609	27.911	23.742	37.11	80.609
Bandwidth < 0	1.249	0.759	1.169	0.843	0.895	0.902	0.592	0.718	0.692	0.595	0.592
Bandwidth ≥ 0	0.489	0.432	0.294	0.267	0.425	0.269	0.240	0.470	0.420	0.240	0.240
Effective obs. < 0	696	625	790	451	435	987	527	703	2331	3082	527
Effective obs. ≥ 0	248	333	260	183	260	286	220	500	1546	1447	220

Each column reports regression discontinuity results for a school or stacked sharp sample. Note, School R and the high cut-off sharp samples are identical. 'Bias-corrected' and 'Robust', but not 'Conventional' specification point estimates are bias-corrected for the selected bandwidth following Calonico et al. (2014b). Estimates are non-parametrically estimated using local linear regressions and a triangular kernel. Optimal bandwidths are calculated for each sharp sample based on the MSE-optimal selector separately estimated above and below the cut-off following Calonico et al. (2014b) and Calonico et al. (2016) as described in the text. Observations is the total number of observations available for estimation. The sum of effective observations on either side of the cut-off are those used for estimation. Standard errors are shown in parentheses. Only standard errors for the 'Robust' specification are robust and clustered on Primary FOEI. * significant at 10%, ** significant at 5%, *** significant at 1%.

D.2.2 Common coverage error (CER-optimal) selector

Table D.4: Fuzzy regression discontinuity estimates of effects of an offer on TES (common CER-optimal bandwidth selector)

Specifications	A	B	C	D	E	F	G	H	I	J	K
Conventional	11.24	-18.81	41.28*	5.602	-13.03	49.44**	-1.447	-22.46*	-22.77	-1.821	-17.63
	(16.61)	(16.68)	(22.93)	(16.06)	(11.14)	(22.57)	(18.63)	(12.41)	(15.20)	(17.57)	(12.02)
Bias-corrected	13.14	-22.84	42.83*	5.092	-15.51	52.58**	-3.406	-25.29**	-25.43*	-2.775	-20.05*
	(16.61)	(16.68)	(22.93)	(16.06)	(11.14)	(22.57)	(18.63)	(12.41)	(15.20)	(17.57)	(12.02)
Robust	13.14	-22.84	42.83*	5.092	-15.51	52.58**	-3.406	-25.29*	-25.43	-2.775	-20.05
	(18.36)	(18.37)	(25.25)	(17.25)	(12.24)	(24.19)	(20.61)	(13.76)	(16.46)	(19.63)	(13.23)
Observations	697	1,234	1,022	1,511	1,533	1,098	1,408	1,728	2,222	1,494	2,711
First-stage instrument t-statistic	11.572	19.132	13.782	9.711	15.21	5.658	7.329	10.425	13.563	14.142	18.832
Bandwidth $\gtrless 0$	0.671	0.291	0.376	0.294	0.353	0.345	0.253	0.319	0.147	0.300	0.224
Effective obs. < 0	144	117	132	131	173	81	96	151	94	155	187
Effective obs. ≥ 0	176	127	145	160	205	115	153	189	107	189	242

Specifications	L	M	N	O	P	Q	R	Low	Mid-low	Mid-high	High
Conventional	4.957	9.464	-1.795	8.284	-15.33	20.79**	13.57*	11.99	-2.644	-0.908	13.57*
	(9.373)	(7.528)	(10.13)	(11.67)	(10.99)	(10.01)	(8.230)	(13.03)	(5.619)	(5.775)	(8.230)
Bias-corrected	5.232	9.336	-1.896	10.04	-18.66*	22.05**	14.01*	10.66	-3.760	-1.280	14.01*
	(9.373)	(7.528)	(10.13)	(11.67)	(10.99)	(10.01)	(8.230)	(13.03)	(5.619)	(5.775)	(8.230)
Robust	5.232	9.336	-1.896	10.04	-18.66	22.05**	14.01	10.66	-3.760	-1.280	14.01
	(10.39)	(8.347)	(11.62)	(13.28)	(12.07)	(10.69)	(9.231)	(14.47)	(6.155)	(6.107)	(9.231)
Observations	1,805	2,216	2,340	1,610	1,805	3,081	1,623	2,945	10,977	15,548	1,623
First-stage instrument t-statistic	17.381	33.911	23.24	15.312	15.807	15.119	20.697	21.233	30.12	33.243	20.697
Bandwidth $\gtrless 0$	0.353	0.380	0.234	0.309	0.317	0.169	0.166	0.343	0.333	0.158	0.166
Effective obs. < 0	162	353	181	212	165	180	178	342	1079	864	178
Effective obs. ≥ 0	228	305	239	200	202	185	191	368	1284	1008	191

Each column reports regression discontinuity results for a school or stacked sharp sample. Note, School R and the high cut-off sharp samples are identical. 'Bias-corrected' and 'Robust', but not 'Conventional' specification point estimates are bias-corrected for the selected bandwidth following Calonico et al. (2014b). Estimates are non-parametrically estimated using local linear regressions and a triangular kernel. Optimal bandwidths are calculated for each sharp sample based on the common-bandwidth CER-optimal selector following Calonico et al. (2016) as described in the text. Observations is the total number of observations available for estimation. The sum of effective observations on either side of the cut-off are those used for estimation. Standard errors are shown in parentheses. Only standard errors for the 'Robust' specification are robust and clustered on Primary FOEI. * significant at 10%, ** significant at 5%, *** significant at 1%.

D.2.3 Separate coverage error (CER-optimal) selectors

Table D.5: Fuzzy regression discontinuity estimates of effects of an offer on TES (separate CER-optimal bandwidth selectors)

Specifications	A	B	C	D	E	F	G	H	I	J	K
Conventional	10.48	2.118	41.71**	9.316	-1.476	36.55**	1.848	-26.19**	-11.08	-3.862	-13.67
	(16.90)	(15.05)	(19.75)	(12.37)	(9.082)	(18.12)	(13.83)	(10.57)	(13.47)	(16.47)	(9.102)
Bias-corrected	12.03	-0.210	45.02**	9.964	-3.229	42.18**	2.628	-29.92***	-14.88	-3.705	-15.68*
	(16.90)	(15.05)	(19.75)	(12.37)	(9.082)	(18.12)	(13.83)	(10.57)	(13.47)	(16.47)	(9.102)
Robust	12.03	-0.210	45.02**	9.964	-3.229	42.18**	2.628	-29.92***	-14.88	-3.705	-15.68
	(18.69)	(16.85)	(21.93)	(13.56)	(10.02)	(20.17)	(15.31)	(11.59)	(14.63)	(18.14)	(10.26)
Observations	697	1,234	1,022	1,511	1,533	1,098	1,408	1,728	2,222	1,494	2,711
First-stage instrument t-statistic	10.038	24.717	27.756	11.066	16.963	8.463	7.619	8.345	25.849	12.097	23.321
Bandwidth < 0	0.639	0.518	0.554	0.582	0.754	0.643	0.575	0.638	0.586	0.651	0.744
Bandwidth ≥ 0	0.717	0.333	0.389	0.351	0.412	0.361	0.250	0.244	0.150	0.213	0.232
Effective obs. < 0	133	210	197	262	352	198	226	302	441	331	620
Effective obs. ≥ 0	183	149	148	196	230	118	149	155	109	136	247

Specifications	L	M	N	O	P	Q	R	Low	Mid-low	Mid-high	High
Conventional	9.277	9.930	1.185	9.983	-6.784	19.34**	10.86*	16.49	0.920	1.482	10.86*
	(8.112)	(7.395)	(6.477)	(11.69)	(8.821)	(7.521)	(5.643)	(11.49)	(4.982)	(4.494)	(5.643)
Bias-corrected	11.13	10.39	1.239	10.72	-9.798	20.74***	10.72*	17.03	-0.0136	1.544	10.72*
	(8.112)	(7.395)	(6.477)	(11.69)	(8.821)	(7.521)	(5.643)	(11.49)	(4.982)	(4.494)	(5.643)
Robust	11.13	10.39	1.239	10.72	-9.798	20.74**	10.72*	17.03	-0.0136	1.544	10.72*
	(9.038)	(8.153)	(7.336)	(12.83)	(9.960)	(8.105)	(5.974)	(12.60)	(5.431)	(4.826)	(5.974)
Observations	1,805	2,216	2,340	1,610	1,805	3,081	1,623	2,945	10,977	15,548	1,623
First-stage instrument t-statistic	10.421	29.893	28.396	13.879	17.769	15.884	80.609	27.911	23.742	37.11	80.609
Bandwidth < 0	0.937	0.557	0.859	0.629	0.664	0.673	0.445	0.524	0.492	0.425	0.445
Bandwidth ≥ 0	0.366	0.317	0.216	0.200	0.315	0.201	0.180	0.343	0.299	0.172	0.180
Effective obs. < 0	499	498	600	369	335	756	438	506	1663	2304	438
Effective obs. ≥ 0	234	278	225	146	202	214	198	368	1172	1099	198

Each column reports regression discontinuity results for a school or stacked sharp sample. Note, School R and the high cut-off sharp samples are identical. 'Bias-corrected' and 'Robust', but not 'Conventional' specification point estimates are bias-corrected for the selected bandwidth following Calonico et al. (2014b). Estimates are non-parametrically estimated using local linear regressions and a triangular kernel. Optimal bandwidths are calculated for each sharp sample based on the CER-optimal selector separately estimated above and below the cut-off following Calonico et al. (2016) as described in the text. Observations is the total number of observations available for estimation. The sum of effective observations on either side of the cut-off are those used for estimation. Standard errors are shown in parentheses. Only standard errors for the 'Robust' specification are robust and clustered on Primary FOEI. * significant at 10%, ** significant at 5%, *** significant at 1%.

D.3 Other alternate regression specifications

D.3.1 Parametric (quadratic), over full sample, uniform kernel

Table D.6: Fuzzy regression discontinuity estimates of effects of an offer on TES (parametric)

Specifications	A	B	C	D	E	F	G	H	I	J	K
Conventional	2.686	22.34***	27.51***	5.879	10.49	7.143	5.617	5.071	21.62***	-3.856	-7.787
	(12.62)	(8.529)	(10.27)	(8.657)	(6.958)	(11.58)	(9.080)	(8.174)	(7.051)	(9.626)	(6.624)
Bias-corrected	3.644	21.89**	35.39***	4.954	0.261	17.41	-1.460	-9.057	12.35*	-10.16	-8.605
	(12.62)	(8.529)	(10.27)	(8.657)	(6.958)	(11.58)	(9.080)	(8.174)	(7.051)	(9.626)	(6.624)
Robust	3.644	21.89*	35.39**	4.954	0.261	17.41	-1.460	-9.057	12.35	-10.16	-8.605
	(12.78)	(11.33)	(14.78)	(10.96)	(8.946)	(13.22)	(12.23)	(9.006)	(10.13)	(14.39)	(8.104)
Observations	697	1,234	1,022	1,511	1,533	1,098	1,455	1,774	2,222	1,533	2,711
First-stage instrument t-statistic	14.588	32.654	18.207	10.372	22.509	7.289	9.641	15.57	15.118	19.205	21.51

Specifications	L	M	N	O	P	Q	R	Low	Mid-low	Mid-high	High
Conventional	3.610	8.549	3.716	11.02	10.88	11.64**	11.11**	19.79***	7.723**	4.249	11.11**
	(7.340)	(5.859)	(4.960)	(6.802)	(7.120)	(5.528)	(4.846)	(6.007)	(3.053)	(2.787)	(4.846)
Bias-corrected	11.24	6.861	3.931	11.87*	3.606	12.45**	11.12**	23.68***	2.498	5.278*	11.12**
	(7.340)	(5.859)	(4.960)	(6.802)	(7.120)	(5.528)	(4.846)	(6.007)	(3.053)	(2.787)	(4.846)
Robust	11.24	6.861	3.931	11.87	3.606	12.45*	11.12*	23.68***	2.498	5.278	11.12*
	(9.168)	(7.368)	(6.448)	(9.056)	(8.486)	(7.403)	(6.176)	(7.818)	(3.685)	(3.511)	(6.176)
Observations	1,857	2,286	2,416	1,661	1,868	3,081	1,623	2,945	10,977	15,548	1,623
First-stage instrument t-statistic	17.346	49.297	24.201	18.074	16.462	24.388	28.356	26.038	53.86	44.633	28.356

Each column reports regression discontinuity results for a school or stacked sharp sample. Note, School R and the high cut-off sharp samples are identical. 'Bias-corrected' and 'Robust', but not 'Conventional' specification point estimates are bias-corrected for the selected bandwidth following Calonico et al. (2014b). Estimates are parametrically estimated over the entire sharp sample using a quadratic polynomial that is allowed to differ on either side of the cut-off and a uniform kernel. Observations is the total number of observations available and used for estimation. Standard errors are shown in parentheses. Only standard errors for the 'Robust' specification are robust and clustered on Primary FOEI. * significant at 10%, ** significant at 5%, *** significant at 1%.

D.3.2 Uniform kernel

Table D.7: Fuzzy regression discontinuity estimates of effects of an offer on TES (uniform kernel)

Specifications	A	B	C	D	E	F	G	H	I	J	K
Conventional	3.080	21.73*	42.97**	5.296	-5.900	51.38**	3.188	-8.244	5.051	-6.107	-16.63
	(14.35)	(11.21)	(19.31)	(16.94)	(8.787)	(20.20)	(17.30)	(10.98)	(13.03)	(15.31)	(10.71)
Bias-corrected	5.407	22.55**	46.21**	4.065	-8.932	54.92***	-3.414	-17.61	4.800	-5.774	-21.49**
	(14.35)	(11.21)	(19.31)	(16.94)	(8.787)	(20.20)	(17.30)	(10.98)	(13.03)	(15.31)	(10.71)
Robust	5.407	22.55*	46.21**	4.065	-8.932	54.92**	-3.414	-17.61	4.800	-5.774	-21.49*
	(16.64)	(13.65)	(22.57)	(18.46)	(9.907)	(22.84)	(20.41)	(13.00)	(15.41)	(18.57)	(12.74)
Observations	697	1,234	1,022	1,511	1,533	1,098	1,455	1,774	2,222	1,533	2,711
First-stage instrument t-statistic	14.588	32.654	18.207	10.372	22.509	7.289	9.641	15.57	15.118	19.205	21.51
Bandwidth $\gtrless 0$	0.704	0.651	0.425	0.246	0.391	0.381	0.308	0.428	0.203	0.337	0.209
Effective obs. < 0	153	266	144	113	189	91	119	213	145	174	177
Effective obs. ≥ 0	182	255	159	127	225	123	184	239	142	217	231

Specifications	L	M	N	O	P	Q	R	Low	Mid-low	Mid-high	High
Conventional	2.565	12.27**	0.314	-0.949	-10.41	14.33	8.707	19.62*	6.182*	1.487	8.707
	(10.17)	(6.239)	(9.997)	(9.495)	(10.61)	(8.870)	(7.989)	(11.40)	(3.611)	(4.576)	(7.989)
Bias-corrected	2.169	13.82**	-1.969	-2.586	-16.05	14.56	11.17	20.02*	6.025*	0.924	11.17
	(10.17)	(6.239)	(9.997)	(9.495)	(10.61)	(8.870)	(7.989)	(11.40)	(3.611)	(4.576)	(7.989)
Robust	2.169	13.82*	-1.969	-2.586	-16.05	14.56	11.17	20.02	6.025	0.924	11.17
	(11.93)	(7.305)	(11.97)	(11.96)	(12.21)	(9.776)	(9.314)	(13.55)	(4.224)	(5.156)	(9.314)
Observations	1,857	2,286	2,416	1,661	1,868	3,081	1,623	2,945	10,977	15,548	1,623
First-stage instrument t-statistic	17.346	49.297	24.201	18.074	16.462	24.388	28.356	26.038	53.86	44.633	28.356
Bandwidth $\gtrless 0$	0.270	0.593	0.225	0.354	0.270	0.221	0.176	0.363	0.558	0.188	0.176
Effective obs. < 0	123	524	169	237	152	254	184	357	1897	1026	184
Effective obs. ≥ 0	178	389	235	217	174	235	197	399	1845	1191	197

Each column reports regression discontinuity results for a school or stacked sharp sample. Note, School R and the high cut-off sharp samples are identical. ‘Bias-corrected’ and ‘Robust’, but not ‘Conventional’ specification point estimates are bias-corrected for the selected bandwidth following Calonico et al. (2014b). Estimates are non-parametrically estimated using local linear regressions and a uniform kernel. Optimal bandwidths are calculated for each sharp sample based on the common-bandwidth MSE-optimal selector following Calonico et al. (2014b) and Calonico et al. (2016) as described in the text. Observations is the total number of observations available for estimation. The sum of effective observations on either side of the cut-off are those used for estimation. Standard errors are shown in parentheses. Only standard errors for the ‘Robust’ specification are robust and clustered on Primary FOEI. * significant at 10%, ** significant at 5%, *** significant at 1%.

D.3.3 No covariates

Table D.8: Fuzzy regression discontinuity estimates of effects of an offer on TES (no covariates)

Specifications	A	B	C	D	E	F	G	H	I	J	K
Conventional	18.90	-6.472	34.46	-7.593	-6.092	55.43**	6.396	-11.11	-30.07*	-8.011	-17.76*
	(16.59)	(17.84)	(21.10)	(13.78)	(9.833)	(22.73)	(17.65)	(10.92)	(17.12)	(16.09)	(10.39)
Bias-corrected	23.32	-14.64	34.92*	-12.30	-9.524	60.69***	4.482	-19.73*	-37.44**	-10.95	-23.24**
	(16.59)	(17.84)	(21.10)	(13.78)	(9.833)	(22.73)	(17.65)	(10.92)	(17.12)	(16.09)	(10.39)
Robust	23.32	-14.64	34.92	-12.30	-9.524	60.69**	4.482	-19.73	-37.44*	-10.95	-23.24*
	(19.65)	(20.21)	(24.58)	(15.56)	(11.27)	(25.08)	(20.90)	(13.20)	(19.47)	(19.48)	(12.56)
Observations	721	1,280	1,052	1,639	1,588	1,179	1,455	1,774	2,342	1,533	2,804
First-stage instrument t-statistic	10.761	18.276	16.841	12.255	17.471	6.032	8.271	13.072	13.273	16.766	21.255
Bandwidth $\gtrless 0$	0.630	0.366	0.513	0.406	0.437	0.424	0.348	0.498	0.171	0.421	0.312
Effective obs. < 0	138	153	186	200	215	118	142	250	116	226	274
Effective obs. ≥ 0	168	166	202	221	243	142	206	263	123	253	304

Specifications	L	M	N	O	P	Q	R	Low	Mid-low	Mid-high	High
Conventional	6.823	11.68*	-1.845	3.309	-8.210	18.44**	11.48*	20.24	0.746	-1.521	11.48*
	(8.802)	(6.774)	(5.574)	(8.834)	(9.691)	(8.015)	(6.979)	(13.08)	(5.113)	(4.655)	(6.979)
Bias-corrected	7.009	12.29*	-3.905	0.400	-11.73	20.50**	12.65*	18.21	-0.821	-2.508	12.65*
	(8.802)	(6.774)	(5.574)	(8.834)	(9.691)	(8.015)	(6.979)	(13.08)	(5.113)	(4.655)	(6.979)
Robust	7.009	12.29	-3.905	0.400	-11.73	20.50**	12.65	18.21	-0.821	-2.508	12.65
	(10.36)	(8.100)	(6.945)	(11.17)	(11.37)	(9.174)	(8.651)	(15.28)	(5.693)	(5.261)	(8.651)
Observations	1,857	2,286	2,416	1,661	1,868	3,195	1,658	3,044	11,493	16,067	1,658
First-stage instrument t-statistic	18.733	37.937	39.478	17.328	16.318	18.241	24.277	22.501	38.053	39.534	24.277
Bandwidth $\gtrless 0$	0.425	0.556	0.744	0.485	0.379	0.231	0.240	0.398	0.457	0.219	0.240
Effective obs. < 0	214	500	532	310	203	264	257	393	1576	1202	257
Effective obs. ≥ 0	243	379	357	257	231	247	221	440	1662	1350	221

Each column reports regression discontinuity results for a school or stacked sharp sample. Note, School R and the high cut-off sharp samples are identical. 'Bias-corrected' and 'Robust', but not 'Conventional' specification point estimates are bias-corrected for the selected bandwidth following Calonico et al. (2014b). Estimates are non-parametrically estimated using local linear regressions and a triangular kernel. Optimal bandwidths are calculated for each sharp sample based on the common-bandwidth MSE-optimal selector following Calonico et al. (2014b) and Calonico et al. (2016) as described in the text. Observations is the total number of observations available for estimation. The sum of effective observations on either side of the cut-off are those used for estimation. Standard errors are shown in parentheses. Only standard errors for the 'Robust' specification are robust and clustered on Primary FOEI. * significant at 10%, ** significant at 5%, *** significant at 1%.

D.4 Acceptance of selective school offer as treatment indicator

Table D.9: Fuzzy regression discontinuity estimates of effects of accepting a selective school offer, compared to non-acceptance or non-offer, on TES

Specifications	A	B	C	D	E	F	G	H	I	J	K
Conventional	16.27	18.45	53.30**	4.226	-11.45	83.04**	6.126	-32.17*	1.471	-24.35	-15.88
	(24.38)	(16.73)	(25.17)	(18.14)	(11.09)	(35.90)	(16.55)	(17.66)	(13.36)	(26.34)	(9.687)
Bias-corrected	24.74	13.37	60.50**	2.177	-16.31	99.10***	5.612	-41.44**	-3.047	-33.50	-20.65**
	(24.38)	(16.73)	(25.17)	(18.14)	(11.09)	(35.90)	(16.55)	(17.66)	(13.36)	(26.34)	(9.687)
Robust	24.74	13.37	60.50**	2.177	-16.31	99.10**	5.612	-41.44**	-3.047	-33.50	-20.65*
	(30.39)	(20.37)	(30.76)	(20.63)	(12.98)	(39.69)	(19.69)	(20.47)	(16.02)	(30.97)	(11.80)
Observations	697	1,234	1,022	1,511	1,533	1,098	1,455	1,774	2,222	1,533	2,711
First-stage instrument t-statistic	4.634	8.460	6.547	6.269	11.181	2.558	8.197	6.878	8.790	7.03	17.06
Bandwidth $\gtrless 0$	0.684	0.643	0.708	0.346	0.406	0.446	0.410	0.266	0.302	0.279	0.344
Effective obs. < 0	146	263	250	150	201	121	166	124	216	147	295
Effective obs. ≥ 0	179	252	252	193	229	147	240	170	197	180	319

Specifications	L	M	N	O	P	Q	R	Low	Mid-low	Mid-high	High
Conventional	7.588	13.10*	-1.867	5.301	-4.400	20.79**	12.78*	31.55**	-5.146	0.528	12.78*
	(11.62)	(7.490)	(7.402)	(8.178)	(9.599)	(8.419)	(7.551)	(14.45)	(7.226)	(4.514)	(7.551)
Bias-corrected	6.963	14.09*	-2.019	1.313	-10.08	23.53***	13.93*	31.71**	-8.410	-0.259	13.93*
	(11.62)	(7.490)	(7.402)	(8.178)	(9.599)	(8.419)	(7.551)	(14.45)	(7.226)	(4.514)	(7.551)
Robust	6.963	14.09	-2.019	1.313	-10.08	23.53**	13.93	31.71*	-8.410	-0.259	13.93
	(13.45)	(8.928)	(9.211)	(10.21)	(11.47)	(9.504)	(8.672)	(17.11)	(8.148)	(5.099)	(8.672)
Observations	1,857	2,286	2,416	1,661	1,868	3,081	1,623	2,945	10,977	15,548	1,623
First-stage instrument t-statistic	10.469	18.907	19.69	15.065	10.947	13.764	21.806	10.365	16.426	33.592	21.806
Bandwidth $\gtrless 0$	0.258	0.572	0.444	0.656	0.546	0.226	0.211	0.570	0.297	0.267	0.211
Effective obs. < 0	119	510	348	387	272	264	226	550	959	1460	226
Effective obs. ≥ 0	171	382	310	309	310	241	212	592	1168	1565	212

Each column reports regression discontinuity results for a school or stacked sharp sample. Note, School R and the high cut-off sharp samples are identical. 'Bias-corrected' and 'Robust', but not 'Conventional' specification point estimates are bias-corrected for the selected bandwidth following Calonico et al. (2014b). Estimates are non-parametrically estimated using local linear regressions and a triangular kernel. Optimal bandwidths are calculated for each sharp sample based on the common-bandwidth MSE-optimal selector following Calonico et al. (2014b) and Calonico et al. (2016) as described in the text. Observations is the total number of observations available for estimation. The sum of effective observations on either side of the cut-off are those used for estimation. Standard errors are shown in parentheses. Only standard errors for the 'Robust' specification are robust and clustered on Primary FOEI. * significant at 10%, ** significant at 5%, *** significant at 1%.

Bibliography

Abdulkadiroğlu, A., J. Angrist, and P. Pathak (2014, January). The Elite Illusion: Achievement Effects at Boston and New York Exam Schools. *Econometrica 82*(1), 137–196.

Abdulkadiroğlu, A., P. A. Pathak, A. E. Roth, and T. Sönmez (2005, May). The Boston Public School Match. *American Economic Review 95*(2), 368–371.

Angrist, J. D. (2014). The perils of peer effects. *Labour Economics 30*(C), 98–108.

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

Arcidiacono, P., G. Foster, N. Goodpaster, and J. Kinsler (2012, November). Estimating spillovers using panel data, with an application to the classroom. *Quantitative Economics 3*(3), 421–470.

Atkinson, A., P. Gregg, and B. McConnell (2006, April). The Result of 11 Plus Selection: An Investigation into Opportunities and Outcomes for Pupils in Selective LEAs. The Centre for Market and Public Organisation 06/150, Department of Economics, University of Bristol, UK.

Austen-Smith, D. and R. G. Fryer (2005). An economic analysis of "acting white". *The Quarterly Journal of Economics 120*(2), 551–583.

Australian Bureau of Statistics (2013). 2011 Census QuickStats. `http://www.censusdata.abs.gov.au/census_services/getproduct/census/2011/quickstat/1?opendocument&navpos=220`. All people — usual residents, New South Wales Code 1 (STE). Accessed: 2016-10-07.

Australian Curriculum Assessment and Reporting Authority (2008). 2008 NAPLAN national report. Technical report.

Australian Curriculum Assessment and Reporting Authority (2009). 2009 NAPLAN national report. Technical report.

Australian Curriculum Assessment and Reporting Authority (2010). 2010 NAPLAN national report. Technical report.

Australian Curriculum Assessment and Reporting Authority (2014, March). ICSEA 2013: Technical Report. Technical report.

Australian Scholarships Group (2016). ASG's Education Costs Estimates — Metropolitan New South Wales Estimated schooling costs in 2016. Summary sheet, `http://www.asg.com.au/doc/default-source/Media-Releases/Planning-for-Education-Index-2016/ASG_EdCosts_SchoolCosts_2016_NSW_Metro.pdf`. Accessed: 2016-10-21.

Azmat, G., C. Calsamiglia, and N. Iriberri (2014, November). Gender differences in response to big stakes. LSE Research Online Documents on Economics 60607, London School of Economics and Political Science, LSE Library.

Betts, J. R. and J. L. Shkolnik (2000). The effects of ability grouping on student achievement and resource allocation in secondary schools . *Economics of Education Review 19*(1), 1 – 15.

Broinowski, A. (2015, January). Testing times: selective schools and tiger parents. *Sydney Morning Herald*. `http://www.smh.com.au/good-weekend/testing-times-selective-schools-and-tiger-parents-20150108-12kecw.html`. Accessed: 2016-10-18.

Bronx Science Alumni Association (2011). The Campaign for the Bronx High School of Science. `http://bronxscienceendowmentfund.org/capital/`. Accessed: 2016-08-31.

Bui, S. A., S. G. Craig, and S. A. Imberman (2014). Is Gifted Education a Bright Idea? Assessing the Impact of Gifted and Talented Programs on Students. *American Economic Journal: Economic Policy 6*(3), 30–62.

Bureau of Studies, Teaching and Educational Standards NSW (2011, February). Explanation of aligning and moderating procedures for the Higher School Certificate. `http://www.boardofstudies.nsw.edu.au/hsc-results/moderation.html`. Accessed: 2016-10-25.

Bureau of Studies, Teaching and Educational Standards NSW (2014, November). 2015 Higher School Certificate Rules and Procedures.

Calonico, S., M. D. Cattaneo, and M. H. Farrell (2015, August). On the Effect of Bias Estimation on Coverage Accuracy in Nonparametric Inference. *ArXiv e-prints*.

Calonico, S., M. D. Cattaneo, and M. H. Farrell (2016). Coverage Error Optimal Confidence Intervals for Regression Discontinuity Designs. working paper.

Calonico, S., M. D. Cattaneo, and M. H. Farrell (2016). rdrobust: Software for Regression Discontinuity Designs. working paper.

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik (2016, March). Regression Discontinuity Designs using Covariates. working paper.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014a, December). Robust data-driven inference in the regression-discontinuity design. *Stata Journal 14*(4), 909–946.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014b, November). Robust Nonparametric Confidence Intervals for Regression–Discontinuity Designs. *Econometrica 82*, 2295–2326.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2015). Optimal Data-Driven Regression Discontinuity Plots. *Journal of the American Statistical Association 110*(512), 1753–1769.

Cameron, A. C. and D. Miller (2015). A Practitioner's Guide to Cluster-Robust Inference. *Journal of Human Resources 50*(2), 317–372.

Card, D. and L. Giuliano (2014, September). Does Gifted Education Work? For Which Students? NBER Working Papers 20453, National Bureau of Economic Research, Inc.

Card, D. and L. Giuliano (2015, September). Can Universal Screening Increase the Representation of Low Income and Minority Students in Gifted Education? NBER Working Papers 21519, National Bureau of Economic Research, Inc.

Card, D. and L. Giuliano (2016, March). Can Tracking Raise the Test Scores of High-Ability Minority Students? NBER Working Papers 22104, National Bureau of Economic Research, Inc.

Carrell, S. E., M. Hoekstra, and E. Kuka (2016, February). The Long-Run Effects of Disruptive Peers. NBER Working Papers 22042, National Bureau of Economic Research, Inc.

Cattaneo, M. D., M. Jansson, and X. Ma (2016a). rddensity: Manipulation Testing based on Density Discontinuity. working paper.

Cattaneo, M. D., M. Jansson, and X. Ma (2016b). Simple Local Regression Distribution Estimators with an Application to Manipulation Testing. working paper.

Centre for Eduation Statistics and Evaluation (2014, January). Language and diversity in NSW government schools in 2013. Technical report, NSW Department of Education.

Centre for Eduation Statistics and Evaluation (2015, September). Schools and Students: 2014 Statistical Bulletin. Technical report, NSW Department of Education.

Cherrybrook Technology High School (2015). Cherrybrook Technology High School 2015 Annual Report.

Cherrybrook Technology High School (2016a). English. `http://cths.nsw.edu.au/curriculum/english/`. Accessed: 2016-10-03.

Cherrybrook Technology High School (2016b). Year 9. `http://cths.nsw.edu.au/curriculum/mathematics/year-9/`. Accessed: 2016-10-03.

Clark, D. (2010). Selective schools and academic achievement. *The BE Journal of Economic Analysis & Policy 10*(1).

Clark, D. and E. Del Bono (2016, January). The Long-Run Effects of Attending an Elite School: Evidence from the United Kingdom. *American Economic Journal: Applied Economics 8*(1), 150–76.

Cohodes, S. (2015). The Long-Run Impacts of Tracking High-Achieving Students: Evidence from Boston's Advanced Work Class. working paper.

Coleman, J. S., E. Campbell, C. Hobson, J. Partland, A. Mood, F. Weinfeld, and R. York (1966). Equality of educational opportunity.

Cross, S. E. and L. Madson (1997). Models of the self: self-construals and gender. *Psychological bulletin 122*(1), 5.

Ding, W. and S. F. Lehrer (2007, May). Do Peers Affect Student Achievement in China's Secondary Schools? *The Review of Economics and Statistics 89*(2), 300–312.

Dobbie, W. and R. G. J. Fryer (2011, August). Exam High Schools and Academic Achievement: Evidence from New York City. NBER Working Papers 17286, National Bureau of Economic Research, Inc.

Dobbie, W. and R. G. J. Fryer (2014, July). The Impact of Attending a School with High-Achieving Peers: Evidence from the New York City Exam Schools. *American Economic Journal: Applied Economics 6*(3), 58–75.

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

Dustan, A. (2010). Have Elite Schools Earned their Reputation? High School Quality and Student Tracking in Mexico City. working paper.

Dustan, A., A. de Janvry, and E. Sadoulet (2015, March). Flourish or Fail? The Risky Reward of Elite High School Admission in Mexico City. Vanderbilt University Department of Economics Working Papers 15-00002, Vanderbilt University Department of Economics.

Epple, D. and R. Romano (2011). Peer effects in education: A survey of the theory and evidence. *Handbook of social economics 1*(11), 1053–1163.

Figlio, D. N. and M. E. Page (2002). School choice and the distributional effects of ability tracking: does separation increase inequality? *Journal of Urban Economics 51*(3), 497–514.

Fort, M., A. Ichino, and G. Zanella (2016, June). On the perils of stacking thresholds in RD designs. working paper.

Gale, D. and L. S. Shapley (1962). College admissions and the stability of marriage. *The American Mathematical Monthly 69*(1), 9–15.

Gelman, A. and G. Imbens (2014, August). Why High-order Polynomials should not be used in Regression Discontinuity Designs. NBER Working Papers 20405, National Bureau of Economic Research, Inc.

Gneezy, U., M. Niederle, and A. Rustichini (2003). Performance in competitive environments: Gender differences. *The Quarterly Journal of Economics 118*(3), 1049–1074.

Hahn, J., P. Todd, and W. van der Klaauw (2001). Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design. *Econometrica 69*(1), 201–209.

Hanushek, E. A., J. F. Kain, J. M. Markman, and S. G. Rivkin (2003). Does peer ability affect student achievement? *Journal of applied econometrics 18*(5), 527–544.

High Performing Students Unit (2016a). Information about applying for Year 5 entry to an opportunity class in 2017. Technical report, NSW Department of Education.

High Performing Students Unit (2016b). Information about applying for Year 7 entry to selective high schools in 2017. Technical report, NSW Department of Education.

Ho, C. (2016, October). Hothoused and hyper-racialised: the ethnic imbalance in our selective schools. *The Guardian*. `https://www.theguardian.com/commentisfree/2016/oct/27/hothoused-and-hyper-racialised-the-ethnic-imbalance-in-our-selective-schools?CMP=soc_568`. Accessed: 2016-10-27.

Houng, B. (forthcoming). The Effect of Selective High Schools on University Entrance Results: Exploratory Analysis from Australia. Doctoral dissertation.

Houng, B. and M. Justman (2013). NAPLAN Scores as Predictors of Access to Higher Education in Victoria. Melbourne Institute Working Paper Series, Working Paper No. 22/14.

Hoxby, C. (2000, August). Peer Effects in the Classroom: Learning from Gender and Race Variation. NBER Working Papers 7867, National Bureau of Economic Research, Inc.

Hoxby, C. and G. Weingarth (2005). Taking Race Out of the Equation: School Reassignment and the Structure of Peers Effects. Technical report, MIT mimeo.

Husbands, C. (2016, September). Dear Theresa May, this is what you need to know about grammar schools. *The Conversation*. `https://theconversation.com/dear-theresa-may-this-is-what-you-need-to-know-about-grammar-schools-65360`. Accessed: 2016-10-07.

Imbens, G. and K. Kalyanaraman (2012). Optimal Bandwidth Choice for the Regression Discontinuity Estimator. *Review of Economic Studies 79*(3), 933–959.

Jackson, C. K. (2010, December). Do Students Benefit from Attending Better Schools? Evidence from Rule-based Student Assignments in Trinidad and Tobago. *Economic Journal 120*(549), 1399–1429.

Jackson, C. K. (2013). Can higher-achieving peers explain the benefits to attending selective schools? Evidence from Trinidad and Tobago. *Journal of Public Economics 108*(C), 63–77.

Klugman, J. (2013). The advanced placement arms race and the reproduction of educational inequality. *Teachers College Record 115*(5), 1–34.

Lazear, E. P. (2001). Educational production. *The Quarterly Journal of Economics 116*(3), 777–803.

Lee, D. S. and T. Lemieux (2010). Regression Discontinuity Designs in Economics. *Journal of Economic Literature 48*, 281–355.

McCrary, J. (2008, February). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics 142*(2), 698–714.

North Sydney Girls High School (2016). Years 11 & 12 HSC. `http://web2.northsydgi-h.schools.nsw.edu.au/public/years-11-12-hsc`. Accessed: 2016-09-21.

NSW Department of Education (2015a). Secondary Schools - Selective High Schools.

NSW Department of Education (2015b, October). The Resource Allocation Model (RAM) in 2016. Technical report.

NSW Department of Education (2016, April). Selective High Schools Entry in year 7 2016 - minimum scores.

Perry, L. (2015, February). The lesson from Canada: why Australia should have fewer selective schools. *The Conversation*. `https://theconversation.com/the-lesson-from-canada-why-australia-should-have-fewer-selective-schools-35534`. Accessed: 2016-10-07.

Planning and Innovation, NSW Department of Education (2008). Statistical Bulletin: Schools and Students in New south Wales, 2007. Technical report, NSW Department of Education.

Pop-Eleches, C. and M. Urquiola (2013, June). Going to a Better School: Effects and Behavioral Responses. *American Economic Review 103*(4), 1289–1324.

Pope, D. G. and J. R. Sydnor (2010, Spring). Geographic Variation in the Gender Differences in Test Scores. *Journal of Economic Perspectives 24*(2), 95–108.

PwC (2013, March). Local Schools, Local Decisions — Resource Allocation Model. Technical report, NSW Department of Education and Communities.

Rickard, K. and L. Lu (2014, August). Family Occupation and Education Index (FOEI) 2013. Technical report, Centre for Education Statistics and Evaluation, NSW Department of Education.

Sacerdote, B. (2011, May). *Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?*, Volume 3 of *Handbook of the Economics of Education*, Chapter 4, pp. 249–277. Elsevier.

Sacerdote, B. (2014). Experimental and Quasi-Experimental Analysis of Peer Effects: Two Steps Forward? *Annual Review of Economics 6*(1), 253–272.

Salam, R. (2014, July). Close Stuyvesant High School: Why super-elite public magnet schools arent necessary anymore. *Slate*. http://www.slate.com/articles/news_and_politics/politics/2014/07/the_case_for_shutting_down_stuyvesant_high_school_the_best_public_school.html. Accessed: 2016-10-23.

Shue, K. (2013). Executive Networks and Firm Policies: Evidence from the Random Assignment of MBA Peers. *Review of Financial Studies 26*(6), 1401–1442.

Smith, A. (2014, April). Selective schools 'the most socially exclusive' in NSW. *Sydney Morning Herald*. http://www.smh.com.au/national/education/selective-schools-the-most-socially-exclusive-in-nsw-20140405-365r8.html. Accessed: 2016-10-18.

Staiger, D. and J. Stock (1997). Instrumental variables regression with weak instruments. *Econometrica 65*(3), 557–586.

Thaler, R. and C. Sunstein (2008). *Nudge: Improving Decisions about Health, Wealth, and Happiness*. Yale University Press.

The Economist (2016, August). Grammatical Error. *The Economist*. http://www.economist.com/news/britain/21704837-lifting-ban-new-selective-schools-would-damage-social-mobility-grammatical-error. Accessed: 2016-10-07.

Thistlethwaite, D. L. and D. T. Campbell (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology 51*(6), 309.

Tincani, M. M. (2015). Heterogeneous Peer Effects and Rank Concerns: Theory and Evidence. working paper.

Ting, I. and E. Bagshaw (2015, December). HSC Results 2015: Top 20 HSC schools revealed. *Sydney Morning Herald*. http://www.smh.com.au/national/education/hsc-results-2015-top-20-hsc-schools-revealed-20151216-glory7.html. Accessed: 2016-10-18.

Universities Admission Centre (NSW & ACT) (2015, May). Calculating the Australian Tertiary Admission Rank in New South Wales. Technical report.

UNSW Australia (2014). Science Undergraduate Guide 2015. Brochure, http://www.science.unsw.edu.au/files/UNSW-Science-UG-Guide-2015.pdf. Accessed: 2016-10-01.

Urquiola, M. and E. Verhoogen (2009, March). Class-Size Caps, Sorting, and the Regression-Discontinuity Design. *American Economic Review 99*(1), 179–215.

Wilson, R., B. Dalton, and C. Baumann (2015, March). Six ways Australias education system is failing our kids. *The Conversation*. `https://theconversation.com/six-ways-australias-education-system-is-failing-our-kids-32958`. Accessed: 2016-10-07.

Yager, R. (2016, May). Report on the Scaling of the 2015 NSW Higher School Certificate. Technical report, Universities Admission Centre (NSW & ACT).

Zhang, H. (2008). Magnet Schools and Student Achievement: Evidence from a Randomized Natural Experiment in China. working paper.