Achievement Gains from Attendance at Selective High Schools

Brendan Houng

ORCID: 0000-0002-0858-7876

Submitted in partial fulfilment of the requirements of the degree of

Doctor of Philosophy

March 2018

Melbourne Institute of Applied Economic and Social Research

THE UNIVERSITY OF MELBOURNE

Abstract

Academically selective high schools are a polarizing topic in education policy, despite only having a small presence in some Australian states. They appear successful. The schools regularly top annual school rankings of university entrance results, but this is perhaps unsurprising given that their students are admitted based on their performances on an entrance exam.

This thesis asks whether selective high schools improve their students' university entrance results beyond what they would have achieved otherwise. The main chapter is a case study from an anonymized Australian state that follows high-achievement students through high school. The key challenge is finding a group of non-selective students comparable to those who attend selective schools. For additional background, the thesis explored the following themes: the historical development of selective high schools, the premise that the schools cater to gifted and talented students, and the high levels of demand for the schools within current trends in educational policy.

The thesis provides the first estimates of the selective school effect (roughly contemporaneous with Zen 2016[1]) from matching and regression discontinuity approaches in the Australian context, which are improved statistical methods compared with that of previous research (e.g. regression analyses from Lu and Rickard, 2014). The estimates point to small positive effects at best on university entrance results from attending the selective schools.

Overall, the small selective school effect appears to reflect the high levels of educational aspiration of both selective students as well as applicants who attended other schools. Both groups of students appear to be among the most driven and motivated, being disproportionately from immigrant and socio-economically advantaged backgrounds and having implicitly signaled an aspirational intent by applying to the schools.

Lastly, the thesis expands on one aspect of the selective schools, whereby many of their students experience a decrease in within-school achievement ranks from attending a school with high-achievement peers. In a more general context, the thesis assesses the effect from changes in local ranks on later achievement for students who transitioned from primary to secondary school. The results indicate that perceived increases in local rank have a negative

[1] Zen (2016) conducted similar regression discontinuity analyses of 18 selective schools in New South Wales, reflecting a broad range of academic selectivity. Zen finds limited and mostly insignificant effects consistent with the research here and previous studies from the UK and the USA. Note that Zen's study was not accessible prior to March 2018.

effect on standardized test scores, suggesting that students reduced their allocation of effort in response to random increases in rank.

The new empirical evidence from the thesis supports the view that selective schools represent a positive achievement ideal for their students. Recent public policy discourse on the selective schools has included calls for expansion of the system to the primary school level in one state, and criticisms of a hyper-competitive culture at the schools, including suggestions of unfair entry due to excessive tutoring on the part of applicants. The research positively contributes to the discourse by providing historical context, identifying the relevant issues and articulating the potential indirect consequences of these policies.

Declaration

This is to certify that:

1. the thesis comprises only my original work towards the PhD except where indicated in the Preface,
2. due acknowledgement has been made in the text to all other material used,
3. the thesis is fewer than 100,000 words in length, exclusive of tables, maps, bibliographies and appendices.

Brendan Houng

Brendan Houng, March 2018

Preface

My experience attending a selective high school inspired my interest in this topic.

Chapter 4 is an iteration of a research project conducted in 2016 and is joint work with my supervisor Associate Professor Chris Ryan. It is also the second set of analyses relating to the main research question, reflecting the historical development of the research. Chapter 3, the preliminary set of analyses, was presented at an Economics PhD conference at the University of Queensland in November 2015.

Data, Privacy and Anonymization

The main research uses data from an Australian state that is anonymized for privacy concerns regarding the identities of both the state and individual selective schools. The number of selective high schools in the thesis is not specified and certain details are intentionally less precise to prevent the identification of schools. The research from Chapter 5 uses data from the Victorian Department of Education and Training. The findings and views reported in this thesis are those of the author(s) and should not be attributed to any branches of state or Australian Commonwealth governments.

Acknowledgements

I am deeply grateful to my supervisors Associate Professor Chris Ryan, Professor Guyonne Kalb, and Professor Moshe Justman. I thank them for exceeding their duties, guiding and supporting me towards raising the quality of my work. I especially thank Chris as my main supervisor for helping me develop an intuition for narrative, and for his inexhaustible patience and insights. I thank Guyonne for stepping up at the crucial end period to shoulder a large share of the burden, helping me with improving the clarity of my writing and getting all the small details right. I thank Moshe for showing me the importance of sustained effort in the Sisyphean struggle that is the creative process. I thank: Dr. Cain Polidano for thought-provoking discussions on education and selective schools, committee member Professor John Haisken-DeNew for his contribution to this work, and Associate Professor Gigi Foster for providing helpful comments on an earlier version of Chapter 3.

I am indebted to the Melbourne Institute of Applied Economic and Social Research (Melbourne Institute), where I have worked on public policy research and on this thesis. I thank my colleagues and fellow students at the Melbourne Institute and at the Faculty of Business and Economics who I have gotten to know so well and helped me along the way, and with whom I shared a common enjoyment in the exchange of ideas and furthering research in public policy. I am thankful for crossing paths with Dr. Mike Helal and Lama Andari. I am thankful for the University of Melbourne for providing an ideal environment for intellectual endeavours with beautiful surrounds.

For Chapters 3 and 4, I thank the state education department for providing access to the data that is used in this research, and thank the individuals at the state department who helped facilitate the research. I thank also the principals of the selective schools for their valuable input. I thank seminar and conference participants at the Melbourne Institute and at the state education department for their helpful feedback on earlier versions of this research.

I thank the Victorian Department of Education and Training (formerly the Department of Education and Early Childhood Education) and the Victorian Curriculum and Assessment Authority, for providing access to the data that is used in Chapter 5. I would like to thank the individuals involved in setting up the research partnership between the Department of Education and Training and the Melbourne Institute, including Professor Deborah Cobb-Clark and Professor Duncan McVicar. I thank the individuals who helped facilitate my research under the research partnership and provided their comments and feedback on the veracity of the details in my work, including: Rose Peterson, Leona Smith, Catherine Rule, Joe Bui, Michael

Dalton, Emma Katauskas, Scott Ware, Jamie Jacob, Sam Dwyer, Colleen Howell, Joyce Cleary, and Andrew Nipe. Apologies to others who I have missed.

I thank my loving wife Saori for her unwavering belief in me, and my children, Yuka, Kenji and Sota, for inspiring me. I thank my parents, Simon and Susan Houng, for their encouragement and support. I thank all my friends and family, who helped provide motivation when progress stalled. Thank you to you all.

Table of Contents

List of Figures

List of Tables

Copyright Material

Citation Information for Third party copyright material	Location of item in thesis	Permission granted Y/N
Top of the charts, figure from Clynes, T. (2016). How to raise a genius: lessons from a 45-year study of super-smart children, Nature News.	p. 29	Y

Chapter 1. Introduction

Academically selective high schools[2], which admit students on the basis of an entrance exam, make up a small part of education systems in countries like Australia, the United Kingdom and the United States. They appear successful. Selective high schools regularly top school rankings for university entrance results reported in metropolitan newspapers, while some selective schools are known for their successful alumni, who include among them prominent public figures. Nevertheless, there is evidence from New York, Boston and the United Kingdom that selective high schools produce only scattered additional gains in high school test outcomes and have little impact on university attendance for those who attend them (Dobbie and Fryer, 2014, Adulkadiroglu et al., 2014, and Clark, 2010).

This thesis has several aims. As the primary focus of the thesis, I ask: What is the effect of selective school attendance for the academic achievement of students on the end-of-high school assessment used to determine university entrance? The key challenge is finding a group of non-selective students comparable to those who attend selective schools to subsequently compare their outcomes. I undertake this task in two sets of analyses (parts one and two), each corresponding to a different cohort of students from the same anonymized Australian state. I follow the students through high school and apply different methods with comparison groups for each cohort.

The main analyses provide the first estimates from matching and regression discontinuity approaches (roughly contemporaneous with Zen 2016 [3]) in the Australian context; the estimates point to a small effect at best, consistent with studies from other high-income countries. These improved statistical methods produced more precise estimates of the selective school effect compared with that of previous research (e.g. value-added regression analyses from Lu and Rickard, 2014). Additional exploratory analyses help explain the differences in results between methods and support the interpretation that selective students are highly motivated.

Second, the thesis seeks to understand the role of selective high schools in the Australian education system from an economic perspective. The selective schools are both polarizing and interesting for being at the intersection of several research areas in education, psychology, and

[2] Also known as select-entry high schools or exam schools in the USA.

[3] Zen (2016) conducted similar regression discontinuity analyses of 18 selective schools in New South Wales, reflecting a broad range of academic selectivity. Zen finds limited and mostly insignificant effects consistent with the research here and previous studies from the UK and the USA. Note that Zen's study was not accessible prior to March 2018.

economics. To that end, I trace the historical development of the selective high schools and review the literature on academic achievement and academic selection across schools and related research in educational psychology. I also explore the premise that the selective high schools cater to gifted and talented and high-achievement students, and in addition, provide commentary on the apparent high levels of demand for the schools in the context of current trends in educational policy.

Lastly, I investigate an indirect feature of the selective high schools which potentially helps explain differences in the results from the two methods in the main analyses. Due to the concentration of high-achieving peers, many students who attend selective high schools experience a downgrade in their relative academic status, which is likely to lead to lower self-concept and may negatively affect their later achievement. For this analysis, I estimate the effect of changes in local achievement ranks on later academic achievement in the general context, when students transition from primary to secondary school.

The thesis contributes to the public discourse surrounding the selective schools' policies by providing the first RD estimates of the selective school effect (roughly contemporaneous with Zen 2016) from an anonymized Australian state, and placing the selective schools' policies in their historical and conceptual contexts. I discuss several themes including potential explanations for the high levels of demand for selective schools, the argument that the schools increase social inequality, and the premise that the schools cater to gifted students. Reviewing the literature on peer effects and academic achievement, I also relate academic selection to students' academic self-concept, their belief in their own ability.

Overall, the small selective school effect appears to reflect the high levels of educational aspiration of selective students as well as applicants who attended other schools. Both groups of students are disproportionately from immigrant and socio-economically advantaged backgrounds[4], and they appear to be among the most driven and motivated, having implicitly signaled an aspirational intent by applying to the schools.

From a policy perspective, there is an open question as to the optimal number of selective schools, which varies across states and territories in Australia, the level at which education policy is administrated. Policies regarding selective schools attract significant media attention and are hotly debated, perhaps because the schools are seen to reflect broader societal values

[4] For the selective schools analysed in Chapter 3, over 75% of students were from a foreign language background compared to around one-quarter in the broader population.

such as the encouragement of competitive striving for achievement or, through a different lens, the promotion of elitism and social inequality.

1.1. Structure

The structure of the thesis is as follows. There are six chapters, including the Introduction and Conclusion, which are Chapters 1 and 6, respectively. Chapter 2 is the literature review which places the selective school policies in the broader research and policy context. Chapters 3 and 4 correspond to Parts 1 and 2 of the main analyses, and together comprise a case study of achievement gains at selective high schools from an anonymized Australian state. Chapter 5 assesses the effect of changes in local achievement ranks (relative academic status), between primary and secondary school, for students' later achievement.

Note that all chapters are largely self-contained but there is some referencing across chapters. Part 2 of the main analyses (Chapter 4) references some sections from Part 1 (Chapter 3 - literature review, methodology, and data description), while there are also some references to the literature review chapter (Chapter 2) from other chapters.

1.2. Background

This background section describes the representation of selective schools across states and territories in Australia, briefly reviews their historical development, and discusses the argument that selective schools increase social inequality in the education system and the counter arguments that selective schools improve social mobility and encourage the aspirations of their students. I note that their distribution across geographic areas is correlated with population size, while fluctuations in their development also appear to be associated with trends in social inequality, which the selective schools are perceived to symbolize.

1.2.1. Selective High Schools in Australia

Australia has a choice-based education system, with most students typically attending their local government schools, and a large proportion of students of up to 40% attending private schools. Selective high schools generally have a small presence but there is some variation across the states and territories depending on their education policies.

In 2015, there were 25 selective high schools across the states of New South Wales (NSW)[5], Queensland, Victoria and Western Australia[6]. Most of these schools admit students based on entrance exams or tests, while a few schools that focus on certain areas of study, such as science or mathematics, also include interviews as part of their admissions process. There were also 70 other high schools in these same states offering select-entry programs, where one or more classes of students (as compared with all students) are admitted in a similar process involving tests and interviews.

In Australia changes to selective school, policies appear to have been made with the goal of educating gifted and talented children. At times, it appears that high-achievement and gifted students are equated (e.g. Chan, 1996 and Gross, 2005), while changes in selective school policies in Australia parallel a similar pattern for the gifted and talented programs in the USA, whose policy support has been influenced by cultural sentiment:

> '…gifted and talented students become a national priority when excellence is sought and a critical need is perceived. However, as equity becomes a predilection, gifted students' needs are seen as an elitist luxury and are replaced with the priorities of students within other subpopulations.'
>
> (Jolly, 2009)

Geographically, most of the selective schools are located in the more populous states, reflecting the preferences of residents and the level of government at which education policy is administrated. This is perhaps unsurprising as under a choice-based education system, the benefits from specialization of schools are likely to be greater in larger populations, while it is also possible that demand for the schools is related to increased social inequality more commonly found in urban areas as suggested by Blow, Blundell and Machin (2010), cited in Blundell, Dearden and Sibieta (2010). The correlation between number of selective schools and population size is an instance of the sociological phenomena of geographic sorting of individuals, a natural process attributed to individual preferences and residential mobility; the early influential theoretical models were Schelling (1969) and Tiebout (1956).

The sorting of students across schools which results from the selective school policy can be perceived as decreasing social equity due to the increased segregation of students by socioeconomic background (e.g. Edwards, 2005), regardless of potential benefits to the selective students. The issue of their existence is therefore a polarizing one, and the history of

[5] Fully selective schools comprised 2.1% (17 of 815) of secondary schools in NSW, including combined primary/secondary schools. Australian Bureau of Statistics, 4221.0 Schools, Australia 2016. Released 2 February 2017.
[6] See Braggett and Moltzen (2000) for differences in gifted education policy across states and territories.

expansions of selective schools reflects this. A re-occurring argument against the streaming of students is the impact it has on local schools, by weakening ties to local community[7], and by promoting social inequality and residualisation through concentrating socioeconomic disadvantage in schools (Lamb, 2007 and 2008).

1.2.2. Historical Development[8]

The number of selective high schools has grown unevenly over time, with many introduced after the late 1980s. The original selective high schools, of which there were few, commenced as the first government high schools in their respective states, over the period spanning the middle of the 19th century to the early 20th century. Their purpose was to provide a public alternative to private schools to allow students without sufficient financial resources to undertake secondary schooling. That is, the selective schools had the historical function of promoting social equality through providing access to secondary schooling, and indirectly university education. This is a sentiment echoed today in relation to providing opportunities to working-class students[9] and for providing an alternative to private schooling[10].

Free and compulsory primary school education was introduced in the latter half of the 19th century, with basic assumptions of uniformity and conformity and with standards set in terms of minimum attainments (Braggett 1985). Despite research on gifted students in the 1920s, Australian education was generally not progressive until the 1950s, although there were some cases of catering to gifted students. For example, special classes for gifted students began in 1932 in a small number of high schools in NSW. By 1938, secondary school enrolments varied between 3% and 20% across the states for children attending state schools.

There was a large period of change between the 1950s and 1970s resulting from the move towards an egalitarian secondary school system, which coincided with increases in the minimum school leaving age, the removal of tuition fees, and increased rates of immigration and rising birthrates. The subsequent pressures on the secondary schools led to an increased demand for catering to students with higher levels of academic ability.

[7] Inquiry into the provision of public education in NSW, Teachers Federation & Federation of P & C Associates of NSW, 2002.

[8] This subsection draws on: Braggett (1985); Education and Training Committee (2012), Inquiry into the Education of Gifted and Talented Students, Parliamentary Paper of Victorian Government; Teachers Federation & Federation of P & C Associates of NSW (2002), Inquiry into the provision of public education in NSW; and brief histories provided on websites of some of the selective schools.

[9] First Speech - Senator and Minister, Hon Mitch Fifield, http://www.mitchfifield.com/Media/Speeches/tabid/71/articleType/ArticleView/articleId/238/First-Speech.aspx, accessed 28 September 2016.

[10] In a class of their own, Jewel Topsfield, The Age, 5 February 2011. http://www.smh.com.au/national/education/in-a-class-of-their-own-debate-20110204-1agy0.html

The culmination of pent-up demand for the education of gifted children[11] (and selective schools) resulted in the late 1980s expansion of 8 selective schools in NSW, which almost doubled the number of applications to the schools. This contrasts with the preceding period which was characterized by a lack of policy support for selective high schools. In the late 1970s, some selective schools were closed, and the existing selective schools were not selective in practice due both to the use of catchment zones and preferential treatment to sons and daughters of alumni. In another state, one of the establishment selective high schools became a comprehensive (non-selective) high school in the 1950s, before reverting to a selective high school in the 2000s.

From the history of selective high schools in Australia, policy support and opposition to selective schools appears to trend over spans of decades. At present time there appears to be support for selective schooling, with expansions in one state in recent years[12], while in NSW boarding positions are available at some of the selective schools, and a remote access selective school has also been created[13]. Recent acceptance rates among applicants were over 30% in NSW[14], with similar levels from applicants for the state on which the case study is based. As part of the ongoing trend of increasing academic selection, a new selective boarding school has also been planned in NSW[15].

1.2.3. Competition and Social Inequality

The role of selective schools in choice-based education systems is an unexpectedly complex one despite its simple aim of catering to high-achievement students. It has been suggested that school choice increases school competition by allowing students to attend high performing schools and consequently improve underperforming schools (Cullen, Jacob and Levitt, 2005 and Bradley, Johnes and Millington, 2000[16]).

[11] Braggett (1985) identifies the beginning of a low-key movement in education policy to provide for gifted children in 1976-77.

[12] Consistent with Chapters 3 and 4, states other than NSW are not identified for privacy concerns relating to specific selective schools.

[13] Source: Information about applying for Year 7 entry to selective high schools in 2019. NSW, Department of Education Website (https://schoolsequella.det.nsw.edu.au/file/eafa8804-3a28-4730-9fc0-0ee49dd90f83/1/SHS-application-information.pdf), accessed 14 November 2017.

[14] Education and Communities, NSW Government (2015). Selective High Schools Entry in Year 7 2015 – minimum scores, (April 2015). https://www.det.nsw.edu.au/media/downloads/about-us/statistics-and-research/key-statistics-and-reports/selectivehs.pdf

[15] Source: NSW government to build new co-ed selective boarding super-school. Eryk Bagshaw. http://www.smh.com.au/national/education/nsw-government-to-build-new-coed-selective-boarding-superschool-20151118-gl22u8.html, accessed 27 September 2016.

[16] As measured by school exam performances and attendance rates from all secondary schools in England, 1993-1998.

At the societal level, although the immediate impact of the selective schools is small due to their limited size, they also symbolize issues of elitism and social inequality. It is argued that the specialization introduced by formal streaming undermines equity in education and the inclusive ideal behind the comprehensive system. The main argument is that selective schools increase social inequality, disadvantaging other students through the loss of high achievers from other schools, which in turn leads to wider inequalities between performing and underperforming schools without benefiting the state in aggregate (Lamb, 2007 and 2008)[17].

A second related argument against selective schools is that their existence weakens the ties of students to their local communities through the concentration of high-achieving students in a few locations. The social inequality that is perceived to be perpetuated by the selective schools is potentially accentuated by the high representation of students from an immigrant background at these schools[18] (see section 2.5.3 Hyper-Selectivity and Polarization).

The policy of streaming students across schools is analogous to the issue of school choice and the social segregation across schools and neighbourhoods. The segregation that is reflected in house prices is part of a natural sorting process occurring under residential mobility and has theoretical foundations in sociology (Schelling 1969 and Tiebout 1956).

The prevalence of selective school policies is perhaps unsurprising in the context of a choice-based education system that implicitly encourages the specialization of schools, particularly in larger populations. Edwards (2005) provides the example of Melbourne, between 1993 and 2003 where there was a distinct geographical pattern in secondary schools, between the north and west as compared with the eastern suburbs, increasing in specialization towards vocational or academic outcomes.

Although the selective school policies in Australia have been criticized for increasing inequality, the schools are small in direct impact when compared to that of countries where there is a system-wide allocation of students across schools. Through international comparisons in 2015 PISA[19], Australia placed near the OECD[20] average in both socioeconomic inequality and academic selection.

---

[17] Lamb (2007) also argues that increasing the number of selective schools moves the education system away from a comprehensive model and towards a segregated one, which results in a lower overall performance, as demonstrated by countries with comprehensive systems (Sweden, Finland and Iceland) having higher university entry rates than the OECD average; Germany and Austria are provided as examples of countries with greater levels of segregation.

[18] For the selective schools analysed in this case study (Chapter 3), over 70% of students were from a non-English speaking background, compared to around one-quarter in the broader population.

[19] Programme for International Student Assessment (PISA) measures the performances of 15-year-olds on international standardized tests.

1.2.4. Improved Social Mobility and Public Benefits

The main argument in support of selective public schools is for the promotion of social mobility, with the schools providing equality of opportunity through a meritocratic admission process. In the UK context, selective schools were also presented as way out of poverty and disadvantage through access to quality schools (Coe et al. 2008). In Australia, selective schools have been described as providing opportunities to working-class students and as providing an alternative to private schooling (see section 1.2.2).

The public benefit identified in this thesis is that selective schools provide a goal for students at which to aim at, potentially reinforcing and raising students' aspirations and encouraging their longer-term successes (discussed in sections 2.5.4 and 6.2). This function is may be particularly important in creative domains where top performers have extremely large impacts due to imperfect substitution (discussed in section 2.4.4), leading to a net positive benefit from a societal perspective.

1.3. Setup of the Main Analyses

This section describes the setup of the main analyses of the thesis, which addresses whether selective high schools improve their students' university entrance results. There are two sets of analyses, Parts 1 and 2 - presented in Chapters 3 and 4, which are concerned with identifying the counterfactual for high-achievement students who attend selective schools with respect to their high-school achievement outcomes. Both parts contribute to the same research question and use similar data, however, there are several differences between the two sets of analyses which involve intricate details relating to the admissions process of the selective schools.

Part 1 establishes the foundation by identifying a comparison group of non-selective students from the general population with selective school potential, as indicated by Year 9 achievement on standardized tests. The end-of-high school outcomes of selective and non-selective students are then compared using a matching method that accounts for similarities in their prior achievement and their background characteristics. Part 2 builds on the earlier analyses by identifying an improved comparison group of non-selective students who sat the entrance exam but did not attend selective high schools. Receipt of access to data from the admissions process allows for the estimation of the selective school effect using a regression discontinuity approach, in addition to the matching approach using standardized tests from Year 7.

[20] Organization for Economic Co-operation and Development with 35 member countries.

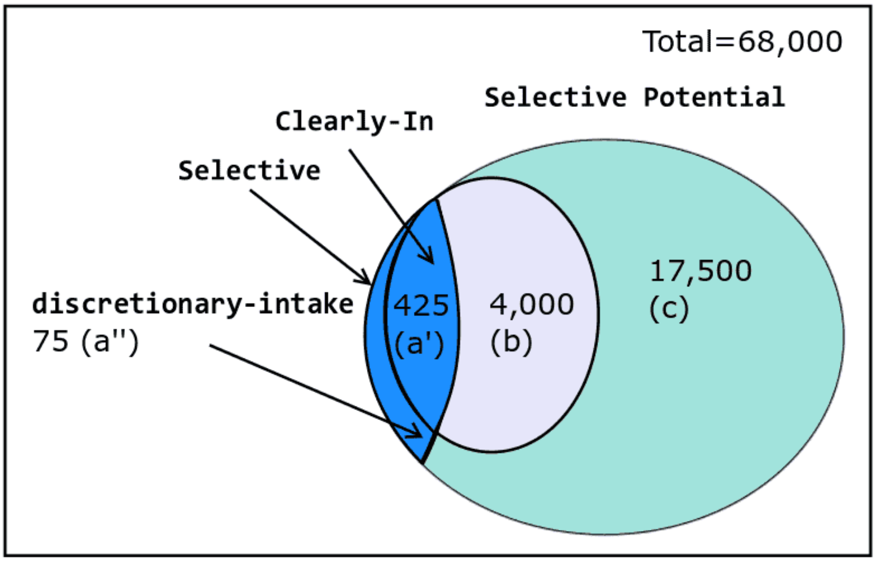
1.3.1. A General Description of Samples and Methods

In this section, I describe the relevant samples and methods applied in the Parts 1 and 2 of the main analyses, using Venn diagrams.

Part 1: Matching from the General Population (Chapter 3)

Figure 1.1 presents the Venn diagram for Part 1, which compares selective and non-selective students from the general state population for one cohort, of around 68,000 students in Year 9. First, I identify the 'selective potential' sample of students who scored above the minimum level at the selective schools on standardized tests. In the diagram, this group of around 22,000 students is represented by the larger ellipse comprising all areas: a', a'', b and c.

Figure 1.1: Venn Diagram of the Student Groups in Part 1



Note: student numbers are rough approximations for ease of explanation.

Next, the selective students are represented by the blue lens (areas a' and a''). The blue lens of selective students is intersected by the smaller ellipse, separating the selective students into the clearly-in sample (a') or a discretionary intake (a''). The smaller ellipse represents the clearly-in sample and is intersected by the blue lens, thereby creating areas a' and b. Together, areas a' and b are the final data sample, reflecting selective and non-selective students, respectively. The intention of defining the clearly-in sample is to exclude students admitted to the selective schools through a discretionary intake, as they may have been filtered for positive unobservable traits, such as stronger motivation.

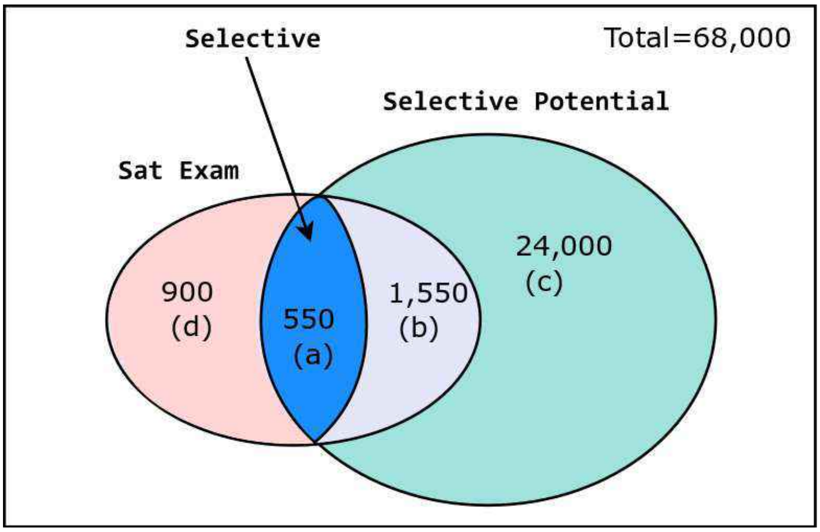
Without access to specific information on which selective students were admitted through the discretionary intake, the clearly-in sample was approximated using the 15th percentile on standardized test scores at selective schools, separating the bottom 15% (area a'') and the top

85% of selective students (area a'). This restriction resulted in 4,000 non-selective students in the clearly-in sample (area b) and 21,500 non-selective students in the broader selective potential sample, the area combining b and c. As both the selective potential and clearly-in samples are defined based on performance levels on the standardized tests, the latter sample is a subset of the former.

Part 2: A Comparison of Two Methods using the Exam Sample (Chapter 4)

Figure 1.2 presents the Venn diagrams for the student groups in Part 2, the second set of analyses. As per the earlier diagram, the blue lens (a) represents the selective students, and the larger ellipse 'selective potential', represents 26,100 students who scored above the minimum level at the selective schools on the standardized tests (areas: a, b, c). The larger number of students in the latter sample results from a lowering of the minimum score obtained by students at selective schools, compared with that from Part 1. This arises from the increase in number of selective schools in the cohort analysed in Part 2, as noted earlier. Note also the slightly larger number of selective students of 550, compared with 425 from the Part 1 Venn diagram, in Figure 1.1.

Figure 1.2: Venn Diagram of the Student Groups in Part 2



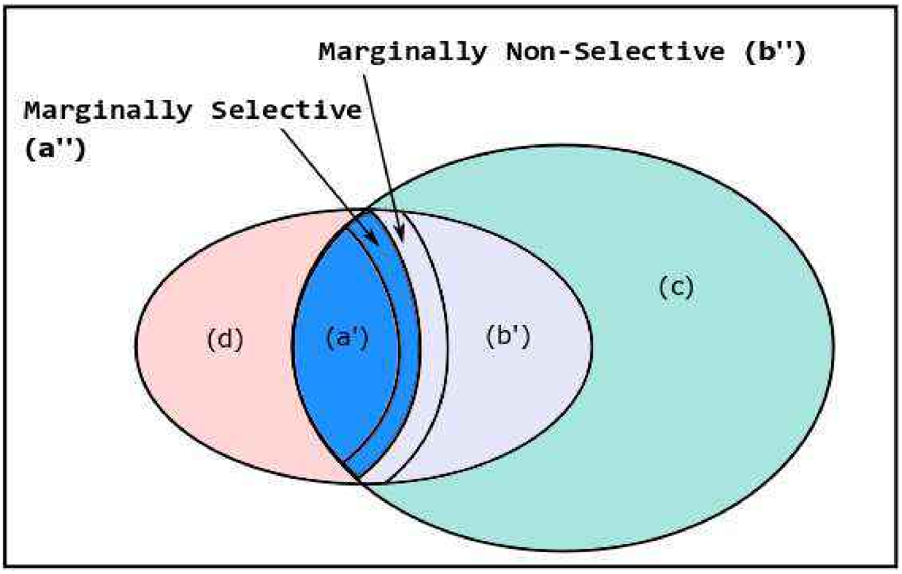
Note: student numbers are rough approximations for ease of explanation.

The selective students (in the blue lens) are a subset of both the larger ellipse and the smaller ellipse 'sat exam', which represents applicants who sat the entrance exam (areas: d, a, and b). The intersection of the 'sat exam' ellipse and the 'selective potential' ellipse (areas a and b) represents the sample for the matching comparison, and is analogous to the clearly-in sample, described earlier for Part 1. Area d represents students who sat the entrance exam but did not receive offers to the selective schools and also did not attain a score on the nationally standardized tests above the minimum observed at the selective schools.

The clearly-in sample[21] of areas a and b are an improvement over the final data sample from Part 1 (Chapter 3) due to the intent signaled by non-selective students in applying for admission to the selective schools. In addition to intent, high acceptance rates from applicants who did receive offers also demonstrate the similarity between applicants who subsequently did (area a) and did not attend the selective schools (area b). The high acceptance rates among applicants who received offers implies, however, a bias in the performance of applicants such that the selective students outperform non-selective students on the entrance exam, with only a small proportion of applicants who received offers, declining them.

The admissions data and entrance exam results allow for a further improved comparison to the data sample obtained from the exam sample. This refers to the comparisons between marginally selective and marginally non-selective students, indicated by areas a’’ and b’’ in Figure 1.3, who were applicants who received results on the entrance exam which were just above, and just below, the admission cut-offs at the selective schools. These areas are within the lines parallel to the line forming the right side of the blue lens, representing the selective students. This line separated the selective (a) and non-selective students (b) in the exam sample, from Figure 1.2.

Figure 1.3: Marginally Selective and Marginally Non-Selective Students (from Part 2)



Note that applicants express preferences for specific selective schools such that Figure 1.3 represents an abstraction of the marginal selective and marginal non-selective students. In

[21] Note selective students who were admitted through the discretionary intake were identifiable in Part 2 and are not represented in Figure 1.2.

practice, there is a pair of selective and non-selective groups for each selective school for the data sample used in the regression discontinuity analyses. Figure 1.3 also shows that these analyses based on marginal students rely on smaller sample sizes than the matching approach.

1.3.2. Key Differences

There are two key differences between the two sets of analyses, which are detailed as follows.

Each chapter follows a different cohort of students through high school from the same anonymized Australian state, graduating in different years: in 2011 and in 2014. This is noteworthy because there are a greater number of selective schools in the second later cohort, with one of the selective schools being in its first year of operation[22], causing the level of academic selectivity of the schools in aggregate to decrease. I.e. the cut-off score from the entrance exam for receiving an offer to the selective schools is lower due to the increase in number of places at selective schools.

The other key difference is that we received information relating to the admissions process to the selective schools for the second set of analyses, Part 2 (Chapter 4). This has several consequences:

a) first, this information allows us to exclude with certainty the discretionary intakes of selective students from the analyses, who were not admitted strictly on an academic basis, thus addressing potential selection issues. In Part 1 (Chapter 3), students admitted through the discretionary intake were approximated with achievement;
b) second, the admissions data allow for the identification of a more comparable group of non-selective students. These are non-selective students who sat the entrance exam who have similar levels of academic achievement on standardized tests to the high-achievement students at selective schools; and
c) third, we obtained access to results from the entrance exam used to determine offers to the applicants to the selective schools. This allows us to estimate a quasi-randomised comparison between applicants who were marginally successful in attending the selective schools and those applicants who were marginally unsuccessful. An arbitrary limit in the academic threshold used to make offers to students, which is determined by the number of places available at the selective schools, means that applicants just above and below the threshold are both randomly selected and alike in observable and unobservable characteristics.

[22] This is intentionally vague to preserve the anonymity of the individual selective schools, which are limited in number.

Chapter 2. Literature Review

2.1. Introduction

This chapter explores several themes and presents a broad perspective on the selective high schools, providing context for the later chapters. In section 2.2 I discuss potential explanations for the high levels of private demand for attendance of selective schools, which are related to the signalling of academic ability.

Narrowing the focus, in section 2.3 I outline research on academic achievement at selective high schools, and explain how the sorting of students across schools by academic aptitude (i.e. academic selection) can influence self-concept, students' belief in their own ability. I extend the discussion to self-identity and achievement motivation which potentially offer explanations for why students attend selective schools.

In the next section (2.4), I explore the premise that selective schools cater to academically gifted students, identifying the personality characteristics associated with gifted students and their typical experiences. As a bridge from selective schools and giftedness to the economic narrative of education, I discuss the link between giftedness and expert performance and discuss creative domains which disproportionately reward the top performers.

The fourth section (2.5) discusses the potential social and cultural influences at selective schools, including social benefits to gifted students, and perceived hyper-selectivity from aspirational immigrant culture. The section concludes by discussing the extent to which the selective schools cater to academically gifted and high-achievement students, as informed by the research on giftedness.

The chapter concludes with commentary on the apparent demand for selective high schools in the context of current trends in education policy that are supported by the economic narrative of education (section 2.6). These include a trend towards an expansion of the education system and policies oriented towards higher rates of high school completion, an increased use of standardized testing, and an increased specialization of schools which appears to be premised on the ability of competition to improve standards of achievement.

2.2. Demand for Academic Selection

The small selective school effects on achievement documented in other high-income countries (i.e. the UK and the USA) by the international literature suggests that there may be other explanations for the high levels of demand for selective schools by parents and students. Explanations relating to individual private benefits of: whether the demand can be explained by the reputations of the schools; viewing the schools as positional goods; and/or by information asymmetry in identifying academic ability in students, are considered in turn.

2.2.1. A Reputational Explanation

Besides the academic successes of their students, another possible reason for the high levels of demand that informs peoples' beliefs about the selective schools are the reputations which are formed by the successes of their alumni. The reputations of the schools convey the impression that students who attend these schools go on to experience success in life, specifically in their careers and occupation, and is demonstrated especially well by alumni of the selective schools with prominent roles in society. This view is supported by the earlier research by Braithwaite and Kensell (1992), who showed that the established selective schools held an advantage over the new schools which had apparently yet to create their own narratives of success.

The perceived transmissions of success through the school one attends, via access to social networks or from reinforcement of aspirations is more apparent at individual schools, each of which appear to have their own traditions and narratives. For instance, in Western Australia there is a view that many successful individuals have followed the pathway of attending the Perth Modern school (a selective school), followed by attending the University of Western Australia, before obtaining a Rhodes scholarship; former Prime Minister Bob Hawke is the best-known figure who followed this path.

In NSW, the Fort Street high school was known for providing working class students with the possibility or opportunity for attaining a university level education. This narrative has historical origins, with the earlier intention of the policies for providing equity of opportunity to individuals of disadvantaged background, when the rate of university completion was far lower. In Victoria, Mac Robertson Girls' High School, as a single-sex girls' school, is known for encouraging the career aspirations of their students through celebrating the public achievement of female role models.

Although the significance of school culture and reputations are not confined only to selective schools, it is possible that they help explain the high levels of demand for attending these schools.

2.2.2. Positional Goods

The explanation that the high levels of demand for selective schools can be largely attributed to reputational effects is essentially the idea that students and parents have preferences for the schools that are unrelated to their effectiveness, but which rather derive from concerns of social status. That is, schools are a form of conspicuous consumption, or are positional goods.

The concept of positional goods has been explored in a line of research pursued by Robert H. Frank. Frank (1985) asks whether it is better to be a big frog in a small pond or a small frog in a big pond. Frank built on Hirsch's definition of positional goods (Hirsch, 1976), which is closely related to Veblen's idea of conspicuous consumption (Veblen, 1899).

Frank (2007) suggests that people care about relative consumption in some domains more than others. Houses and schools are positional goods in that the visibility of these goods is such that people tend to spend more on these goods to obtain better relative positioning, perhaps fueling larger and more expensive houses and more prestigious schools. Part of the spending is due to the evaluation of the good being made in relative terms.

The application of positional goods to the selective schools is straight-forward, as there is transparency in the high levels of academic aptitude of students attending the schools, including the expectations of successful longer-term career outcomes. School reputations may also be more important to parents who are themselves aspirational and motivated.

In a similar convincing way, Frank (1985) relates relative status to wages, and the extent to which it reflects the negative effect on well-being from having lower status. He posits that the distribution of wages within companies tends to be flatter than would be predicted when wages reflect productivity because there is a wage premium to lower ranked individuals for feeling worse.

When applied to educational institutions, Frank notes that there are two countervailing pressures which influence the wage structure. Halo effects are the reputational benefits from being associated with an institution of high standing, while learning effects are the learning benefits from being in the presence of higher performing colleagues or peers. In addition, the slope of the hypothetical wage distribution should also be flatter in the educational context because the nature of the work tends to be self-directed unlike in other areas of the labour force, where functions are more hierarchical in nature.

2.2.3. Information Discovery

A third possibility for the high levels of demand for attending the selective schools is due to information asymmetry in the academic ability of students. The idea that educational attainment is a costly signal of productivity was established by Spence (1973) in the context of hiring by employers under uncertainty. Applied to this context, students are interested in attending selective schools because it signals their academic ability to others. Rather than communicating reputational status, attendance of the schools signals direct ability, including an indication of likely future educational and career outcomes. Admittedly, there is a close relationship between individual ability and being a member of a group defined by their ability.

A perhaps unintended consequence of increasing the information transparency in education or achievement is that the incentive for performance is diminished, as once a signal is attained, then the information content is transferred to the signal (e.g. education qualification, school reputation, etc.). In other words, once external markers of success have been obtained, such as attendance at an academically selective high school, the exertion of ongoing effort (such as maintaining grades) is less necessary because one’s ability can already be determined.

In addition to signaling to others, the academic ability of students may also be unknown to students themselves and their parents. As a result, being admitted to a selective school may provide an indication to students about their likely later educational and career outcomes. A simple calculation from the case study of the state cohort of students from Chapter 4 demonstrates this idea. For students placing in the top quartile of achievement in Year 7 standardized tests, the difference from attending a selective and attending a non-selective school in attaining results in the top 5%, 10% and 15% of Year 12 outcomes is large, at between 27 and 32 percentage points[23]. Note that this ignores all other factors, like sex, socioeconomic and language background, and differences in school environment. From only information on Year 7 achievement, attending a selective school suggests an increase in the probability of attaining a result in the top 5% in Year 12 by 26.7ppt, increasing from 16.6% to 43.3%.

If receiving an offer to attend selective schools is treated strictly as a form of information discovery (of students’ own academic ability), with the assumption that there are no

[23] For reference, 12.5% of this top quartile (15,127 students) apply for the selective schools and sit the entrance exam, while 5.3% of these students attend the selective schools.

differences in potential gains or decreases in achievement from attendance, then it would not be necessary for them to attend the schools. However, this seems to be an unlikely explanation since most successful applicants accept their offers and choose to attend the selective schools, while it is also probable that applicants who decline, decline for other reasons, possibly because they prefer their incumbent school.

2.3. Academic Achievement & Educational Psychology

This section summarizes the research on the influence of selective schools on the academic achievement of their students and reviews the related literature on self-concept and frame-of-reference effects. I additionally relate academic selection and self-concept to other research areas that have greater implications for longer-term outcomes in: self-identity, academic specialization and achievement motivation.

2.3.1. Academic Achievement at Selective High Schools

Much of the research on selective schools has concentrated on estimating the effect on the academic achievement of their students, and the associated mechanisms for such effects. Several more recent studies have taken advantage of access to admissions data to apply more advanced statistical techniques, finding little effect on achievement of students in high-income countries with choice-based education systems (from the USA and the UK), but there is evidence of positive effects in developing countries (i.e. Romania, Trinidad and Tobago) with system-wide academic selection and from a Chinese case study (details in section 3.2).

The focus of the empirical research on selective schools has also been extended to include the impact on student achievement from the expansion of the number of the selective schools and the link between selective schooling education systems and income inequality. Guyon, Machin and McNally (2012) found a negative effect for the academic achievement of students attending non-selective schools from an expansion in the number of students admitted to selective schools in Northern Ireland. In the UK, Burgess, Dickson, and Macmillan (2014) showed that income inequality within local government districts is higher for individuals educated under a selective schooling system, as opposed to a comprehensive system.

Clark and Del Bono (2016) noted the somewhat puzzling combination of ‘strong preferences and weak impacts’, in relation to the study of achievement effects at selective high schools, whereby the demand for attendance of selective schools is high but there has been little documented effect on achievement. They posited three explanations for the weak impacts and strong preferences, namely students and parents placing an overestimated importance on:

peer effects on achievement; non-academic outcomes, such as youth outcomes like crime; and longer-run labour market, education, and family outcomes, which are the focus of their UK study. They find positive effects on educational attainment for both men and women, but negative effects on fertility for women (a 20% decrease from a mean of 2 children).

The related research on the link between academic achievement and longer-term outcomes is consistent with the limited effects on achievement from attending selective schools (Clark and Del Bono 2016, citing others). Broader measures of school quality, including more advanced peers, higher paid teachers and more advanced curriculum, have similarly been shown to have little effect on longer term labour force and educational outcomes using instruments of student dates of birth (Dustmann, Puhani, Schonberg, 2012).

Earlier UK studies rely on regression and matching techniques (e.g. Coe et al. 2008, Schagen and Schagen 2002, and Fogelman 1984 reviewing others), which are subject to issues of positive unobservable selection in the characteristics of students attending the selective schools. Other earlier studies also observed that social or academic selection explained the differences in aggregate achievement. For example, Bonhomme and Sauder (2011) noted that the better performance of selective schools relative to non-selective ones was due to differences in pupils’ composition; they use difference in differences to compare the effects of selective and non-selective secondary education on student test scores in the UK. Similarly, Husen (1960) contrasted the highly selective filtering in the French education system with that of Sweden, noting that there were losses of talent due to social selection between stages of schooling, which he suggested would be less prevalent in comprehensive education systems, like that of the US.

2.3.1.1. Academic Selection & Tracking

Selective high schools and the system-wide allocation of students on an academic basis relate more generally to the degree of academic selection across education systems. Countries including Japan, Singapore and Croatia appear to allocate many of its students to high schools based on performance on achievement tests[24], while Germany, for example, tracks its students into different types of schools with different curricula: vocational and academic, as opposed to in comprehensive systems in which students follow the same curriculum. The degree of academic selection across education systems is an area of focus for the OECD in their PISA

[24] Figure IV.2.5. Chapter 2, Selecting and grouping students, PISA 2012 Results: What Makes a School Successful (Volume IV).

reports, parts of which concentrate on differences in socioeconomic inequality and academic achievement.

With a study of Kenyan primary school students, Duflo, Dupas and Kremer (2011) found tracking raised scores for all students, including those with lower achieving peers, from allocations into classes based on initial achievement compared with random allocations. From difference in difference comparisons of international test data for students in primary and secondary school, Hanushek and Woessman (2005) find that education systems with early tracking have greater educational inequality. In her review, Oakes (1987) suggests that some tracking systems confer a cognitive advantage to students placed in top tracks, while tracking systems appear to be detrimental to students not in the top track, meaning that the overall effect is to increase the initial differences in students.

2.3.2. Desire for Status and Relative Status

The importance of relative status to individuals, including their influence on individuals' performance outcomes, has been documented in many contexts in the domains of sports, education, and in the workplace. The influence of relative status in the competitive context is supported by the idea that the desire for status is a fundamental human motive, as explored by Anderson, Hildreth and Howland (2015) in their synthesis of the relevant literature. Chapter 5 builds on the idea that relative status has an influence in the competitive context, analysing the effect of perceived changes in local within-school ranks on later achievement for students when they transition from primary to secondary school.

There is persuasive evidence, reviewed in Frank (1985), that there is a strong biological and evolutionary basis for this universal phenomenon[25]. For example, individuals' blood pressure and heart rates have been shown to rise when speaking to individuals of higher status (Long et al., 1982); for groups of vervet monkeys, McGuire et al. (1982) found concentrations of serotonin, a chemical messenger which transmits signals between neurons, were almost 50 percent higher for dominant monkeys than for non-dominant monkeys[26].

The distinction between the desire for status and the influence of relative status on performance outcomes has been less well articulated and is discussed in the Chapter 5. Essentially, performance-related status implies social status to the extent that there is

[25] In support of humans' sensitivity to relative status is Dunbar's finding (1992) that the group size in primates, which for humans is 150, is a function of brain size (neocortex volume).
[26] Dominant monkeys exhibited more resting and eating behaviour and less vigilant and solitary behaviour. Demonstrating a causal mechanism between status and the biological response, re-arrangement of the 19 groups resulted in changes in serotonin, 72 hours after the change.

perceived instrumental value. i.e. that is individuals defer to others who appear to have greater utility (Anderson, Hildreth and Howland 2015). It is however possible that individuals' responses are purely a competition- rather than social status- related phenomena.

The importance of the desire for status in the education context, whereby students and parents choose schools based on their reputations, is discussed earlier in Demand for Academic Selection (section 2.2), and later in Self-Identity and Academic Selection (section 2.3.4).

2.3.3. Self-Concept and Rank-Order

Marsh (1987) and Marsh and Parker (1984) paved the way on the importance of academic self-concept, students' beliefs in their own abilities, when informed by a group frame-of-reference. Marsh identified the 'big-fish-little-pond-effect' (BFLPE), focusing on a specific instance of the group frame-of-reference effect, when an individual's achievement is high relative to the group. The effect, which is also described as a negative contrast effect (e.g. Marsh, Trautwein, et al., 2007), is typically demonstrated by measures of group achievement in regressions of individuals' self-concept, controlling for prior achievement (e.g. Marsh and Hau, 2003).

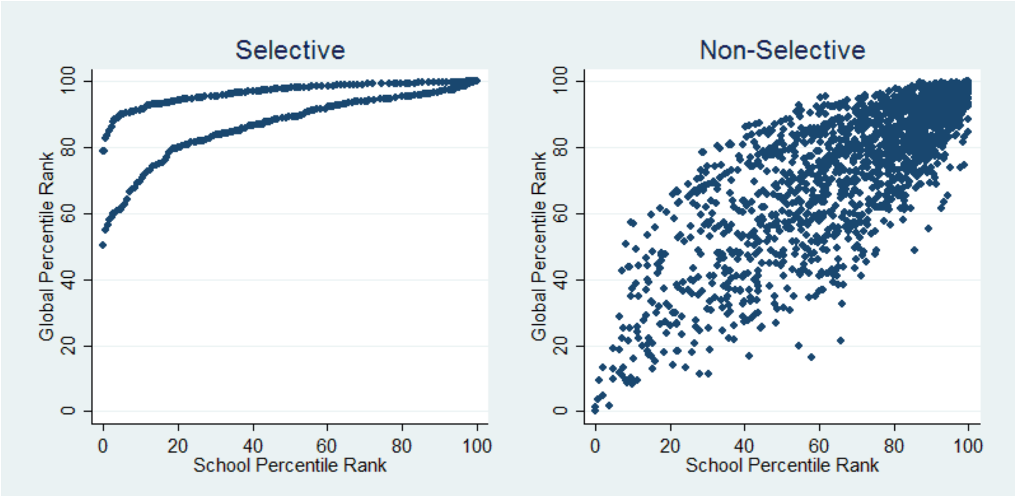
The research on self-concept and the BFLPE followed earlier work by Marsh, Byrne and Shavelson (1988), who demonstrated subject-specific self-concept for math and verbal achievements, and Shavelson and Bolus (1982), who initially found support for multiple facets of hierarchical self-concept in addition to general academic self-concept among junior high students. The theoretical basis for the group frame-of-reference research has been attributed to derivations from adaptation level, psychosocial judgement, social psychology, sociology, social comparison theory, and relative deprivation (Marsh, et al. 2008).

The influence of self-concept on achievement has relevance for students who attend selective schools due to the potentially large changes in relative academic status that they experience. An example of this is provided from the case study from Chapter 4 of selective schools from the anonymized Australian state. Using standardized test scores taken from before and after the selective school applicants sit the entrance exam, Figure 2.1 provides a sense of the magnitude in downgrades of status[27]. The scatter plots show both the global ranks defined over the state population and within-school ranks of Year 9 achievement of applicants who

[27] The percentile ranks are based on the average standardised Numeracy and Reading scores for each given year level, Year 7 or Year 9. Further details can be found in section 3.4.1 (referenced in section 4.3.1).

subsequently attended and did not attend the selective schools; the global ranks are calculated over the state population and only the ranks of the most and least selective schools are shown.

Figure 2.1: Global & Within-School Ranks of Year 9 Achievement of Selective School Applicants

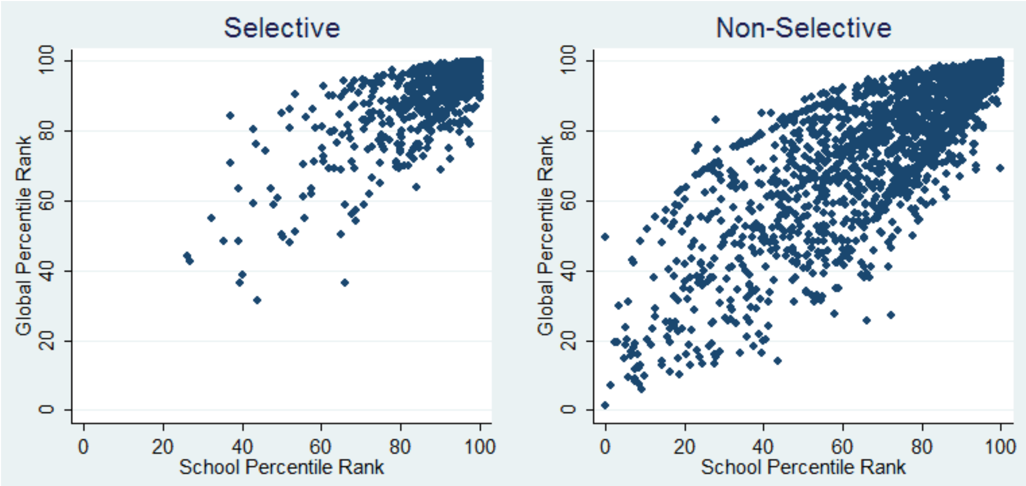


Note: Global ranks (defined over the state population) and within-school ranks of achievement are based on the average of standardized Year 9 Numeracy and Reading test scores, which are taken a few months after the selective students start at the selective schools. Only the most and least selective schools are shown.

By attending schools with high-achievement peers, selective students receive much lower within-school ranks (left panel), despite all having global ranks greater than the average of 50th percentile. This is quite different from the pattern in global and school ranks for the non-selective students (right panel). The difference between global and school ranks for non-selective students tended to be smaller than for the selective students and the differences were also less systematic, as the non-selective students attend many different schools.

Figure 2.2 presents the equivalent scatterplot for Year 7 test scores. For selective students, the mass of observations in the middle to upper right indicates that there is a degree of academic selection but that school percentile ranks follows a similar form to that of the non-selective students, being shaped liked a lens from the bottom-left to top-right.

Figure 2.2: Global & Within-School Ranks of Year 7 Achievement of Selective School Applicants



Note: Global and Within-school ranks of achievement are based on the average of standardized Year 7 Numeracy and Reading test scores.

The importance of academic self-concept in selective schools has been studied by Marsh, Trautwein, Ludtke, Baumert and Koller (2007), who examined whether a negative contrast effect persisted among selective students after graduating from high school in Germany. From the study of two samples, they found that mathematical self-concept was negatively associated with school-average mathematics scores at the end of high school, with also negative effects from academic school type. A small sample of students in gifted programs (29) from grade 4 in an Australian primary school were similarly found to experience declines in their academic self-concept (reading, math, and school), relative to a matched comparison group, but not in their non-academic self-concept (Marsh, Chessor, Craven and Roche, 1995).

The BFLPE or group frame-of-reference effect, is one of two identified effects on self-concept. The other effect is an internal/external frame-of-reference effect, whereby individuals focus on an internal comparison and conceive of their strengths in a relative sense across subject domains, despite potentially small differences in performance. Using international standardized test data, Chiu (2012) estimates together both group and internal/external frame-of-reference effects.

In a separate direction, the positive influence of academic achievement on self-concept and the interdependency between the two, described as 'reciprocal' effects, has also been investigated (e.g. Marsh, Trautwein, Ludtke et al. 2005 and Niepel, Brunner and Preckel 2014). This area of research in self-concept is closely related to the research on peer effects, which posits that exposure to high-achievement classmates is expected to improve student's achievement; the evidence for peer effects has been mixed, with typically weak effect sizes

(Sacerdote, 2011). The group frame-of-reference effect on achievement can be interpreted as the reverse in direction anticipated by peer effects.

The reciprocal nature of self-concept and achievement implies that the group frame-of-reference effect on achievement would be evident indirectly, whereby a student's self-concept is influenced by group achievement, which then affects their subsequent achievement. Marsh, Kong and Hau (2000), in their Hong Kong study, however, found a lack of effect from the inclusion of a measure of group achievement on later achievement, which they suggested could be explained by other factors relating to ability grouping, including the quality of education such as resources, curriculum and teaching expertise. They suggested that where negative achievement effects were present despite probable positive bias from other factors, these could be attributed to school-average achievement.

Later, Murphy and Weinhardt (2014) introduced a school-level rank-order measure of achievement to reflect the impact of self-concept on later achievement, finding evidence in support of group frame-of-reference effect. Their approach followed a line of research in economics that relates rank-order measures with individual outcomes. E.g. rank-order of income within a reference group and subjective measures of economic status (Powdthavee, 2009), or subjective happiness and changes in relative income status (Di Tella, et al. 2010). Chapter 5 of this thesis builds on Murphy and Weinhardt (2014), estimating the effect of changes in relative status (rank-order of achievement) on later achievement.

2.3.4. Self-Identity and Academic Specialization

Self-concept potentially reflects one facet of an individuals' set of beliefs relating to achievement motivation and could be linked with identity formation. Self-concept changes over time, decreasing from early pre-adolescence to middle adolescence, then increasing thereafter (as noted in Marsh, 1989).

Early adolescence, the time in which students enter the selective schools, is an important time for developing self-concept, identity and behaviours, attitudes and values (Savin-Williams 1979, citing Hartup 1970, Erikson 1959 and Kelley 1952). Savin-Williams (1979) found stable dominance hierarchies over time and settings for 8 groups of 12 to 14 year olds at summer camp. The status positions correlated strongly with rank orders of pubertal maturation, athletic ability, and group leadership.

With respect to occupational aspirations, Gottfredson (1981) proposed the formation of four stages of self-concept, with orientations to: size and power (age 3 to 5), sex roles (6-8 years), social valuation (9-13 years) and internal self (from 14 years). This supports the argument that attending the selective schools itself is a goal, as it helps with identity formation. In the opposite direction, Ahmavaara and Houston (2007) noted the potential for schools with poor reputations to negatively impact their students' self-esteem.

There is an argument that attending a selective school may be more likely to be a negative experience for students who have invested their sense of self into their intellectual or academic ability. For example, De Botton (2004) interprets William James as saying that individuals' goals will dictate their disappointment, and the ability to be satisfied with oneself does not depend on all areas of endeavor. Relatedly, Burhans and Dweck (1995) proposed the concept of 'contingent self-worth' to explain young children's achievement-based helplessness, whereby self-worth is derived from performance goals to obtain positive and avoid negative judgements. There is also anecdotal support for this idea at the college level, discussed by Gladwell (2013) in his book 'David and Goliath', whereby the director of admissions at Harvard in the 1960s implemented a 'happy-bottom-quarter' policy, which admitted students who have sufficient non-academic accomplishments to withstand the stresses of having lower relative academic status[28].

Cicala, Fryer and Spenkuch (2017) document short-term behavioural responses in support of individual specialization according to peer group from an experimental game-based study of middle school students from Texas. They find that participants respond to receiving low ranks either positively, by trying harder, or negatively, by affecting other participants, depending on the available built-in options in the game.

The influence of social context on achievement performance through individuals' identities is also demonstrated by the stereotype threat, which is when individuals are at risk of confirming a negative stereotype about one's group (Steele and Aronson 1995). Test scores of African American students were negatively affected when tests were presented as evaluative of verbal ability, as when compared with when presented as a psychology of problem solving (Steele and Aronson, 2007, describing their earlier work from 1995).

In the same vein, African American students were found equally likely at the individual level to drop out of the University of Michigan independent of SAT performance (Steele, 1997 cited in Dweck, 2000). On the same issue, Sander and Taylor (2012) find negative effects from affirmative action at elite law schools, whereby African American students with similar grades and test scores perform worse on several outcomes from attending more elite schools.

[28] p291 of Jerome Karabel (2005), cited in Gladwell (2013).

Elliot, Strenta, et al. (1996) and Smyth and McArdle (2004) have also documented the increased drop-out rates of minority students from universities with greater levels of selectivity (by SAT) in the Science, Technology, Engineering and Mathematics (STEM) fields. Cole and Barber (2003) similarly examine the issue from the perspective of the under-representation of minority individuals teaching at university faculties.

The racial stereotype threat appears to be related to a general social comparison phenomena; Davis (1966) found that college majors at liberal arts colleges were associated with students' relative academic performance (GPAs), which suggests that the group frame-of-reference effect is significant for decision making for academic specialization and potentially occupational choice. Cole and Barber (2003) suggest that Davis' finding on the importance of social comparisons (1966) was supported for African American students in their survey data of liberal arts college students, with both test scores and grades, but not for students in other ethnic groups.

The racial stereotype threat appears to share many similarities with negative stereotypes that girls may hold about STEM-subject choices. For example, Mouganie and Wang (2017) find that having a higher proportion of high-performing female peers increases STEM major choices for girls, while there is an opposite effect from having more high-performing male peers. Further, in a fascinating study of female primary school teacher's math anxiety, Beilock, Gunderson, Ramirez and Levine (2010) found that girls in first and second grade were more likely to endorse stereotypes that boys were good at maths and girls were good at reading, the higher the levels of teacher anxiety. Leslie, Cimpian, Meyer and Freeland (2015) find supporting evidence for the hypothesis that: women are under-represented in attaining PhDs in fields where the belief that natural ability is important for success is more commonly held.

Another possibility distinct from the explanation provided by social comparisons among students is that the pattern of academic specialization and GPAs is a consequence of students' exposure to university level humanities subjects, which they had previously thought not be interested in; this is consistent with the aims of liberal arts colleges to provide a broad education.

2.3.5. Achievement Motivation

Dweck (2000) and colleagues posit two theories of intelligence which fit in with her meaning-system approach to achievement motivation. That is, individuals lean towards holding one of two types of beliefs: a theory that intelligence is a fixed entity, or conversely that intelligence is malleable such that it can be increased with effort; an 'entity' or 'incremental' theory of intelligence, respectively. This appears counter to the definition of general intelligence, the abstracted factor from intelligence tests comprising sets of different types of questions, there is reason however to think that the tested levels of intelligence can be developed, particularly given that there have been large increases in aggregate levels over time (Flynn, 1999).

Diener and Dweck (1978, 1980) identified two types of responses to failure: master-oriented and helpless patterns of behaviour, referring respectively to a process of learning for increasing competence, and an orientation to achievement goals that concern the winning of judgement of individuals' competence. There is a strong resemblance of the learning and performance goals to the concepts of intrinsic and extrinsic motivation, which place value on the act of engaging in activities and on external rewards, respectively (Deci and Moller, 2007, citing White, 1959 and Deci, 1975).

There are implications for high-achieving students from Dweck's theories of intelligence: Dweck (2000) comments that when students are labelled (for example, as gifted or high-achieving), they may become overconcerned with the label and begin to react badly to setbacks. In addition, under an entity-theory framework student peers are competitors for self-esteem, meaning that the negative effects from entity-theories may be particularly harmful at selective high schools.

Models of achievement motivation are particularly relevant in the study of selective high schools due to the grouping of high-achieving students, as well as the typically large changes in peer group environment from attending the schools. Based on Dweck's theories, Ahmavaara and Houston (2007) interpreted admission to selective schools for students as supporting individuals' future aspirations, and failure to receive admission as negative feedback on later achievement. They compared the aspirations and self-concept related measures of 11-12 year old and 14-15 year old students from both selective and non-selective schools, finding support for Dweck's implicit theories. Ahmavaara and Houston find a significant effect for the interaction between incremental theories of intelligence and future aspirations for students at selective schools, but not for students at non-selective schools. They suggest this to mean that entry to selective schools reinforces students' initial beliefs, leading them to have even higher levels of aspiration.

2.4. Gifted and High-Achievement Students

In this section I distinguish between gifted and high-achievement students and discuss the level of academic selectivity at the schools, which I suggest informs the selective schools' relevant measures of success. The former is relevant as changes to selective school policies appear to have been made in response to demand for educating gifted students, while at times references to academically gifted and high-achievement students are equated in policy descriptions (such as in NSW, for example). I also discuss creativity and expert performance, which are closely related fields to giftedness research, and may help identify possible benefits to gifted students from attending selective schools. As a bridge to the later section on the apparent demand for selective schools within trends in education policy (section 2.6), I explain how skewed rewards to top performers in creative fields can help motivate the economic narrative of education.

2.4.1. Selectivity

Different states fund different numbers and types of selective schools, and accordingly, the level of academic selectivity observed in Australian selective schools varies by state. From the case study in Chapter 4 that is the focus of this thesis, half of the students in the most competitive selective school were in the top 2% of students statewide on the Year 9 standardized tests, while a third were in the top 1%. Even at the least competitive selective school[29], the median student was at the top 11% statewide, while the academically weakest test performance was still at the median statewide, meaning that the school was still very selective.

Taking the top 11% level of achievement as a benchmark in relation to the university entrance results, variation in the outcome have potentially important consequences for the course and university institution that students could attend in Australia; entry to most courses are conditional on university entrance results and the minimum requirements are determined by student demand relative to the number of positions available. For example, receipt of a Year 12 result in the top 5% would be close to the cut-off for entry to a Bachelor of Commerce at a top research institution.

At the very high end of the achievement distribution, however, an argument can be made that the Year 12 results have less importance as they are expected to be high regardless of the school that the student attends. At that level of achievement, longer-term educational and

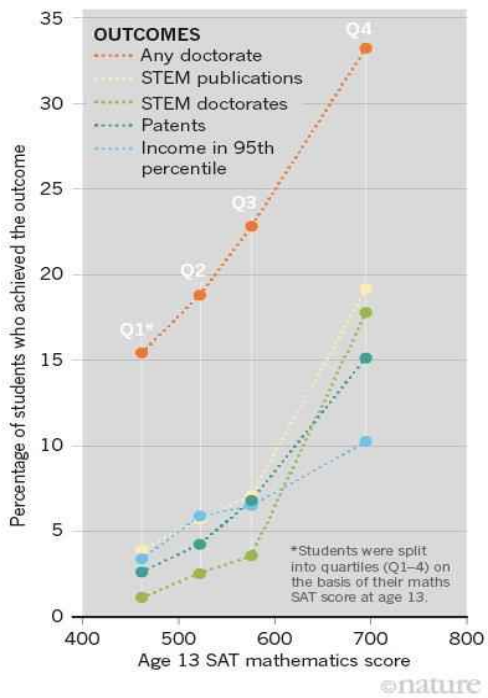
[29] The student intake appears to have had atypically low levels of achievement because the selective school was in its first year of operation.

labour market outcomes would appear to have greater relevance. It is probable that a significant proportion of students attending the academically selective high schools are drawn towards research careers.

An indication of possible later educational outcomes for the highest-achieving students at selective schools can be drawn from results from the Study of Mathematically Precocious Youth (SMPY), where subjects were chosen for scoring in the top 1% of the Mathematics section of the SAT[30] at age 13. As well as attaining a university degree, other relevant educational indicators of success could include postgraduate qualifications, patents and publications, while career success could be measured in terms of types of occupation, career satisfaction, or financially.

Figure 2.3 presents the relationship between quartiles of test score and long-term educational outcomes of: publications, doctorates, patents; and attaining income in the 95th percentile; the figure is reproduced from a Nature news feature and is derived from Robertson et al. (2010). Even in the bottom quartile of SMPY subjects, a significant percentage of 15% achieved the longer-term outcome of 'any doctorate'.

Figure 2.3: Long-Term Educational Outcomes and SAT Mathematics Performance



Source: reproduced from 'How to raise a genius: lessons from a 45-year study of super-smart children (nature), originally from Robertson et al. (2010). How to raise a genius: lessons from a 45-year study of super smart children, Tom Clynes, Nature news feature, 7 September 2016.

30 the Scholastic Assessment Test, a test of academic ability.

The results from the standardized tests are not directly comparable to those from the SAT because the former is intended to measure academic progress; Sternberg (2007) makes the point that ability and achievement-oriented tests are different qualitatively due to their purpose but are essentially the same. As the standardized tests place a greater emphasis on accumulated curriculum knowledge, it potentially also reflects conscientiousness to the extent that the content required recent learning. Conventional intelligence tests measure achievement that individuals should have accomplished several years back (Sternberg 2007, citing Anastasi and Urbina 1979).

2.4.2. Gifted and Talented

This section discusses whether gifted and talented students are relevant to selective schools as implied by some research (e.g. Chan, 1996 and Gross, 2005) and selective school policies in NSW.

2.4.2.1. Background

Gifted and talented students are students who are usually identified by their intelligence as observed from academic performances or tests, and sometimes by additional personality characteristics. There is, however, some ambiguity surrounding the definition of gifted and talented, and particularly its identification. In the United States, there is a trend away from relying on intelligence tests only towards more inclusive policies for determining gifted eligibility (Card and Guilano, 2014 and McClain and Pfeiffer, 2012)**.**

The cause of the ambiguity can be partially explained by tracing the development of the research on giftedness that commenced with the identification of intellectual talent in children through intelligence testing in the 1920s, by Terman and Hollingworth (see e.g. Jolly 2009, Fox, 1981 and Chapman 1988); Terman (1926) sought children who scored in the top 1% of the Stanford-Binet Intelligence Scale (cited in Jolly, 2009), and Hollingworth (1942) undertook case studies of children who tested with IQs above 180. The definition of giftedness was subsequently expanded in later decades to include other identifying traits or factors, including a link to creativity. More recently, several models of giftedness have emerged which relate giftedness to additional domains or categories beyond intellectual ability.

For example, the U.S Office of Education adopted a definition of giftedness relating to six categories: general intellectual ability, specific academic aptitude, creative or productive thinking, leadership ability, visual and performing arts aptitude, and psychomotor ability (Renzulli 1978). Fox (1981) noted that the emphasis in policy practice was on the first two categories of giftedness, as is the case here with respect to the academically selective high schools.

Another perspective of giftedness is provided by Gardner (introduced in Gardner 1983, cited in Karolyi, Ramos-Ford and Gardner, 2003) who defined 8 multiple intelligences, consisting of: linguistic, logical-mathematical, musical, spatial, kinesthetic, interpersonal, intrapersonal, and naturalist.

The more widely cited models of giftedness have the common goal of explaining the personality traits and behaviours associated with individuals who attain the highest levels of performance (see related section on expert performance, section 2.4.3); Renzulli (1978) describes giftedness as the interaction between superior general ability, task commitment and creativity; Gagne (1985) identifies motivation as the catalyst for ability to manifest as talent, interpreting the term 'gifted' as reflecting ability or potential, and interpreting 'talent' as reflecting performance or achievement. Tanenbaum (2003) classifies developed talent into categories of produced or performed, and creative or proficient.

The view of gifted individuals and their personality traits has largely been informed by both longitudinal studies of children identified as gifted through intelligence tests and from studying historical geniuses or eminent personalities[31]. The salient personality traits of gifted students in the view of their parents are: sensitivity, perfectionism, intensity and introversion (Silverman, 1997). Another characteristic that is used to describe both creative and gifted students is a tolerance for ambiguity, although evidence for the support of the trait is not strong (Merrotsy, 2013); Merrotsy traced the origins of this notion back to Frenkel-Brunswick (1948, 1949), who studied its opposite, 'dichotomous conceptions', as part of an anti-fascist agenda.

Like Silverman (1997), Piechowksi (1997) connects emotional giftedness to moral sensitivity and developmental potential through Dabrowski, who defined, in his concept of developmental potential, five components of psychic life: psychomotor, sensual, intellectual, imaginational, and emotional overexcitabilities[32].

[31] Dabrowski (1970) analyses biographies of eminent personalities, cited in Silverman (2008). Catherine Cox, a doctoral student of Lewis Terman's, also reviewed the biographies of eminent adults to study giftedness, noted in Simonton (2003).

[32] Dabrowski (1992, 2006, and 1970) developed his psychological theory of positive disintegration based on clinical experience with intellectually and creatively gifted individuals, cited in Silverman (2008).

2.4.2.2. High-Achievers

Given a definition of giftedness based on intellectual ability, the concept of giftedness is likely most relevant at the most selective of selective high schools. That is, both the number of students who are gifted, and the extent to which they are gifted, are correlated with the level of selectivity of each specific school. For example, at the most selective school in NSW, the entry level was estimated to have an IQ equivalent of 145[33], placing their students roughly in the top 0.5% of the population.

Referencing a guide for levels of intellectual giftedness as defined by IQ ranges (refer to Table 2.1), at least half of the students who attend selective schools in the case study fall into the categories of mildly or moderately gifted, which are IQ ranges of 115 to 129 and 130 to 144, respectively[34]. This is based on a rough approximation of IQ using selective students' results on Year 9 standardized test scores. i.e. assuming a normal distribution for the Year 9 test scores, a score in the top 11% of students, which half of students at the weakest selective school attained, translates to an IQ of 118 (roughly the mean plus 1.23 SD).

Table 2.1: Guideline Categories of Intellectual Giftedness and IQ Ranges

Categories of Giftedness	IQ Range
Mildly Gifted	115 – 129
Moderately Gifted	130 – 144
Highly Gifted	145 – 159
Exceptionally Gifted	160 – 179
Profoundly Gifted	> 180

The difference between the categories of moderately gifted and highly gifted, separated by an IQ of 145, can be likened to the distinction between gifted and high-achieving students in personality descriptions which are independent of IQ. The general description of high-achievers (e.g. motivated, conscientious) is counter to that of gifted students (e.g. creative, sensitive, introspective). The perception is that high-achieving students attain the same levels of performance as gifted students through effort, discipline and conscientiousness. One view was that high-achievers were fast learners who were easily programmed, and potentially

[33] Daily Telegraph (January 18, 2009), the 'school where every kid's a genius'. Jasmine Le Rade and Sharon Labi, The Sunday Telegraph.
[34] Gross (2005) expanding on Kline and Meckstroth (1985) and Webb, et al. (1982); also attributed to Gagne (2000, 2003) by Gross (2005), p27.

creative in a predictable manner, whereas gifted students were creative about a better developed body of knowledge[35].

2.4.2.3. Difficulties Experienced

Gifted students appear to face obstacles of a different nature from non-gifted students, such as social interactions, boredom or motivational problems, and learning difficulties. Towers (1987) discussed the finding that after a certain point IQ does not correspond with longer-term career outcomes, in an article called 'The Outsiders'[36]. Simonton (2003) makes a similar point regarding exceptional achievement and IQ, pointing out that William Shockley did not score high enough to be admitted into Terman's study but eventually co-invented the transistor, for which he received a Nobel Prize. Drawing from Terman's longitudinal study of children with IQs above 140, Towers noted that the incidence of social maladjustment was positively correlated with performances on verbal intelligence tests.

Towers also pointed out that a socially optimal IQ range of around 125 to 155 was identified by Hollingworth. Hollingworth described a limit in communication of about 30 points in IQ, which Towers terms a 'communication range', meaning that: a) extremely gifted students are more likely to be socially isolated; and b) leadership breaks down when there is a large difference in intelligence between the group and the leader. Simonton (1985) formalized this concept, suggesting models of influence based on intelligence, accounting for communications limits imposed by comprehension difficulties and openness to criticism from intellectual superiors. His model suggests a socially optimal IQ of 119, and he estimates leader-follower gaps of around 8 and 20 points.

At very high levels of giftedness, Gross (2005) citing Hollingworth (1942) suggests that students are likely bored with school, while several researchers have also described the benefits to exceptionally gifted students from following an accelerated program (Colangelo, Assouline, and Gross, 2004). For example, among exceptionally gifted students from the SMPY study, with mean IQs above 180 (1 In 10,000 people), above almost 40% thought educational acceleration had a positive influence on educational and career planning[37]. In her discussion of appropriate curriculum for gifted students, VanTessel-Baska (1985) suggested that successful gifted

[35] Paraphrased quote from Associate Professor John Munro. Education and Training Committee (2012), Inquiry into the education of the gifted and talented students, Parliamentary paper No. 108, Session 2010-2012.
[36] It is possible that he was drawing from personal experience; he was a security guard and member of a high IQ society. 'In Memoriam: Grady Towers', Kevin Langdon. http://megasociety.org/noesis/149/towers.html, accessed 23 November, 2017.
[37] Lubinski (2004) describing findings from Lubinski, Webb, et al. (2001).

programs included content acceleration and would emphasize inter-relationships on bodies of knowledge, self-directed learning, and producing rather than consuming knowledge. She also argued that depth of learning was important to gifted students.

The positive association between SAT performance and longer-term outcomes from Robertson et al. (2010), discussed earlier, can be explained by the likely positive selection of those students from their socioeconomic backgrounds, which likely assisted their intellectual development. In her case study of 15 exceptionally gifted students, Gross (2005, Chapter 11) noted that very few of her subjects would have been identified as gifted from models that relied on motivation or task commitment. The theme of unrealized potential or underachievement reoccurs in descriptions of gifted students (e.g. motivation chapter of Webb et al. (1982), or the Psychosocial Development chapter in Gross, 2005). Though dated, Lemov (1979) estimated that up to 15-30% of high school dropouts were gifted and talented, cited in Webb et al. (1982).

Aside from the social difficulties described above, the method of instruction and simple material may contribute to the lack of motivation in gifted students. Lovecky (1994) suggests that exceptionally gifted students (IQ > 170) find the complex to be simple, because they understand the underlying pattern and cannot break down the component parts, and that the simple is complex due to the possibility of multiple interpretations to questions and from their preference for precision.

Finally, it appears that some gifted children have learning disabilities, such as dyslexia and ADHD, or prefer a visual-spatial style of learning, which Silverman found from her research on the visual-spatial abilities of gifted children (Silverman 2002). Silverman attributes the visual-spatial learning style to the strength of the right hemisphere of the brain, and the auditory-sequential learning style to the left hemisphere, which is responsible for word retrieval. Silverman suggests around one third of the school population are visual-spatial learners[38].

[38] pxvii Silverman (2002) and Silverman (2000).

2.4.2.4. Creativity

Creativity is frequently described as an important facet of giftedness. For example, Feldhusen (1985) suggests that creativity and giftedness are synonymous in the sense they have in common the capacity to produce ideas or solutions which are unique and worthwhile[39]. However, creativity is arguably poorly understood, as was identified by Gruber (1982), as he observed that the study of creativity was approached through the study of: giftedness; exceptionally creative individuals; and the creative processes of ordinary individuals[40].

Amabile (1983) defined creativity as displaying novelty and usefulness, with the additional constraint that the task or product is not algorithmic in nature. Carson, Peterson and Higgins (2005) described creative achievement as being facilitated by a confluence of intrapersonal factors, such as familial resources, societal and cultural factors, and interpersonal factors, such as the capacity for divergent thinking[41], intelligence, intrinsic motivation and talent. Many, including Rogers (1962) and Maslow (1954), have equated creative development to self-actualisation, striving for one's potential, cited in Davis (2003).

Perhaps more contentiously, there appears to be some association between the incidence of mental disorders and distinguished achievers, who are by definition creative (Simonton 2003, citing others). In contrast, academic performance in early high school was found to be positively associated with conscientiousness and negatively associated with psychoticism, which is one of Eysenck's three dimensions of personality (Heaven, Ciarrochi, and Vialle, 2007), the others being neuroticism and introversion-extraversion; Eysenck (1992) suggests that low psychoticism is associated with empathy, socialization and co-operativeness, while traits associated with high psychoticism included being impulsive, aggressive and schizophrenic.

It has also been posited that creativity is facilitated by intelligence, but the empirical research is mixed. For example, Jauk, et al. (2013) provide evidence from IQ and creativity tests for the threshold hypothesis that an above-average level of intelligence is necessary for high-level creativity. Amabile (1983), citing others, also noted that levels of creativity appear to be low at lower values of intelligence, from correlations of IQ and creativity scores. In a meta-analyses, Kim (2005) found that the mean correlation between creativity test scores and IQ scores was

[39] See Introduction section of Feldhusen (1985).
[40] Gruber (1982) defined an approach to view creative individuals as loosely coupled systems evolving through their life history, as organisations of knowledge, affect and purpose.
[41] Demonstrating the difficulty in this area, there is little empirical evidence of divergent thinking correlating with creative performance (Gruber 1982, citing Barron and Harrington, 1981). See also Runco's (1993) discussion of the significance of divergent thinking for creativity and giftedness despite criticisms of the validity of the tests.

small at 0.174 across 21 studies, with no support for the threshold theory that intelligence is positively associated with creativity up to a certain threshold.

Creatively gifted individuals were found to have acuity of the senses, with pronounced neurological responses to certain types of stimuli, while giftedness can also be thought of as asynchronous development arising from advanced cognitive abilities and heightened intensity (Dabrowski 1965, 1972, and 1974, cited in Silverman 1994). A study comparing a group of artistic and intellectually gifted individuals with a comparison group of graduate students supported the interpretation that the groups could be distinguished by differences in the dimensions of overexcitability and (emotional) stability (Piechowski, Silverman and Falk, 1985)[42], with larger values for the former group.

Additional empirical support for differences in measures of overexcitability between gifted and non-gifted students is provided by several studies and graduate theses using a 21-item overexcitability questionnaire (Silverman 2008). The idea that creativity may be associated with neurological responses accords with other research which shows that personality traits have a biological basis. e.g. Kagan, Reznick, and Snidman (1988) distinguished between quiet and spontaneous children at 21 and 31 months, and the differences in behaviour were also observed in derivative forms at later ages of up to 7.5 years old. The biological basis for differences in IQ and other traits of gifted children is also consistent the observation that the characteristic modes of thinking differ between moderately gifted and exceptionally gifted children as suggested by Lovecky (1994).

[42] A third of graduate students in the survey of 42, showed an identical overexcitability profile to that of the intellectually gifted. Piechowski and Colangelo (1984) have a similar small study of gifted adolescents, gifted adults (identified from intelligence tests), graduate students and artists.

2.4.3. Expert Performance and Giftedness

Both research on giftedness and expert performance are interested in the transformation of talent to performance or achievement. These two areas of research reflect an apparent dichotomy between innate ability ("giftedness") identified from tests and expert performance which attributes success to deliberate practice and the development of ability. The significance of persistence is also recognized in the latter, with a consensus view that persistence plays a greater role than intelligence in attaining great levels of achievement (e.g. Simonton 2003, citing Cox 1926).

There are however some unresolved ambiguities between the areas as demonstrated by a conflict regarding the importance of natural ability between Gagne (2009) and Ericsson, Nandagopal and Roring (2009). It is probable that the emphasis on deliberate practice in the studies of expertise results from the large degree to which there is pre-existing selection such that most participants already have a reasonably high level of underlying ability.

From a meta-analyses of studies, Macnamara, Hambrick and Oswald (2014) found that deliberate practice accounted from between 1% and 26% of variance in performance in individual-oriented domains, with: 21% for music, 26% for games, 18% for sports, 4% for education and 1% for professions. The low variance explained appears to be associated with measures of performance that are standardized in nature, and not reflective of creativity; studies in the education category mostly measured course grades or grade point averages, while the professions category included: computer programming, military aircraft piloting, soccer refereeing and insurance selling.

Sternberg (2000) interprets the tests of abilities used to define giftedness as reflecting developing expertise, observing that neither classification implies the other, and suggesting that gifted students require the continual development of abilities to remain gifted. He describes as an example that a student who can cite historical facts may appear gifted but is unlikely to be later identified as a gifted historian, without the accompanying analytical or creative skills.

With applicability to the selective schools, Sternberg suggests the conception of giftedness is very narrowly defined, whereby ability-test scores predict a similar kind of competence, in school grades or achievement test scores; although this competence is less important for job performance, these measures reflect overlapping expertise in the kinds of skills used to define experts; the characteristics used to define expertise are multi-faceted, and in relation to performance on ability tests includes determining how to represent test problems in addition to searching and executing a strategy.

Researchers have found expert performance to be associated with the development of talent over long periods of time, with estimates of several hours a day for over a decade or 10,000 hours (Simonton 2003 citing others, and Runco 1993 citing Hayes 1978 and Simon and Chase, 1973). Simonton (1991a and 1991b) suggests a model of career development where productivity is a function of onset of individuals' creative potential, onset of career, age, and a two-step process of ideation and elaboration.

In a widely-cited study, Ericsson et al. (1993) emphasized the importance of deliberate practice for the acquisition of expert performance, from their study of violinists at the Music Academy of Wester Berlin. They find that students with the potential for careers as international soloists practice the same amount as violinists with lesser ability, but show greater intent in their habits. They spend more time practicing alone, have more sessions of practice, slept more, and practiced more regularly between 10am and 2pm.

Simonton (2003) also noted some relevant idiosyncracies of distinguished achievers: there appears to be a relationship between birth-order and domain of expertise, e.g. the over-representation of first-borns in Terman's study of gifted children, and the role of traumatic events; a large proportion of eminent personalities experienced traumatic events in their childhood.

The identification of gifted students appears to have potential benefits from a societal and economic perspective. However, an unappreciated aspect of exceptional performance by giftedness research is that it appears to be a result of a rare combination of high intelligence and drive, and possibly creativity, of which there may be little correlation with intelligence. E.g. Getzels and Jackson (1962), cited in Gross (2005), find no correlation between intelligence and creativity for individuals with IQs above 120, which is approximately the top 10% of the population. See also Creativity section (section 2.4.2.4), earlier.

Simonton (2003) asserts that expert performance is a rare occurrence for the reason that several normally distributed factors need to coincide to produce the outcome. As a consequence the distribution of exceptional performance is log-normal, which is the distribution resulting from multiplicative relationship of the normally distributed factors. He provides the example of Price's law, where the square root of the total number of 250 composers of classical music, which is roughly 16, produce half of the total output.

The beliefs that individuals hold or have developed appear to be influential in the translation from initial ability to later performance. Dweck's (2000) distinction between fixed and incremental theories of intelligence is particularly relevant (discussed earlier). Measures of performance in creative or scientific endeavours appear to require a sustained interest that incorporates long-term perseverance and discipline. These habits have not necessarily been developed in individuals with innate ability and there are indications that these challenges may be compounded by additional social problems that are unique to these individuals.

2.4.4. Skewed Rewards to Top Performers

Interest in transforming potential into performance through the education system is a natural extension of a narrative of economic and technological growth; theoretical models of economic growth often assign importance to education through human capital as a key input (e.g. Lucas 1988), which is typically reflected by educational attainment. This is also reflected in the increasing attention paid to results from international and national standardized testing[43], and the longer-term trends of higher rates of high school completion and university participation in developed countries[44]. Interpreting these trends together, the emphasis by policymakers on academic performance and other educational outcomes form part of a narrative where the economic growth of countries is linked to the educational attainment of its populations (e.g. Woessman and Hanushek, 2007).

The importance of achievement motivation has even been posited as an explanation for variation in economic growth between countries and over time; McClelland (1961) ambitiously attempts to show that achievement motivation is associated with economic growth. The research interest in giftedness and expert performance can be further motivated by the economics of markets characterized by skewed rewards to top performers, which has been described as the economics of superstars.

Rosen (1981) identified that certain domains have the common elements of: personal rewards being connected to the size of an individual's market; and both market size and reward to be skewed towards the most talented people; there is imperfect substitution between best and next best performers. These markets are characterized by a small number of individuals

[43] Nationally standardized tests were introduced in 2008 in Australia for year levels 3, 5, 7 and 9, along with associated school level reporting that is publicly available (see Data Description, section 3.4). The Australian government has also incorporated a goal in the Australian Education Act 2013 to place among the top 5 countries in reading, mathematics and science by 2025 in PISA (Masters 2016).

[44] E.g. In the USA, 41% of 18 to 24 year-olds were enrolled in postsecondary institutions in 2012, increasing by 15% since 1970; noted in Mouganie (2015). Kariya (2011) observes that Japan is ahead of other high-income countries in having universal access to higher education policy (reaching near 75%), and discusses social equality in accessing secondary and higher education.

attaining most of the market, or generating most of the output. For instance, the decreasing costs in distribution of entertainment along with the development of more internationalized markets, means that sports and music have expanded the market reach of the participants, resulting in even larger gains to top performers.

Technological developments, in the internet and mobile phones for example, have also created conditions that disproportionately reward the top performers such that marginal producers in such markets are likely to struggle more as the trend continues. A useful example is provided by the software company, WhatsApp, which provides messaging services via the internet. By leveraging a programming language with origins in industry research and a low rate of adoption (Erlang), the company was able to service 450 million users with only 35 engineers in 2014, when it was acquired by Facebook for $19 billion[45].

It is quite possible that the confluence of technological progress, which inherently disproportionately rewards top performers, and the globalization of the higher education industry contributes to the demand for greater specialization in secondary schools and colleges. Note that this may be desirable from a societal perspective with a net positive benefit if specialized schools play a role in helping the top performers to subsequently contribute new innovations.

Another example demonstrating the economic impact from the link between technology and education are patterns in aggregate cross-country data which show that incentivizing individuals towards different occupations may increase economic growth (Murphy, Schleifer and Vishny, 1991). Murphy et al. take Rosen's line of thought further, suggesting that certain occupations are rent-seeking (i.e. lawyers) in nature while others (engineers) contribute to increasing economic output.

[45] Why WhatsApp only needs 50 Engineers for its 900m Users. Cade Metz, 2015.09.15, Wired, https://www.wired.com/2015/09/whatsapp-serves-900-million-users-50-engineers/, accessed 11 September 2017.

2.5. Social & Cultural Influences at Selective Schools

This section discusses the potential social and cultural influences on academically gifted students from attending the selective schools and having similar peers and includes a review of media and sociological commentary on the selective schools relating to the issues of hyper-selectivity and aspirational immigrant culture. The section concludes by discussing the extent to which the selective schools cater to academically gifted and high-achievement students, as informed by the research on giftedness.

2.5.1. Social Benefits

There appear to be social benefits to academically gifted students from attending selective schools. A study by Milgram and Milgram (1976) found that compared to non-gifted students, gifted students from grade 4 to 8 initially had positive self-concept, more internal locus of control, and lower levels of general anxiety and test anxiety. However, they observed a decrease in self-concept of the gifted group in adolescence which they attributed to a shift in attitudes by gifted and non-gifted students towards each other, as interest patterns diverged. Consistently, gifted students in a different study also reported lower levels of self-concept than other students in Year 7 in the domains of social and physical activities (but higher levels for academic self-concept) (Brounstein, Holahan and Dreyden, 1991).

One possible explanation for the decrease in self-concept of gifted students is that there is a social norm towards low achievement among adolescents (Dweck 2000, citing Covington 1992, Juvenon 1995, and Ogbu 1991); Dweck (2000) explains the formation of this peer pressure by the combination of students' heightened sensitivity to academic performance and a fixed mindset towards achievement that leads to low effort.

The perception that gifted or high-achieving students have weaker social skills could also be explained by delayed social development among gifted students, such as from Dabrowksi's concept of asynchronous development. Perceptions of social skills also appear to be informed by subject interests. Handel, Vialle and Ziegler (2013) found through surveyed opinions of high school students of their hypothetical peers that perceptions of high-achievers are informed by the subject in which the achievement is attained. E.g. High-achievers in mathematics and science subjects were thought to be less socially-minded than high-achieving peers in sports.

The prevalence of social isolation in gifted students is also supported by the finding that a friend is a top priority for gifted children (Mendaglio, 1993, cited in Masden, et al. 2015). Gross (2005) also describes, similarly, the tendency for intellectually gifted children to seek out children with similar mental ages or who are at similar stages of intellectual development[46]. This social pattern also appears to hold at the college level; from self-reported data, students at an elite selective college were more likely to be better acquainted with other students with a similar academic background (Arcidiacono, Kahn, and Vigdor, 2011).

Selective schools would then appear to confer social benefits to gifted students, which are larger for students with higher levels of intellectual ability who tend to experience social isolation. These students are, as discussed earlier, also more likely to be socially maladjusted. From a probabilistic perspective, it is reasonable to assert that most selective students are high-achievers and are not gifted (or only mildly gifted), falling into the socially optimal range of IQ[47], meaning that the social benefits are small for most students. They may, however, receive more leadership opportunities at selective schools, as implied by the existence of an IQ communication range.

One deficiency of the selective schools is that the gifted students who would receive the most benefit socially, in a counterfactual sense, are probably least likely to attend because they have not been identified as gifted. Gross (2005) makes the same point that students who are identified for gifted programs are typically from middle-class and professional families, who value learning and scholastic achievement. This positive selection into gifted programs by families from advantaged socioeconomic backgrounds also explains the positive relationship between Mathematics SAT test performances and longer-term educational outcomes.

2.5.2. Student Population & Cultural Influences

For gifted students, if providing social benefits represents one of the strengths of selective schools, then an admissions mechanism that relies on achievement tests contributes one of its main potential weaknesses: a relatively homogenous student population with high levels of educational aspiration and competitiveness. A strong cultural influence is anticipated from having a large proportion of students of both immigrant background and from an advantaged socioeconomic background, independent of the specific school.

Over 75% of selective students from the case study in Chapter 4 were from an immigrant background, which is similar to the levels found in selective schools in NSW[48]. Around 50% of applicants attended non-government schools[49], which are typically attended by students with

[46] p134 and p135, citing: Davis (1924); Burks et al. (1930); Hollingworth (1931); also O’Shea (1960) and Hubbard (1929).
[47] As inferred from their performances on standardized tests (see Selectivity section, section 2.4.1).
[48] Over 80% of selective students were from an immigrant background in NSW (Ho, 2017).
[49] Source: author’s calculations from NAPLAN data.

a more advantaged socioeconomic background, compared with 34.6% for the student population in 2016[50].

Students who attend selective schools appear to be highly motivated as signaled by their intent from applying to academically specialized schools, and because students of immigrant background, for whom studies have documented both higher levels of educational attainment and educational aspiration (e.g. Gemici et al., 2014 and Homel et al., 2012), have a large representation at the selective schools. Several studies have attributed the academic outperformance of students from an Asian background largely to greater effort that is attributed to cultural factors (see for example, Hsin and Xie 2014 and Jerrim 2015).

The academic outperformance of students of immigrant background may be only one instance of broader social phenomena. In the American context, Chua and Rubenfeld (2014) suggest, perhaps contentiously, that the successful outcomes of certain groups of immigrants can be attributed to three psychological factors: a superiority complex that may have religious or historical origins, insecurity that derives from being an immigrant, and impulse control; these factors appear to have a predominantly cultural basis. One example provided is the suggestion that as much as two thirds of black undergraduate students at Harvard were West Indian and African immigrants or their children[51].

Returning to the interpretation of selective schools that cater to gifted students, the educational aspirations and cultural beliefs held by a largely immigrant student population do not seem compatible with some of the ideas described in giftedness research. Specifically, intrapersonal factors like intrinsic motivation, confidence, and non-conformity have been linked with creativity (cited in Carson, Peterson and Higgins, 2005). Contrary to the traits in this description, the most obvious point of conflict is the emphasis on extrinsic markers of success at selective schools, which suggests a reduced emphasis on intrinsic motivation. Similarly, the strength of cultural factors from a homogenous student population is likely to lead to a concentrated school culture that is conformist in nature.

An example of the narrow focus of certain immigrant groups of students is provided by Cole and Barber (2003), who found from surveys that educational success in the view of parents of Asian background was interpreted narrowly in terms of elite (Ivy-school) college attendance and specific occupations, particularly, the field of Medicine.

---

[50] Australian Bureau of Statistics: 4221.0 – Schools, Australia, 2016.

[51] Top Colleges Take More Blacks, But Which Ones. Sara Rimer and Karen W. Arenson. New York Times (June, 2004). http://www.nytimes.com/2004/06/24/us/top-colleges-take-more-blacks-but-which-ones.html, accessed 11 November 2017.

Further, the combination of traits in gifted students is such that the intense competition encouraged by selective school environments is potentially a poor fit. At the very least, a competitive environment is implicit in the intention of the schools. In support of this interpretation, Dabrowski suggested individuals with similar characteristics, i.e. sensitive, nonaggressive, intelligent and creative individuals, were oppressed in societies oriented towards competition, power, status and wealth (1964, 1967, 1972) (cited in Silverman, 1994). Gifted individuals also appear to be more susceptible to existential depression arising from their introspective nature (Webb, 2008).

It is possible that there are other cultural influences from the selection of students into selective high schools that are unrelated to an achievement frame of success. For instance, in their study of self-concept and academic achievement, Marsh, Kong, and Hau (2000) noted the collectivist culture in Hong Kong, which has been shown to be associated with smaller negative group frame-of-reference effects, contrasting it with that of Western individualist cultures.

2.5.3. Hyper-Selectivity and Polarization

The importance of the selective school policies is heightened in NSW, where there is a greater presence of these schools (17 fully selective), and where the controversial and polarizing topic is further developed from the public policy discourse in both media and academia.

Perhaps the main source for controversy of the selective school policies is the extent of social segregation into schools by language background in NSW, despite a long history of students of immigrant background attending selective schools[52]. At the upper limit, 97% of students at one selective school had a foreign language background, which incidentally was the selective school with highest entry requirement[53].

This concentration by language background appears in even starker contrast when considered within certain local areas, than when considered by school sector. By sector, foreign language background levels were 52% for government schools, 22% for independent schools and 37% for catholic schools (Ho, 2011), whereas Ho (2015) calculates that for the local area of Sydney's lower north shore, 'a wealthy and culturally diverse region', 49% of students at public high schools had an immigrant background, compared with 13% for private schools, and around 90% for the two selective schools. The increased attendance of private schools by non-

[52] Many students of European Jewish background attended selective schools as early as the 1930s. Source: Selective schools' long and tangled history with race and class. The Conversation. Helen Proctor and Arathi Sripakash, https://theconversation.com/selective-schools-long-and-tangled-history-with-race-and-class-74614, accessed 10 November 2017.

[53] From Ho (2017), Angry Anglos and the aspirational Asians: everyday multiculturalism in the selective school system in Sydney.

immigrant students in areas of high immigration has been documented previously, by Mavisakalyn (2012) from the 2001 Census.

The prevalence of tutoring in preparation for attending the selective schools by predominantly students of Asian background is another contributing factor to the negative perceptions held by some parents and students (Ho, 2017); from semi-structured interviews of both present students and their parents, and recent students who attended the selective schools. Private academic tutoring in addition to schooling is common practice in the East Asian countries, as noted by Wise (2016) for Singapore. Gee has described the situation in Singapore as an educational arms race[54], citing Frank's research on positional goods (see section 2.2.3.2, earlier).

The polarizing views caused by the selective schools is potentially worse than just a simple transfer of cultural attitudes and behaviours from competitive educational environments. The selectivity of Chinese migrants in the USA has been described as 'hyper-selectivity' by Lee and Zhou (2015, 2017), due to positive selection from immigration and in terms of educational attainment. These immigrants have higher levels of educational attainment than both individuals from their country of origin and the general USA population. They identified that 51% of Chinese immigrants graduated from college, compared with only 4% of individuals in China; a similarly high rate of nearly 40% for China-born and South Korean-born individuals was noted by Ho (2017) for Australia. Lee and Zhou also highlight contrasts of Asians comprising 6% of the US population against a fifth of students at Ivy League universities.

Students attending selective schools who are second generation immigrants could also be considered an accentuated form of Lee and Zhou's (2015, 2017) hyper-selectivity, given the additional level of academic selection through the admissions process. Supported by interviews of second generation immigrants, Lee and Zhou posit a more nuanced explanation for Asian American achievement than Chua and Rubenfeld (2014), who sought to explain immigrant success more generally (see previous subsection, 2.5.2).

Although they identify the prevalence of cultural beliefs for explaining achievement, such as the emphasis on effort in Asian countries, consistent with Chua and Rubenfeld, Lee and Zhou also suggest that the hyper-selectivity is supported by institutions due to reinforcing stereotypes of the 'model minority' (relating to the educational achievement of Asian Americans), and place importance on the concept of ethnic capital (Borjas 1992); Borjas

[54] Christopher Gee, Singapore and the Educational Arms Race, public lecture given at the Human Capital and Education for Asian Development Foundation, 22 September 2016, http://headfoundation.org/2016/09/22/singapore-and-the-educational-arms-race/.

suggests that skill development of immigrants is influenced by the quality of the environment, as measured by mean levels of skill or earnings of an ethnic group.

An interesting implication of Lee and Zhou's work for the selective schools is that when individuals do not share the cultural beliefs of the group, for example, a group of gifted students, then they do not benefit and appear to experience social isolation. Lee and Zhou additionally point out that not matching the success frame (the narrow definitions of success), can result in a perceived sense of failure for individuals.

A theme from Ho's interviews of 'Anglo-Australian' students and parents at selective schools was the lack of a sense of belonging and social isolation they experienced from being a minority group. Other issues noted were the potential for contagion introduced by tutoring (echoing a similar thought to Gee), while a negative impact on quality of life was also noted due to the tutoring and emphasis on extra-curricular activities, which were described as excessively competitive and instrumental in approach. These views were similar in nature to the negative perceptions of tertiary educated parents towards high-achieving Asian students in a gentrifying area (Butler, Ho and Vincent, 2017), who themselves placed a greater emphasis on local community and reciprocity.

2.5.4. Catering to Academically Gifted & High-Achieving Students

Although teaching can be targeted to a narrow ability range[55], contrary to possible expectations, there appears to be little academic benefit for students from attending selective high schools (see Academic Achievement, section 2.3.1). Selective schools may, however, provide social and emotional benefits to their students, particularly academically gifted students. For example, a school environment that encourages academic achievement could help students feel more connected or accepted, as high-achieving and gifted students appear to experience greater social difficulties during secondary school.

The intellectual ability of peers is important to gifted students as they are more likely to experience social isolation, such that they tend to seek out friends who are at similar stages of intellectual development. A similar more general pattern also exists at university, with students being more likely to be drawn towards others with a similar academic background.

One potential source of contradiction in the selective school policies is that the range of academic aptitude of students admitted extends over a large range that does not correspond to the guideline IQ definition of giftedness. As noted earlier, a rough approximation from Year

[55] As commented on by Coe et al. (2008) in their own analyses of selective schools (grammar schools) in the UK.

9 standardized scores suggests that many selective students can be placed in the mildly gifted category, denoted by a range of 115-129 of IQ. At this level of academic ability, students are less likely to experience the social difficulties discussed earlier. At exceptionally high levels of intellectual ability, researchers suggest that gifted students receive an accelerated educational program, which is not a function of the selective schools.

A better generalization may be that the selective schools cater to high achieving students who have high levels of educational aspiration and motivation. This is a view which is supported by the documented relationship between educational aspiration and students of immigrant background, who have a (disproportionately) large representation at the selective schools, and are of predominantly Asian background. There appear to be cultural factors behind the educational aspirations of these students, including the narrow interpretations of success and the competitive educational environments associated with additional tutoring. This social phenomenon is the result of at least two levels of selection, selection at immigration and a second level of selection in terms of educational aspiration.

The cultural factors that influence selective students of immigrant background, which also affects other selective students indirectly, conflicts with descriptions of giftedness related to creativity and intrinsic motivation due to an excessive emphasis on competition and extrinsic markers of success. The shortcomings of the aspirational culture are also relevant to the goal of transforming potential into performance for individuals, depending on their achievement motivations. While an academic environment may suit some highly motivated individuals, it is possible from an expert performance perspective that formal education interferes with more important pursuits. For example, Simonton (2003) notes that voracious reading, which he describes as an undisciplined activity, is correlated with success in many fields, offering the interpretation that when formal schooling is not directly contributing to mastery, it is hindering it.

On the positive side, students may receive confirmation of their educational aspirations from attending the selective schools, and experience positive reinforcement of their goals from being surrounded by similar, like-minded peers. It is probable that students who attend selective schools have successful longer-term outcomes, such as moderately successful careers in fields like business, law, medicine, the military and academia, which Feldman (1982) suggested was predicted by high IQ in Terman’s foundational study of intellectually gifted students (top 1% IQ). Finally, there may be benefits from the selective schools to competitive individuals, who are motivated by the characteristic of competitiveness with peers, which is a trait that they share with highly successful artists, pianists, scientists, and swimmers and tennis players, as noted by Bloom (1982), cited by Feldhusen (1985).

2.6. Demand for Selective Schools & Policy Trends in Education

The current high levels of demand for selective high schools can be interpreted in the broader context of several long-term trends. First, the expansion of the education system appears to be continuing following the earlier expansion in the secondary school system, with further increases in the minimum school leaving age, higher benchmarks for high school completion rates, and increased participation in university education[56]. Second, having adopted a competition-based education system with free choice, there is a trend towards greater levels of sorting or specialization. E.g. Meadmore (2001) characterizes this trend as a devolution which undermines the original intentions of the free and compulsory assumptions of the education systems.

A more recent trend is the increased transparency in standardized testing that is accompanied by an implied accountability. In support, nationally standardized testing for which aggregate results are reported was introduced in 2008, while education policy also increasingly emphasizes international standards. For instance, there is a stated target in the Australian Education Act (2013) of reaching performance on PISA among the top 5 countries by 2025 (Masters 2016). The reporting of achievement performances and demographic information of specific schools has arguably led to increased local social segregation, simultaneously increasing and decreasing demand to higher and lower performing schools respectively. This is evidenced by the introduction of catchment areas at certain high demand schools, whose purpose is to restrict the student intake to local residents.

The last trend is technological progress and increasing globalization, whose influence is far less apparent as it is indirect. The discussion on markets that disproportionately reward a small number of individuals indicated that the advances in technology can have a significant impact. It is quite possible that this last trend is one of the reasons for the increasing importance that appears to be placed on education.

Earlier, Butts (1955) provided an outsider's perspective on the assumptions underlying the Australian education system. At the time, secondary schooling was hierarchical in the informal

[56] For example, the Council of Australian Governments set a target in 2008 to lift the Year 12 or equivalent attainment rate to 90% by 2015. Council of Australian Governments Report on Performance 2016, March 2016. https://www.coag.gov.au/sites/default/files/files/report_coag_2016.pdf, accessed 11 November 2017.

streaming of students towards academic or non-academic secondary schools, with types including technical, agriculture, and academic high schools, while there was also a similar hierarchical approach towards subject selection, where it was assumed that the most able students would specialize in English, Mathematics, Science and possibly foreign languages.

In his comments on teacher education, Butts (1955) noted a neglect of the social and philosophical foundations of education. This is similar to the view that there are unappreciated consequences from an overemphasis on an economic interpretation of education, whereby education is viewed as a means for increasing economic growth (see Skewed Rewards to Top Performers, section 2.4.4). An alternative function of education is the instilment of cultural norms that increase social cohesion (Gradstein and Justman, 2000). Rich and De Vitis (1992) similarly discussed the implicit assumption of competition in education, exploring themes from economics, psychology and philosophy, and non-competitive theories of education.

Within the context of the broader policy trends in education, the selective high schools appear well positioned with an argument for increasing the number of the schools being strongly aligned with the present narratives of economic growth and competition. The high levels of demand for the schools are also supported and potentially driven by the economic narrative of education, given that attendance is determined by test performance and because the schools are regularly topping annual school rankings of university entrance results.

The view of education in the early part of the 20^{th} century was more practical, with the establishment of different types of secondary schools, and recognition that the general population had predominantly vocational interests; building character and preparation for citizenship were identified as functions of secondary education (Whiteman 1973, cited in Braggett 1985). Roe (1953), who undertook a psychological study of eminent researchers, also shared a similar view that the aim of education was for the development of citizens for society, and not for producing scientists.

The influence of an egalitarian and anti-intellectual sentiment in Australian culture on education policy has been noted by Gross (2005, Chapter 2) in relation to resistance to the idea of gifted education programs. The expansion of the education system and the trend towards specialization of schools has unintended consequences in that although there may be an egalitarian intention, the result is that greater importance is placed on the reputations of schools, and universities. Braggett (1985, p17) noted a similar irony for the expansion of the secondary system increasing the reputations of academically oriented high schools.

Chapman (1988, p8) noted in his historical review of achievement testing in the United States, that the use of achievement testing had been interpreted as perpetuating, or failing to reduce, social inequality by Bowles and Gintis (1976), who attributed the development to intelligence testing. There is a duality to the standardized testing in that they can help measure progress towards increasing levels of performance, but the counter position is that the value that is placed in the signal or status results in a decrease in incentive for performance once the status has been obtained (e.g. Macleod and Urquiola, 2011).

A duality in interpretation can similarly be applied to the schools themselves. As well as representing a goal for which to strive, once students have been sorted into schools, the reputations of the schools may mean that the subsequent achievement of students within those schools has less importance. Together, the trends of standardized testing and their transparent reporting, along with the increasing specialization of schools and greater importance placed on their reputations, potentially undermines individuals' motivations by placing an overweighting towards extrinsic markers of success.

Policy issues in education are presently primarily concerned with achievement performance and socioeconomic inequality, and are motivated by the success of individuals and the provision of opportunities for success (e.g. Lamb et al. 2015 and Masters 2016). The polarizing nature of education is aptly demonstrated by calls for greater specialization in schools, such as for the creation of gifted primary schools analogous to those at the secondary level[57], particularly when it is contrasted with a relative decline in performance on international tests for Reading, Mathematics and Scientific literacy on PISA, and decreasing participation rates in Year 12 for advanced mathematics and physics (as shown in Masters 2016).

This trend of declining achievement performance in Australia appears to be a worrying accompaniment to the increasing importance of STEM for economic and individual success[58], and inevitably prompts questions about the direction of education policies. However, the costs of increasing competitiveness that are paid by some of the most successful countries are perhaps only starting to be recognized. For example, three of the top five performing countries

[57] Pallavi, Singhal, NSW should have 'selective primary schools' for gifted children: academic. Sydney Morning Herald. July 22, 2017. http://www.smh.com.au/national/education/nsw-should-have-selective-primary-schools-for-gifted-children-academic-20170720-gxev5i.html, accessed 27 July 2017.

[58] See for example, the US Department of Education (2007), cited in Becker and Park (2011) in their meta-analyses of STEM studies, identified that 75% of the fastest growing occupations required STEM skills.

in Science on the 2015 PISA[59] were East-Asian countries (Singapore, Japan and Chinese Taipei) characterized by intense educational competition that includes significant additional tutoring[60]. In particular, Gee (2012) has described Singapore, first-placed on PISA in Science, as engaging in an educational ‘arms race’, referencing Frank’s (2007) discussion of positional goods, whereby parents are pushing up house prices to send their children to good schools and where there is an emergence of academic credentialism. Gee has also pointed out that Singapore has introduced policy with a vision of ‘every school is a good school’ to counteract the pressures of competition[61].

As a final thought, a critique of the focus on economic growth has been offered by Tcitovsky (1992) and Harrod (1958). Tcitovsky notes the limitation of an economic approach which assumes rational individuals, and increasing utility with consumption, due to observations in psychology that it is the contrast in experience that creates satisfaction (e.g. food satiates only when an individual is hungry)[62]. Harrod makes the interesting point that there are limitations in two types of satisfaction which are not available to the majority, these being: wealth consisting of physical material that is defined by scarcity and rents, like land and natural resources; and social scarcity.

Harrod also noted that the standard of living can only be raised to the extent that material goods can be mass produced, as services are priced relative to the labour of other people. One implication of this that has relevance to trends in education policy is that as competition and technology drives down the price of mass produced goods, then it is the positional goods, where there is scarcity and visibility that increases in value, in a relative sense. A culture that has a tendency towards placing increasing importance to positional goods, including in education on school and university reputations and the signaling of academic achievements, may also help explain, along with technological trends and economic incentives, the broader educational policy trends.

[59] Source: PISA 2015 Results in Focus. There are 35 OECD countries and 37 partner countries or economies (p xviii, Thomson, S, De bortoli, L, Underwood, C, PISA 2015: a first look at Australia’s results. Australian Council for Educational Research).
[60] A survey found that 80% of Singaporeans believed that tuition was beneficial to children’s education and 67% have or previously had enrolled their children in tuition (Source: You Know Anot? Private Tuition in Singapore: A Whitepaper Release, Blackbox Research, July 2012). The report also noted that tuition was described as a ‘minor national obsession’ by Prime Minister Lee.
[61] Source: From a description of ‘Singapore and the Educational Arms Race’, public lecture given by Christopher Gee on 22 September 2016. The Human Capital and Education for Asian Development Foundation. http://headfoundation.org/2016/09/22/singapore-and-the-educational-arms-race/
[62] Incidentally, in his later critique of his own work, Tcitovsky (1996) states: “education’s most important function is to civilize, i.e, instruct in the harmless activities of life so as to divert people from harmful, violent ones.”

Chapter 3. Matching from the General Population

3.1. Introduction

This chapter is Part 1 of the main analyses that addresses the question of whether selective schools improve their students' university entrance results. The chapter establishes the foundation by identifying a suitable comparison group to the high-achievement students at the selective schools and introducing a matching method which accounts for differences in the prior achievement and background characteristics between selective and non-selective students.

This first set of analyses follows the cohort of Year 9 students from an anonymized Australian state in 2008, including students at a limited number of selective high schools[63]. The challenge is the counterfactual – finding a group of students most like those that enter selective schools. Given admission to selective schools is conditional on academic performance I limit the control group to those with high prior achievement. For measuring prior achievement, I use nationally standardized tests from 2008, and for the outcome variables I use Year 12 university entrance results from 2011.

For estimation, I adopt Propensity Score Matching over simpler methods as a regression-based approach for its efficiency with small samples and also because it is more intuitive than largely equivalent nonparametric regression approaches. I model attendance of selective high schools separately for girls and boys using information relating to students' language and socioeconomic backgrounds as well as their results on Year 9 standardized tests. I subsequently compare selective and non-selective students on the basis of similar probability of attendance, finding that: students with higher Year 9 test scores, a foreign language background and, to a lesser extent, a higher socioeconomic background (SES) increased the likelihood of attendance.

Previewing the results, I find positive estimates for selective school attendance, with larger effects for girls than for boys, but note that comparisons do not account for the probable higher levels of aspiration and motivation of students attending selective schools. Comparing selective and non-selective students, I find improvements are between 6 and 15 percentage points (ppt) in girls' probability of attaining Year 12 results in the top 5%, 10% and 15% of all students. For the boys I find evidence of smaller effects, of over 5 ppt increases in probability for attaining results in the top 15%.

[63] The number of schools has been anonymized for privacy reasons. Most students are aged 14 or 15 in Year 9 and 17 or 18 in Year 12.

The sample for the main analyses was chosen to balance obtaining the most suitable control group against retaining sufficient observations. To gauge the robustness of the PSM results, I conduct sensitivity analyses by refining the samples to obtain non-selective students who most resemble the selective school students. I also impute missing outcome values and re-run the PSM analyses with the inclusion of these students; in the main analyses I excluded students with missing outcome data rather than treating them as having not achieved a score in the top 5%, 10% or 15%. The robustness checks supported the main results, though the results from the sensitivity analyses were mostly smaller in effect size.

Finally, to complement the PSM analyses, I also undertake difference-in-differences analyses within SES quartiles to produce estimates of local selective school effects. The difference-in-difference analyses suggested that gains for attaining Year 12 results in the top 10% and 15% of all students were large for both girls and boys, and that there was a negative effect for boys in the highest socioeconomic quartile in attaining results in the top 5%.

This chapter is structured as follows. Section 2 is the literature review, which discusses academic outcomes, peer effects, self-concept and rank effects, and the aspirations of immigrants. Section 3 describes the admissions process at the selective high schools, including the different admission intakes, and defines the data sample. Section 4 contains a description of the administrative data for both the standardized tests and the Year 12 assessment, as well as details for the derived socio-economic index. Descriptive statistics are presented in section 5 for: treatment and control groups, achievement outcomes and missing data, sector differences (government and non-government); and also selection into treatment. Section 6 describes the methodological approach, propensity score matching, and the rationale. The results are presented in section 7. The robustness checks are in section 8. Section 9 discusses the results, its limitations and the study's place in the literature. In the conclusion, section 10, the research is summarized and suggestions for future related research are offered.

3.2. Literature Review

A clearer picture of the achievement gains at selective schools has emerged in recent years, with several papers having been published based in different educational institution settings. While research interest prior to this was beset with problems of how to estimate plausible counterfactuals for the academic outcomes of high-achieving selective school students, these recent papers have accounted for the problem of counterfactuals by exploiting the random variation in exams that is created by the thresholds from which offers are made.

The recent studies have produced mixed results, and institutional differences between the studies, including the nuances relating to the high school completion outcomes, mean that interpreting the combination of results can be difficult. Where there have been positive results on achievement (e.g. case studies from Pop-Eleches and Urquiola (2013) for Romania, from Jackson (2010) for Trinidad and Tobago, and from Hoekstra, Mouganie and Wang (2016) for an anonymised Chinese city), it appears that benefits from selective schools are obtained in the context of system-wide academic allocations of students, ranging over the full distribution of achievement, as opposed to the selective schools existing as a small part of comprehensive education systems in high-income countries, like Australia, USA and the UK. [64]

The positive effects found in institutional settings with system-wide academic selection appear to correspond partially to behavioural responses, which were first identified by Pop-Eleches and Urquoila. Pop-Eleches and Urquoila find through surveys that teachers sort in a manner consistent with preferences for high-achieving students. This is demonstrated by Hoekstra et al. (2016), who were able to fully account for their positive effects of selective school attendance with differences in teacher quality between schools. Other behavioural responses identified by Pop-Eleches and Urquoila were that of parents reducing effort when their children attended better schools[65], and also that of students who were successful in attending more selective schools realising that they were weaker and subsequently feeling marginalized.[66]

In comprehensive education systems the existence of selective schools appears to be a puzzle. Despite high levels of demand for attendance of selective schools, little positive effect has been established for academic achievement at high school completion or on college outcomes.

[64] Estrada and Gignoux (2017) observed that the benefits were limited to low- to middle-income countries but do not explicitly make the connection to the differences between system-wide and limited academic selection within choice-based comprehensive systems.
[65] Marginally attending selective students received less homework related help from their parents.
[66] This theme was discussed in Chapter 2: Self-concept and Rank-Order (section 2.3.3).

This combination has been described aptly as ‘strong preferences and weak impacts’ by Clark and Del Bono (2016), who offer several possible explanations for the demand from students and their parents, including: a misunderstanding of achievement growth and peer effects, and the importance of non-academic or longer-term outcomes.

There are indications that all three explanations contribute to the high levels of demand; from an earlier Australian survey (Brathwaite and Kensell, 1992), students reported that an academic emphasis, including benefits for future careers, and also the social, cultural and sporting reputations of the schools were important factors influencing their decisions to attend the schools. In the UK, Clark (2010) documented students taking more advanced subjects in selective high schools, which are pre-requisites for certain university courses, but attributed this potential benefit to the academic curriculums at the schools. Later, Clark and Del Bono (2016) extended the focus of the research from short-term education outcomes to longer-term education, labour market and family outcomes.

Due to the stratification of students introduced by selective schools, whose main consequence for students is a sharp increase in their peer quality, peer effects are inevitably identified as a potential mechanism for explaining the impact of these schools on their students. In comprehensive systems, the high representation of students of immigrant background in selective schools suggests that their higher levels of educational attainment are also relevant to understanding the strong preferences for selective schools. This explanation is also supported by subsample analyses producing positive local effects for minority students in the American case studies from New York and Boston (Adulkadiroglu et al., 2014). There is also evidence that disadvantaged students receive greater achievement benefits from a North Carolina case study (Shi, 2017).

There is growing research that shows there are benefits to the academic achievement of students from attending higher achievement schools in institutional settings with system-wide academic selection[67]. The positive effects appear to reflect a form of resource allocation in countries where there are limited resources or an uneven distribution of resources[68]; the results have been attributed to differences in teacher quality, and also the sorting of teachers consistent with the better teachers being matched with higher achievement students. Across

[67] Lucas and Mbiti (2014) with analyses from Kenya appear to be the exception that proves the rule. For an earlier stage of schooling, Zhang (2013) similarly finds little benefit from attending an oversubscribed selective middle school in China for high school entrance exams or admission outcomes, but relies on a lottery mechanism for estimation rather than variation from exam cut-offs.

[68] Positive results were obtained in low to middle income countries. Park et al. (2015) provide the example of ‘magnet’ selective high schools in rural China.

all studies, differences in peer quality have consistently been shown to not affect achievement. In comprehensive education systems the demand for selective schools poses a different set of questions, with the absence of evidence for positive effects on achievement ruling out the explanation of benefits to students in the form of improved achievement outcomes.

3.2.1. Academic Outcomes

Coe et al. (2008) summarise over twenty studies of the impact of selective schools in the United Kingdom going back to the 1980s. These selective schools, known as grammar schools, operate as part of the public school systems in the United Kingdom. While critical of the methodology of all of the studies summarised in Coe et al. (2008), the authors conclude that the literature points to relatively small positive impacts of selective schools on student achievement. Their own estimates, based on regression and multi-level analyses using sex, prior achievement, ethnic background and school meal status to determine attendance, similarly pointed to positive effects on later school achievement for secondary school students.

Abdulkadiroglu et al. (2014) assessed the academic contribution of selective high schools to the educational outcomes of later test scores and the college quality attended by their students using regression discontinuity methods. For selective schools in New York and Boston, they found no general effect and scattered gains to test scores to students from attendance, which included a modest localized effect on English scores for minority students. Using entrance exam scores as an instrument for peer composition, Adulkadiroglu et al. also found no effect from changes in peer composition on test scores.

In contemporaneous work, Dobbie and Fryer (2014) found little impact of attendance of selective schools in New York on college enrolment, graduation or quality for marginal students. Like Adulkadiroglu et al., Dobbie and Fryer applied a regression discontinuity (RD) approach, where students just above and below the cut-off determining offers are compared. Similarly, Clark (2010) adopted an instrumental variable approach with entrance exam results and assessed the effect of attendance at grammar (selective high) schools in the United Kingdom, and found only small effects from attendance on test scores but some evidence on course-taking and university enrolment.

In some countries, almost all students are allocated to high schools on the basis of admissions exams. Pop-Eleches and Urquiola (2013) and Jackson (2010) investigated the effects of attending higher achievement schools for Romania and Trinidad and Tobago, respectively. Pop-Eleches and Urquiola (2013) find that students with access to higher achievement schools perform better on a high stakes graduation test[69]. With the RD approach, there are 2,000 cut-offs, one minimum score on the entrance exam for each school, or 6,000 cut-offs if tracks within secondary schools are included. Instrumental variable results from Trinidad and Tobago (Jackson, 2010), where all students are assigned to schools after grade 5 on the basis of achievement tests, also showed large positive benefits on examination performance at the end of secondary school from admission to higher performing schools[70].

Hoekstra et al. (2016) conduct regression discontinuity analyses, drawing a sizeable sample from an anonymised Chinese city of 10 million, and find large and significant effects for university entrance results of 0.16 SD for the most selective tier of selective schools. They attribute all of the selective school effect to improved teacher quality, as reflected by meaningful discontinuities in access to better quality teachers (i.e. having obtained the highest ranked and paid class of teacher in the Chinese education system).

Clark and Del Bono (2016) extended the focus on academic achievement towards longer-term outcomes, when individuals were in their mid-to-late forties from cohorts educated in the 1960s. They find large impacts on completed education from selective school attendance. For men, they find little impact on labour market outcomes in income and employment. For women, Clark and Del Bono find increases in income and also document a negative impact on female fertility. In a similar vein, Estrada and Gignoux (2017) obtain approximate effects on earnings by analysing the effects from attendance of 16 selective schools in Mexico City on college outcomes, and hence implied increases in earnings. They find that the selective students are four times as likely to graduate from an elite college, and three times as likely to obtain an engineering degree (off low bases of less than 1% probability).

3.2.2. Peer Effects

The positive selection of students into schools based on academic performance, and in other circumstances socio-economic advantage, might create a positive externality via peer effects, such that it leads to learning-conducive environments and better educational outcomes; peer effects may include a variety of indirect mechanisms relating to teachers, staff and school resources, which are difficult to disentangle. Conversely, the loss or absence of high achieving

[69] The positive effects from attending higher ranked schools and tracks within schools are as high as 0.02 to 0.10 of a standard deviation (SD) on the Baccalaureate, which is a pre-requisite for applying to university.
[70] Students attending secondary schools with peer scores half a standard deviation higher were associated with passing between 0.28 and 0.34 more graduation exams, completing 5 of which is a requirement for attending university.

students from other schools may weaken the factors which are associated with, or contribute, to positive externalities.

The vast literature on peer effects in education finds only modest effects on test scores (see the review in Sacerdote, 2011). Angrist (2014) asserts that the results from many studies are overstated due to methodological deficiencies (bias from weak instruments) that cannot overcome the non-random selection of students into schools. With respect to selective schools, the large change in the composition of peers from attending a selective high school provides an ideal setup for assessing the influence of peer effects, but little evidence of positive peer effects has been found to date (e.g. Adulkadiroglu et al.).

Although the relationship found between academic outcomes and peer effects has not been strong, there is accumulating evidence that there might be heterogeneous effects (although this has not been shown for test results at the end of high school). That is, higher-achieving students are disproportionately influenced by peer effects in a positive way; examples cited by Sacerdote (2011) include: Lavy, Paserman and Schlosser (2012), Duflo, Dupas and Kremer (2011), and Lavy, Silva and Weinhardt (2012) for the case of high achieving girls. Lavy, Silva and Weinhardt (2012) also observe that all students are adversely affected by low-achieving students.

Outside of cognitive outcomes, studies have found substantial effects for social outcomes such as drinking, drug use and criminal behaviour (summarised in Sacerdote, 2011). Regardless of which specific outcome is influenced by peers, students and parents act as if there are positive peer effects. A higher demand for schools with higher test scores is reflected in house prices (Davidoff and Leigh, 2008), and similarly demand for schools with a student population from an advantaged socioeconomic background is evident in commuting behaviour. [71]

The 'Big-Fish-Little-Pond' effect discussed by Marsh and Hau (2003) is a potential explanation for some of the heterogeneous peer effects found in the literature, especially as they relate to selective schools. They suggest that an application of social comparison theory, whereby students compare their own achievement with their peers, leads to students having lower academic self-concept if they attend schools with really high-achieving peers. From the

[71] A significant percentage of students travel long distances to attend primary and secondary schools with higher SES students, in the northern suburbs of Melbourne (Lamb, 2007). On a related note, one of the critiques of selective schools policy is that it results in 'residualisation' or 'cream-skimming', and contributes to the concentration of socioeconomic disadvantage due to the removal of high achievement students from other schools (Lamb, 2008). Residualisation refers to the process arising from the preferences of students for socioeconomically advantaged schools, whereby the socioeconomically disadvantaged schools gradually decrease in size and have high concentrations of disadvantage, which both potentially lower achievement.

Program of Student Assessment (PISA, 2000) that administers tests to around 4,000 fifteen year-old students across 26 countries, Marsh and Hau found a negative association between individual and school average achievement.

Similarly, using administrative UK data from 2001 to 2005, Murphy and Weinhardt (2014) expanded on this concept by estimating a positive effect of classroom rank position of students test results in primary school on their subsequent secondary school test results. In the opposite direction, Jonsson and Mood (2008) found somewhat disconcertingly that having high-achieving peers reduces students' aspirations for attending university, by a small amount.

3.2.3. Aspirations of Immigrants

The high representation of students of foreign language background in Australian selective high schools is also a characteristic of selective schools in New York and Boston, as inferred by the ethnicity of their students. In the three older New York selective high schools, between 57% and 67% of students were Asian, and a further 3% to 8% were Hispanic (Dobbie and Fryer, 2014). American selective students were however from a lower socio-economic background, which is typical of American inner-city residents, than in the Australian context; between 39% and 62% of students at the New York selective schools were eligible for free or reduced lunches. In this case study over 50% of students at selective schools were in the top quartile of SES (Table 3.4, section 3.5 Descriptive Statistics).

It is probable that the high representation of students of immigrant background at selective schools is closely related to the ambitions of these students; the link between the language background of students and their educational attainment has been established by several studies (see also Chapter 4: Educational Aspiration, section 4.6.4). Gemici et al. (2014) found language background and academic performance, along with parental expectations, to be the most influential factors for completing Year 12, from nationally representative survey data. Similarly, from a recent cohort of Australian students, overseas born students with a foreign language background were found to be 6% more likely to complete Year 12, and this effect became statistically insignificant after controlling for surveyed educational aspirations (Homel et al., 2012).

3.3. The Data Sample

3.3.1. Selective High Schools and the Admissions Process

This chapter follows a subset of the Year 9 cohort of students from an Australian state in 2008, including students at a limited number of selective high schools. The selective schools were attended by less than 1% of the student population. There were initially 67,867 students in Year 9 in 2008.

The entrance exam for admission into the selective high schools is held annually and consists of a series of tests. Students sit the exams in the year preceding entry, in Year 8, as the schools cater to students from Years 9 to Year 12. In 2008, results from the entrance exam determined the vast majority of places at selective high schools, while the balance of students were accepted on a discretionary basis, through a subsequent process conducted in addition to the entrance exam.

Students in the discretionary intake may have missed the cut-off score or may have been excluded from the main intake due to the existence of an enrolment cap that is designed to reduce the impact of selective schools on other schools; specifically, the student intake to selective high schools from any one school is restricted to a small percentage of students.

3.3.2. Sample Definition

Given that admission is based on academic performance, the most comparable students to those at selective high schools are students with similar prior achievement. To that end I define the pool of students with prior achievement above the minimum level at selective high schools as the preliminary sample; I obtain a restricted sample of students by excluding students with a Year 9 NAPLAN[72] score below the minimum at a selective high school. The NAPLAN score is calculated by taking the average of standardized normal scores in Numeracy and Reading[73].

From this preliminary sample I then seek to exclude students admitted to selective high schools on a discretionary basis due to concerns they have been selected for favourable characteristics that are not reflected in the data, such as motivation and discipline. Including these positively selected students is problematic for estimating the overall effect from

[72] National Assessment Program – Literacy and Numeracy, see Data Description (Section 3.4.1).

[73] Test scores from the other subject domains (Spelling, Grammar and Writing) were not available, but from previous work (Houng and Justman, 2013), the inclusion of other subject domains is unlikely to have had much influence on the analyses.

selective school attendance due to the anticipated positive selection bias from the unobservables.

Note that due to data limitations the Year 9 NAPLAN tests are taken at a point in time when students have already attended the selective schools for several months, meaning that the estimates may be biased downwards if attending the selective schools, even for a limited time, increased their students' short-term achievement.

In the absence of entrance exam data which identifies students in the discretionary intake, I also use the NAPLAN score to exclude discretionary intake students from the data sample. To exclude discretionary intake students from the data sample with confidence, I drop the bottom 15ppt of students, rather than dropping only the bottom 5ppt of students, which was the maximum percentage of students who could have been admitted on a discretionary basis. Therefore, with a 10ppt margin of error I approximate the admissions cut-off with the 15th percentile of NAPLAN score within each selective school.

Table 3.1 presents the minimum and the 15th percentile of the derived prior achievement score of selective students by sex. The minimum and 15th percentile scores are higher for selective girls, possibly due to having fewer student places: 222 as opposed to 334 for selective boys. Although it appears that it is more competitive for girls to attend the selective schools compared to boys, there are more boys at the top end of the distribution of NAPLAN. From Table 3.2 (later), 9.2% of the non-selective boys attained NAPLAN scores above the 15th percentile score at the selective schools, as compared with 7.7% for girls.

Table 3.1: Cut-off Scores approximated with NAPLAN scores by Sex^

	Minimum	15th Percentile
Girls	0.22	1.36
Boys	0.05	1.30

^ defined as the average of standardized Numeracy and standardized Reading scores (mean 0 and standard deviation of 1)

For girls and boys respectively, Figure 3.1 presents histograms of NAPLAN scores by selective school attendance and non-attendance. It can be seen that attendees of selective high schools are drawn from the high end of the distribution of scores. Red and green vertical lines for each of boys and girls indicate the minimum scores and the 15th percentile cut-off scores respectively (from Table 3.1), which are used to limit the sample of both selective and non-selective students for the subsequent analyses. I.e. in the right panel for students at selective schools, it is assumed that students admitted on a discretionary basis are the students between the two vertical lines.

Figure 3.1: Histograms of Scores for Non-Selective and Selective Boys and Girls

Table 3.2: Clearly-In Data Sample as a Percentage of All Students^

	School Type	Clearly-In / Above Cut-off (%)	Below Cut-off (%)	Below Minimum Score (%)	Total
Boys	Non-Selective	9.2%	38.8%	52.0%	29,894
	Selective	84.7%	15.3%	0.0%	334
	Total	10.1%	38.5%	51.4%	30,228
Girls	Non-Selective	6.1%	17.3%	76.6%	29,828
	Selective	85.1%	14.9%	0.0%	222
	Total	6.7%	17.3%	76.0%	30,050
All	Non-Selective	7.7%	28.1%	64.3%	59,722
	Selective	84.9%	15.1%	0.0%	556
	Total	8.4%	27.9%	63.7%	60,278

^ These are all students in Year 9 aged 14 or 15 and not missing information on language and socioeconomic background, with values for at least one of Numeracy or Reading from NAPLAN.

Frequencies and row percentages of selective and non-selective students falling into the (15th percentile) “clearly-in” and “below cut-off” categories are presented in Table 3.2. The total number represents students aged 14 or 15, not missing information on background characteristics and with at least one of Numeracy or Reading score from NAPLAN. At non-selective schools there were 4,574 clearly-in students above the cut-off and 16,763 students below the cut-off but above the minimum NAPLAN score at selective schools. At selective schools there were 472 students above and only 84 students below the cut-off.

In summary, the main sample is chosen to exclude discretionary-intake students, whose inclusion would be likely to bias the results from the analysis upwards, from the pool of students with high prior achievement.

3.4. Data Description

3.4.1. Sample Definition

The data consists of nationally standardized tests from 2008 and Year 12 results from 2011. Introduced in 2008, "National Assessment Program – Literacy and Numeracy" (NAPLAN) testing is conducted in Years 3, 5, 7 and 9 and covers 5 subject domains. The NAPLAN data contains information relating to students' parental occupation and education and foreign language background, indicated by Language Background Other Than English (LBOTE). From the parental education and occupation information, I derive an index of socioeconomic background (SES) using principal components analysis with both Year 7 and Year 9 students from 2008 (see next section).

The Year 12 results are derived from the state's high school qualification to produce a nationally comparable rank, which this paper uses for the outcome variables. Specifically, Australian Tertiary Admission Ranks (ATAR) values are percentile ranks derived from a combination of the results from subjects completed by students in Year 12. Though the aggregated subject results and the ATARs are equivalent, the ATARs are widely reported and used for admission to university.

I derive from the ATARs three binary Year 12 achievement outcomes: ATAR ≥ 95, ATAR ≥ 90, and ATAR ≥ 85; effectively scores within the top 5%, 10%, and 15% of results, respectively. These derived Year 12 outcomes provide a more meaningful measure of success for the group of students of interest, those with high prior achievement, compared to assessing ATAR directly. For instance, a result of ATAR ≥ 95 would allow a student to enter the majority of courses and universities, although there is still a wide range of courses available at ATAR ≥ 85.

3.4.2. The Derived Socioeconomic Index

Table 3.3 presents the parental education and occupation percentages for the clearly-in sample along with the percentage point differentials between selective and non-selective students. The education and occupation categories are shown along the rows and the percentages and differentials, disaggregated by sex, can be read across the columns.

The SES index was derived to suit the propensity score matching approach and also for ease of interpretation[74]. Reducing the number of dimensions simplifies the analysis as there are eight different education categories and 6 different occupation categories, with one each for each student's mother and father. In relation to the propensity score matching, it is clear that there are differences in parental education and occupation between selective and non-selective students that would violate the balancing hypothesis; the balancing hypothesis requires for selection into treatment to be independent of covariates (see Propensity Score Matching, section 3.7.2, and Balancing Hypothesis Tests, section 3.7.2.2 in this chapter).

The main difference in parental education and occupation between selective and non-selective students was that much fewer of the parents of selective students, ranging between 10.4 ppt and 17.2 ppt, reported themselves to be in the "not stated/unknown" categories for both mothers and fathers. The parents of selective students also appeared to have higher levels of education than the parents of non-selective students, as indicated by the category "bachelor degree or above". The difference was particularly high for the mothers of the boys with a percentage differential of 21.5 ppt.

From the differentials in parental occupation for boys, 7.3 ppt and 7.6 ppt more of mothers of the selective boys were in the "other business manager" and "senior manager" categories, respectively. For the selective girls, 16.9 ppt more of their mothers were in the "no work in the last 12 months" category, noting that fewer mothers were "other business manager" or "senior manager" when compared with selective boys. For fathers' occupation, 11.4 ppt more of the fathers of the selective boys were in the "no work in the last 12 months" category, whereas fathers of selective girls were more likely to be in the "machine operator" or "senior manager" categories.

[74] Calculated using the polychoricpca command in STATA. See Kolenikov, S. and Angeles, G. (2004).

Table 3.3: Parental Education and Occupation (%) and Selective less Non-Selective Differentials

	Education or Occupation	Total (%)	Difference (S – NS) (ppt) Girls	Boys	Total
Mothers' Education	1) Not Stated/Unknown	13.3	-14.6	-10.4	-12.1
	2) Less than or Equal to YR9	2.2	5.5	0.8	2.7
	3) YR10 or Equivalent	2.6	1.6	-1.6	-0.4
	4) YR11 or Equivalent	5.0	-4.6	-3.0	-3.6
	5) YR12 or Equivalent	10.7	8.4	-7.7	-1.3
	6) Certificate I to IV*	10.7	-1.5	-5.0	-3.7
	7) Diploma/Adv Diploma	13.2	0.5	5.5	3.5
	8) Bachelor Degree or Above	42.3	4.7	21.5	14.8
Fathers' Education	1) Not Stated/Unknown	18.9	-14.6	-11.5	-12.7
	2) Less than or Equal to YR9	1.9	3.4	0.2	1.5
	3) YR10 or Equivalent	2.0	0.6	-0.6	-0.1
	4) YR11 or Equivalent	3.1	-2.2	-0.4	-1.1
	5) YR12 or Equivalent	9.3	3.4	0.5	1.6
	6) Certificate I to IV*	13.2	-1.9	-2.7	-2.4
	7) Diploma/Adv Diploma	10.1	3.6	3.2	3.3
	8) Bachelor Degree or Above	41.5	7.6	11.3	9.8
Mothers' Occupation	1) Not Stated/Unknown	16.6	-14.8	-16.8	-16.0
	2) No work 12 mths	14.1	16.9	0.8	7.3
	3) Mach op./ Hospitality	9.3	4.5	3.7	4.0
	4) Tradesmen/ Sales	13.0	-1.0	-2.6	-2.0
	5) Other bus-mgr/associate	19.4	-5.4	7.3	2.2
	6) Senior manager/ etc	27.5	-0.2	7.6	4.5
Fathers' Occupation	1) Not Stated/Unknown	22.2	-15.0	-17.2	-16.3
	2) No work 12 mths	5.4	-1.5	11.4	6.2
	3) Mach op./ Hospitality	8.2	9.2	6.4	7.5
	4) Tradesmen/ Sales/	11.4	0.4	-1.8	-0.9
	5) Other bus-mgr/associate	21.8	1.5	2.2	1.9
	6) Senior manager/ etc	31.0	5.3	-0.9	1.6

Difference (S-NS): difference between selective and non-selective * including trade certificate

Note: occupation classifications: 1 = Not stated or unknown, 2= Not in paid work in last 12 months, 3 = Machine operators, hospitality staff, assistants, labourers and related workers, 4 = Tradespeople, clerks and skilled office, sales and service staff, 5 = Other business managers, arts/media/sportspersons and associate professionals, and 6 = Senior management in large business organisation, government administration and defence, and qualified professionals.

3.5. Descriptive Statistics

3.5.1. Demographic Information

Next, with the subsamples defined earlier, I compare the characteristics of non-selective ("NS") and selective ("S") students in Table 3.4 for girls and for boys. Along the rows are language background and socioeconomic background in quartiles, and also prior achievement from NAPLAN. For students above the minimum scores at the selective schools, the sample is grouped by above and below the admission cut-off.

Table 3.4: Characteristics of Non-Selective ("NS") and Selective ("S") Students by Sex

	Girls				Boys			
	Clearly-In		Below Cut-off^		Clearly-In		Below Cut-off^	
	NS	S	NS	S	NS	S	NS	S
N	1,811	189	5,167	33	2,763	283	11,596	51
LBOTE (%)	26.4	78.7	22.9	70.6	27.9	74.4	23.7	62.3
SES Category (%)								
Quartile 1	6.5	13.3	11	5.9	6.2	7.1	15.4	3.8
Quartile 2	10.9	13.3	17.3	14.7	12.3	10.0	22.1	20.8
Quartile 3	27.7	15.4	29.6	14.7	22.8	15.3	26.8	17.0
Quartile 4	54.9	58.0	42.1	64.7	58.7	67.6	35.7	58.5
NAPLAN Results								
Std. Numeracy								
Mean	1.7	2.5	0.8	1.3	1.9	2.5	0.6	1.3
SD	0.7	0.8	0.5	0.5	0.7	0.8	0.5	0.5
%Missing	2.8	0.0	1.7	0.0	1.3	1.1	1.7	0.0
Std. Reading								
Mean	1.8	1.7	1.0	0.9	1.7	1.7	0.5	0.7
SD	0.6	0.7	0.5	0.5	0.6	0.6	0.5	0.5
%Missing	1.2	0.0	1.2	0.0	1.1	0.4	1.6	2.0
Score								
Mean	1.8	2.1	0.9	1.1	1.8	2.1	0.6	1.0
SD	0.4	0.5	0.2	0.2	0.4	0.5	0.3	0.3

^ above minimum score among students at the limited number of selective high schools by sex.

Notes: Score is the average of Std. Numeracy and Std. Reading (mean 0 and SD 1), or one of Std. Numeracy or Std. Reading where the other test score has a missing value.

LBOTE: Language Background Other Than English.

In relation to language background, selective school students are overwhelmingly from a non-English speaking background; 77.5% of girls and 72.5% of boys at selective schools were LBOTE while less than a quarter of their non-selective school counterparts were LBOTE. Despite this, selective school students were arguably from a higher socioeconomic background, with greater representation in the top quartile of SES values from the principal components analysis[75]. For example, for clearly-in boys, 67.6% of students at selective schools were in the top quartile, compared with 58.7% at non-selective schools. Although Aboriginal and Torres Strait Islander status is reported, there were only 34 students below the cut-off and none in the clearly-in subsample (not shown).

Comparing between girls and boys at selective schools, while prior achievement for selective girls and boys were the same at 2.1, on average boys have higher values of standardized SES than girls by 0.33 standard deviations (at 0.96 for boys compared to 0.73 for girls, from Table 5.3, later).

For the "below cut-off" subsamples, students at selective schools received standardized NAPLAN scores higher by more than one standard deviation. For all subsamples, the mean, standard deviation and percentage missing for standardized Reading were very similar between treatment and control groups. For clearly-in students, the prior achievement was 0.3 higher for selective school students than non-selective school students, for both boys and girls.

3.5.2. Achievement Outcomes and Missing Data

Table 3.5 presents mean ATARs and aggregated year 12 scores and the percentage of missing values for clearly-in students at selective and non-selective schools. There are large differences in the raw outcomes between the treatment and control groups, for both actual and derived outcomes. For example, for clearly-in girls without missing ATARs (first column of first panel), the mean ATAR value for selective girls is 95.9 as compared with 90.8 for the non-selective girls. Excluding students with missing ATARs (in the first column of the second panel), about 33.3 ppt more selective girls (75.8% less 42.5%) attained an ATAR greater than 95, while 14.3 ppt more of selective boys did so (49.3% less 35.0%).

[75] I standardize the SES index to have a mean 0 with 1 standard deviation from 134,909 students across Years 3, 5, 7 and 9 in 2008. See section 3.4.2. The Derived Socioeconomic Index.

Table 3.5: Achievement Outcomes and Selective and Non-Selective Students

	Continuous Outcomes				Derived Binary Outcomes*		
	ATAR	% Missing	Aggregate Year 12 Scores#	% Missing	ATAR ≥ 95	ATAR ≥ 90	ATAR ≥ 85
Girls							
Selective	95.9	5.8	181.1	6.3	75.8	92.7	96.1
Non-Selective	90.8	13.3	167.0	13.4	42.5	66.6	80.1
Difference	5.1	-7.4	14.1	-7.1	33.3	26.1	16.0
Boys							
Selective	92.4	2.5	171.0	2.5	49.3	73.6	87.7
Non-Selective	87.8	10.5	161.5	10.8	35.0	56.1	69.9
Difference	4.6	-8.1	9.5	-8.3	14.3	17.4	17.8

This is the aggregated subject scores *Excludes students with missing ATARs

In relation to the missing data (see second and fourth columns of first panel), for the girls 13.3% of non-selective students and 5.8% of selective students had missing ATARs, while for the boys these percentages are 10.5% and 2.5% for non-selective and selective students, respectively. The larger proportion of non-selective students with missing values compared to selective students potentially biases the selective school estimates, depending on the interpretation of these missing values. Receiving an ATAR could be considered an outcome in itself as it likely partially reflects a desire to attend university[76].

Overall, 63% of the cohort of Year 9 students from 2008 completed the state high school qualification in 2011, while alternative outcomes included: completing vocational qualifications, completing the qualification in the following year, leaving high school early, moving interstate[77], or completing other qualifications, such as the International Baccalaureate (IB). Although it is likely that the reasons for the missing values are different for students in the data sample than those in the general student population, the relevant comparison is how students with missing outcome values and students without missing outcome values (herein "missing" and "non-missing" students) differ among each of selective and non-selective students.

Table 3.6 and Table 3.7 show for girls and for boys how missing and non-missing students differ in terms of prior achievement (score) and socioeconomic background (SES), by selective

[76] The interpretation of achievement measures is discussed further in Chapter 4, section 4.4.3.1.
[77] There are only a relatively small number of students missing from the population that may have relocated interstate or overseas. Numbering around 1,000, this is less than 2% of students, from author's calculations.

and non-selective attendance; standard deviations are included in parentheses below the means. Among non-selective girls (first row), students with missing ATARs had higher standardized NAPLAN scores by 0.06, and also had higher values in SES by 0.26 ($p > 0.05$ and $p > 0.001$ respectively, see columns labelled “Diff”). For selective girls, students with missing ATARs had weaker prior achievement by -0.13 but had a higher but not statistically different mean standardized SES, with a difference of 0.54 standard deviations. For the boys, differences between missing and non-missing students were minimal, though selective students’ missing ATARs had lower SES values, with a difference of -0.30.

Table 3.6: Score and SES by Missing ATAR for Girls by Selective Attendance

Girls	Score Non-Missing	Missing	Diff	SES Non-Missing	Missing	Diff	N Non-Missing	Missing
NS	1.78	1.84	-0.06*	0.76	1.02	-0.26***	1,571	240
	(0.38)	(0.42)	(0.03)	(0.97)	(0.90)	(0.06)		
S	2.10	1.97	0.13	0.69	1.23	-0.54	178	11
	(0.51)	(0.50)	(0.16)	(1.19)	(0.90)	(0.28)		
Diff			-0.18			0.27		
			(0.13)			(0.31)		

NS: Non-Selective, S: Selective, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 3.7: Score and SES by Missing ATAR for Boys by Selective Attendance

Boys	Score Non-Missing	Missing	Diff	SES Non-Missing	Missing	Diff	N Non-Missing	Missing
NS	1.77	1.78	-0.01	0.82	0.79	0.03	2,472	291
	(0.41)	(0.47)	(0.03)	(0.96)	(0.99)	(0.06)		
S	2.07	2.01	0.06	0.97	0.66	0.30	276	7
	(0.50)	(0.67)	(0.26)	(0.99)	(1.00)	(0.38)		
Diff			-0.07			-0.27		
			(0.17)			(0.37)		

NS: Non-Selective, S: Selective, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

While non-selective girls with missing ATARs had significantly higher scores and SES values than girls with ATAR values, difference-in-difference tests[78] indicated that these differences were not statistically dissimilar from those from the selective girls, meaning that differences were on the whole common to both selective and non-selective students.

Given how well prior achievement and SES predict Year 12 outcomes (Houng and Justman, 2014), it is probable that the mean ATARs of both selective and non-selective students were only minimally affected, or affected similarly, by the missing data. The similarity in the

[78] Refer to section 3.8.1.3, Difference-in-Difference.

observable characteristics that are predictive of Year 12 outcomes therefore alleviates the concerns of potential bias so long as missing students are excluded, as is the approach for the analyses. Nonetheless, the issue of missing outcome values is revisited in the robustness checks, where the outcomes of students with missing values are imputed rather than excluding these students from the analysis altogether.

3.5.3. Sector Differences

In this section, I present descriptive statistics for the clearly-in and below cut-off samples by sector to get a better sense of the differences between selective and non-selective students. Specifically, I review the composition of students by sector, their prior achievement, socioeconomic and language background, and also the characteristics of their peers.

Table 3.8 presents the frequencies of clearly-in students by sector. For the clearly-in sample, students at selective schools made up a substantial percentage of 9.4%, while 32% of students attended government schools and the remaining 59.1% attended non-government schools. In the "below cut-off sample"[79], selective students were only 0.5% of the total students, and there was a more even split of representation between government and non-government schools at 46.7% and 52.8%, respectively.

Table 3.8: Sample Frequencies by School Sector

	Clearly-In		Below Cut-off		
School Type	Freq.	(%)	Freq.	(%)	No. Schools
Government	1,594	31.6	7,864	46.7	240
Non-Government	2,980	59.1	8,899	52.8	234
Selective	472	9.4	84	0.5	^
Total	5,046	100.0	16,847	100.0	^

* Government excludes Selective Schools. ^ not shown for privacy reasons.

Next, in Table 3.9 I compare the background characteristics of students by sector for the clearly-in sample. In contrast with students at government and non-government schools, the most striking observation from the selective schools is the large proportion of students from a language background other than English; more than 70% of students at selective schools were LBOTE, as compared with around 36% at government and 23% at non-government schools.

[79] The below cut-off sample was defined in section 3.3.2 as students who received NAPLAN scores below the approximated admissions cut-off and above the minimum score at selective schools.

Table 3.9: Characteristics of Clearly-In Students by School Sector

School Type	Girls					Boys				
	Score	SES	LBOTE (%)	N	(%)	Score	SES	LBOTE (%)	N	(%)
Government	1.7	0.51	35.9	619	30.9	1.7	0.53	36.2	975	32.0
Non-Gov.	1.8	0.94	21.4	1,192	59.6	1.8	0.97	22.7	1,788	58.7
Selective	2.1	0.73	78.8	189	9.4	2.1	0.96	73.9	283	9.3
Total	1.8	0.79	31.3	2,000	100.0	1.8	0.83	31.8	3,046	100.0

Although selective schools had a larger proportion of LBOTE students, their students were from a more advantaged socioeconomic background than students at government schools. As measured by parental education and occupation, government schools had a mean standardized SES value of 0.51 and non-government schools had a mean standardized SES value of around 0.94. The selective boys were very similar to those at non-government schools, with a mean of std. SES of 0.96, while the selective girls, with a standardized SES of 0.73, were more advantaged than students at government schools but less advantaged than those at non-government schools.

Finally, in terms of prior achievement from NAPLAN, the boys and girls at the selective schools were the same, at 2.1, which was better than the mean scores of 1.8 at non-government schools and 1.7 for government schools. Note that there was little or no difference between the boys and girls within sectors (government, non-government, or selective).

Turning to differences in peer groups for the clearly-in sample, Table 3.10 presents the characteristics of the peers of students at government, non-government and selective schools. Though non-selective students in the clearly-in sample share very similar characteristics to selective students in terms of socioeconomic background and prior achievement (see Table 3.9), there are significant peer group differences between sectors.

For each sector the mean characteristics of the median school were taken to represent that sector. This was preceded by limiting the pool of schools to those with clearly-in students. Note the actual characteristics are shown for the selective schools. From Table 3.10, girls and boys in the government sector had lower peer achievement, at -0.1, and lower peer SES at -0.2, and similar proportions of LBOTE peers, at around 25%. Students attending non-government schools had higher levels of peer achievement, at 0.3 for both girls and boys, and peer SES, at 0.5 for girls and 0.4 for boys, but were still much lower than that for students attending selective schools.

Table 3.10: Characteristics of Peers of Clearly-In Students by Sector

School Type	Girls				Boys			
	p50 Peer Score	p50 Peer SES	p50 LBOTE (%)	p50 N	p50 Peer Score	p50 Peer SES	p50 LBOTE (%)	p50 N
Government	-0.1	-0.2	26.5	311.7	-0.1	-0.2	24.8	286.8
Non-Government	0.3	0.5	21.1	261.2	0.3	0.4	21.6	261.8
Selective	1.9	0.7	77.5	225	1.9	0.9	72.5	336
Total	0.1	0.1	24.0	286.8	0.1	0.1	23.5	275.7

3.5.4. Selection into Treatment

The earlier descriptive statistics showed that students at selective schools were more likely to be from a language background other than English, have higher prior achievement scores, and were more likely to be from a higher socioeconomic background. Modelling selection into treatment, I estimate a logistic regression on selective school attendance with standardized Numeracy and Reading, language background, and socioeconomic background as explanatory variables. The average marginal effects for these variables are presented in Table 3.11.

Column 1 pools boys and girls together and columns 2 and 3 show the results separately for girls and then boys. For all students, Numeracy has a stronger effect than Reading on selection into treatment; a one standard deviation increase in standardized Numeracy translates to a 6.4 ppt increase in probability of attendance, while the equivalent increase for standardized Reading is 2.3 ppt. Language background other than English has the largest effect size with a 15.1 ppt increase in probability of attendance. Being male or female was unimportant in size and statistical significance, while having higher values in socioeconomic background marginally improves the likelihood of attendance for boys or for both girls and boys combined; a one standard deviation increase translates to less than 1 ppt increase in probability.

Table 3.11: Logistic Regression on Selective School Attendance

	All	Girls	Boys
	(1)	(2)	(3)
Std. Numeracy	0.064***	0.070***	0.060***
	(0.005)	(0.008)	(0.007)
Std. Reading	0.023***	0.023*	0.025**
	(0.006)	(0.010)	(0.008)
LBOTE	0.151***	0.156***	0.147***
	(0.010)	(0.016)	(0.012)
Std. SES	0.009*	0.005	0.013*
	(0.004)	(0.006)	(0.006)
Male	-0.009		
	(0.009)		
Success Rate	0.103	0.105	0.101
Tjur Statistic	0.161	0.190	0.144
N			
Non-Selective	3,936	1,518	2,418
Selective	451	178	273
Total	4,387	1,696	2,691

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Note: tests scores are from Year 9 NAPLAN and the data sample is from a different cohort than from the samples in Chapter 4.

The Tjur statistic measures goodness of fit by a difference-in-means of the predicted probability between the selective and non-selective groups, with higher values indicating better fit. Its appeal is the simplicity of its definition. Unlike conventional measures, such as the R-square for linear regression, Tjur values are unlikely to tend towards 1 as this is only possible when there is perfect prediction; all treated (selective) observations are assigned a probability of 1 and all untreated (non-selective) observations are assigned a probability of 0.

Estimating selection for girls and boys separately provides a Tjur of 0.190 for the girls and a Tjur of 0.144 for the boys. The Tjurs suggests that with only basic information of prior achievement and language and socioeconomic background attendance can be better modeled for the girls.

3.6. Methodology

The advantage of using Propensity Score Matching (PSM) over other largely equivalent methods is its intuitiveness; the modelling of selection into treatment is separate from the weighting of the outcomes. In contrast, nonparametric kernel estimators and saturated models (Woolridge 2002, Ch. 18) flexibly model the full set of covariates but both methods estimate selection and outcomes in one step.

In the first stage of PSM, selection into treatment is modelled to produce predicted probabilities (propensity scores), and in the second stage the control observations are matched with treatment observations using these probabilities. To provide a point of reference for interpreting the magnitude of the estimated effects, I also conduct probit regressions to obtain the expected success rates for each of the binary Year 12 outcomes: ATAR ≥ 95, ATAR ≥ 90 and ATAR ≥ 85. I.e. an estimated effect of 5 percentage points may be less meaningful if the expected success is low.

3.6.1. Selection into Treatment

"…matching amounts to covariate-specific treatment-control comparisons, weighted together to produce a single overall average treatment effect" – (Angrist and Pischke, 2009)

Propensity score matching makes like with like comparisons, where the probability of treatment reflects likeness. Formally, the propensity score is defined by Rosenbaum and Rubin (1983) as the probability of treatment given **x**, the function p(**x**):

$$p(\mathbf{x}) = \Pr(D = 1|\mathbf{x})$$

where D = {0, 1}, and **x** are the pre-treatment characteristics (Becker and Ichino, 2002). Propensity score matching requires the *ignorability of treatment* assumption, such that treatment and outcomes are independent, conditional on **x.**

In the first stage, a probit or logit model is fitted to estimate the propensity score (or predicted probability) from the selection equation:

$$\Pr(D = 1|\mathbf{x}) = cdf\{h(\mathbf{x})\}$$

where *cdf* is the normal or logistic cumulative distribution function, and the specification of $h(\mathbf{x})$ is such that selection is independent of **x** given p(**x**), a requirement known as the balancing hypothesis. The estimation details relating to satisfying the balancing hypothesis as well as the first stage specifications are provided in the results section (3.7.2.2 The Balancing Hypothesis Tests).

3.6.1.1. The Balancing Hypothesis

Under Propensity Score Matching, the specification of the selection model requires the balancing hypothesis to be satisfied: selection is independent of covariates, given the predicted probability of treatment. That is, when observations with a same probability of treatment are compared, the group that obtained treatment cannot be distinguished from the group that didn't.

To satisfy this requirement, I follow Becker and Ichino (2002) and test for various specifications that the characteristics of treatment and control groups are not different from each other within a given number of subgroups; tests are conducted at the 1% level of statistical significance, and subgroups may be split into smaller units until the requirement is satisfied. These subgroups are also used within stratification matching to produce an estimate by averaging the within-subgroup estimates.

3.6.1.2. Intuition

Figure 3.2 displays a scatterplot of the propensity scores (predicted probabilities) from the selection equation and the ATAR outcome for girls and boys, by selective high school attendance. As a continuous variable, ATARs are more suited to visualisation than the derived binary variables but is omitted from the statistical analyses as selective school effects in terms of the underlying continuous variable are arguably less informative. For this group of students with high academic potential, it matters where in the distribution of outcomes the benefits accrue. Namely, the difference between receiving an ATAR of 95 and receiving 90 is greater than the difference between 85 and 80.

The matching process is as follows. For each selective school student, a comparison group of non-selective students is determined by their proximity to that student, along the scale of predicted probability (the x-axis). Subsequent to this, the difference between their outcomes (represented on the y-axis) is taken and then averaged. The methods differ by which non-selective students are chosen for comparison (see the Types of Matching, section 3.6.1.3). For example, under stratification matching the selective and non-selective students are compared within contiguous subgroups of predicted probabilities (propensity score values) that satisfy the balancing hypothesis, and the differences are then averaged.

Figure 3.2: Scatterplot of ATARs and Propensity Scores by Sex and Selective Attendance^

^ note that for clarity around half of randomly chosen non-selective students are shown.

3.6.1.3. Types of Matching

In the second stage, I adopt several approaches to matching control with treatment observations: a) nearest neighbour; b) kernel based matching with a bin width = 0.06; and c) stratification matching. These various approaches weight the difference in outcomes according to predicted probability to produce a treatment effect.

Nearest Neighbour

Nearest neighbour pairs the closest control observation in terms of predicted probability with each treatment observation. Following the notation from Becker and Imbens' (2002) documentation of their STATA program, the control observation C(i) for treatment *i* from the nearest neighbour method is identified by the following:

$$C(i) = \min_j (p_i - p_j)$$

where pi and pj are the propensity scores for treatment *i* and control observations *j*, and min *j* refers to the control observation, or control observations, that minimize(s) the difference in propensity score.

Each control observation is then weighted as:

$$w_{ij} = \frac{1}{N_i^C} \text{ if } j \in C(i)$$

where *i* is the treatment observation and N is the number of control observation(s).

The average treatment effect is then the average difference between treatment observations (the first summation, below) and their associated set of control observations (the second summation):

$$\tau^M = \frac{1}{N^T} \sum_{i \in \mathrm{T}} (Y_i^T - \sum_{j \in \mathrm{C(i)}} w_{ij} Y_j^C)$$

M denotes nearest neighbour, *T* denotes treatment observations, and *N* is the total number of treatment or control observations. As above, *i* refers to the treatment observation and *j* refers to control observations. Note that nearest neighbor matching is performed with replacement throughout the main analyses and the robustness checks.

Kernel Matching

Kernel matching weights control observations according to their distance from each treatment observation. It relies on a kernel function and bandwidth parameter to weight all control observations, where those that are in closer proximity to the treatment observations are given greater weighting. The estimator of the counterfactual outcome is:

$$\frac{\sum_{j \in \mathrm{C}} Y_j^C \, G\left(\frac{p_j - p_i}{h_n}\right)}{\sum_{k \in \mathrm{C}} G\left(\frac{p_k - p_i}{h_n}\right)}$$

For each treatment observation *i,* a kernel weighting has been given to all control observations (j ∈ C) based on the bandwidth $h_n$and the distance of each control observation $(p_i - p_j)$. A normal distribution has been chosen for the kernel function, G, while the denominator is required to weight the overall term to one.

Stratification Matching

Formally, the estimator for stratification matching is (for within subgroup and overall):

$$\tau_q^S = \frac{\sum_{j \in \mathrm{I(q)}} Y_i^T}{N_q^T} - \frac{\sum_{j \in \mathrm{I(q)}} Y_j^C}{N_q^C} \text{ and } \tau^S = \sum_{q=1}^{Q} \tau_q^S \frac{\sum_{i \in \mathrm{I(q)}} D_i}{\sum_{\forall i} D_i}$$

For stratification matching, differences are calculated within subgroups over ranges of predicted probabilities (for example 0% to 20% and 20% to 40% and so on), and then averaged. The process to determine the ranges is described in section 3.6.1.1, The Balancing Hypothesis, requiring that there is no statistical significant difference (p-value of 0.05) in pre-treatment characteristics between treatment and control groups within these ranges. Once the subgroups (q) and ranges have been determined, the stratification matching estimator is then the weighted average of difference between treatment (Y_i^T) and control outcomes (Y_j^C) within each subgroup, then subsequently weighted again over subgroups with a weighting based on the number of treatment observations in each subgroup ($\sum_{i \in \mathrm{I(q)}} D_i$). *I(q)* is the set of units in block *q*.

3.7. Results

3.7.1. Probit Regressions

Probit regressions are estimated to provide a point of reference for interpreting the magnitude of the estimated effects from the matching approach. Table 3.12 presents average marginal effects from probit regressions of each of the categorised Year 12 outcomes, controlling for prior achievement, language and socioeconomic background and metropolitan residency. Unlike the PSM, the regressions do not account for selection, whereby students from a non-English speaking background with higher prior achievement are more likely to attend selective schools. The expected success rates, of between 46.6% and 82.3% for girls and between 36.7% and 72.2% for boys, provide an indication of the high level of academic ability of students in the data sample. As with the PSM analyses students without ATARs are excluded.

Table 3.12: Average Marginal Effects from Probit Regressions of Year 12 Outcomes^

	Girls			Boys		
	ATAR ≥ 95	ATAR ≥ 90	ATAR ≥ 85	ATAR ≥ 95	ATAR ≥ 90	ATAR ≥ 85
Std. NUM	0.207***	0.198***	0.141***	0.235***	0.246***	0.225***
	(0.017)	(0.018)	(0.018)	(0.011)	(0.013)	(0.013)
Std. RDG	0.109***	0.085***	0.075***	0.097***	0.103***	0.096***
	(0.019)	(0.020)	(0.018)	(0.014)	(0.015)	(0.015)
Std. SES	0.081***	0.081***	0.060***	0.056***	0.067***	0.059***
	(0.011)	(0.010)	(0.009)	(0.009)	(0.009)	(0.008)
LBOTE	0.070*	0.049	0.054*	0.057**	0.054**	0.053**
	(0.028)	(0.026)	(0.023)	(0.019)	(0.021)	(0.019)
Selective	0.142***	0.183***	0.113*	-0.062*	-0.024	0.015
	(0.041)	(0.049)	(0.045)	(0.029)	(0.032)	(0.032)
Metropolitan	0.089**	0.079**	0.044*	0.140***	0.136***	0.122***
	(0.030)	(0.025)	(0.021)	(0.024)	(0.023)	(0.019)
Tjur Statistic	0.166	0.141	0.095	0.189	0.178	0.174
Success Rate	46.6%	70.0%	82.3%	36.7%	58.2%	72.2%
N	Non-Selective	Selective	Total	Non-Selective	Selective	Total
	1,501	178	1,679	2,417	273	2,690

* p < 0.05, ** p < 0.01, *** p < 0.001. ^ Students with missing ATARs and missing geography data are excluded.

All regression estimates for the girls were statistically significant; the selective school estimates suggest that attending a selective school results in a 14.2 ppt increase in likelihood of achieving ATAR ≥ 95 for girls, while the estimate for ATAR ≥ 85 was similarly large and even larger for ATAR ≥ 90. The regression results for boys of selective school attendance were only statistically significant for ATAR ≥ 95, with a decrease in probability of attainment of 6.2 ppt.

Note that the probit regressions do not account for the likelihood that the students attending selective schools were selected for favourable characteristics that are not observed in the available data. It is reasonable to suggest sitting the entrance exam reflects additional motivation. That is, it is probable that the regression estimates were biased upwards, although the results for the boys suggests otherwise.

The Tjur statistics, the difference in predicted probability of success between those attaining the outcome and those unsuccessful, suggest that the goodness-of-fit of the regressions for the boys, at between 0.174 and 0.189, is greater than that for the girls, at between 0.095 and 0.166. For both boys and girls, accuracy in statistical modelling increased with increasing difficulty level in attaining the Year 12 outcome.

3.7.2. Propensity Score Matching

3.7.2.1. The Selection Model

For the basic selection model, I consider boys and girls separately as earlier logistic regressions (section 3.5.4) indicated that socioeconomic background may be less important for girls. For covariates I include standardized Numeracy and standardized Reading as linear terms, language background as an indicator variable, and socioeconomic background as a continuous derived variable of parental education and occupation.

Recall, the premise of the propensity score matching approach is that individuals are observed to be the same based on their likelihood of treatment. Unfortunately, for the girls the basic selection model did not satisfy the balancing hypothesis requirement according to tests comparing the mean characteristics of students in both treatment and control, within subgroups categorised by the likelihood of attendance. Specifically, the mean SES values were different for the treatment and control groups with the lowest likelihood of attendance. See Balancing Hypothesis Tests (section 3.7.2.2) for further explanation.

Note also that differences between selective and non-selective students in raw parental education and occupation (section 3.4.2) mean that the balancing requirement would not be met for selection models with indicators for each education and occupation classification.

Adopting a continuous derived SES variable is a convenient abstraction which simplifies the conceptualisation of the problem, but also draws attention to limitations in the matching approach and the likely bias in the estimates[80].

With a secondary objective for specifying the selection model of maximising the goodness-of-fit, in addition to satisfying the balancing hypothesis requirement, I estimate two other specifications. Given the somewhat surprising lack of SES effect for girls, I posit there to be relationship between SES and prior achievement in its influence on selection and estimate the selection model with interactions for SES and prior achievement. In the other variation, I estimate the selection models with quadratic terms for the NAPLAN scores.

Table 3.13 presents the estimates from both the basic model and the basic model including interactions of SES and prior achievement. The interactions are between quartiles of SES and three equal sized categories of prior achievement (low, medium, high), where prior achievement was defined as the average of standardized Numeracy and standardized Reading.

Table 3.13: Logistic Regressions of Selection into Treatment by Sex

	Girls		Boys	
	(1)	(2)^	(3)^	(4)
Std. NUM	0.0696***	0.0456***	0.0595***	0.0367***
	(0.0082)	(0.0123)	(0.0068)	(0.0099)
Std. RDG	0.0234*	0.0022	0.0245**	0.0034
	(0.0097)	(0.0128)	(0.0080)	(0.0107)
LBOTE	0.1556***	0.1515***	0.1471***	0.1462***
	(0.0157)	(0.0152)	(0.0120)	(0.0118)
Std. SES	0.0050	0.0067	0.0133*	0.0100
	(0.0063)	(0.0142)	(0.0056)	(0.0113)
SES X Score categories	N	Y	N	Y
Tjur	0.190	0.208	0.144	0.150
Success Rate	0.105	0.105	0.101	0.101
N				
Selective	178	178	273	273
Non-Selective	1,518	1,518	2,418	2,418
Total	1,696	1,696	2,691	2,691

* p < 0.05, ** p < 0.01, *** p < 0.001. ^ preferred specifications for propensity score analyses.

[80] Without random assignment in treatment, it is expected that the more granular the information used to model selection and subsequently make comparisons, the less likely the underlying assumption (that individuals are observed to be the same based on their likelihood of treatment) will hold.

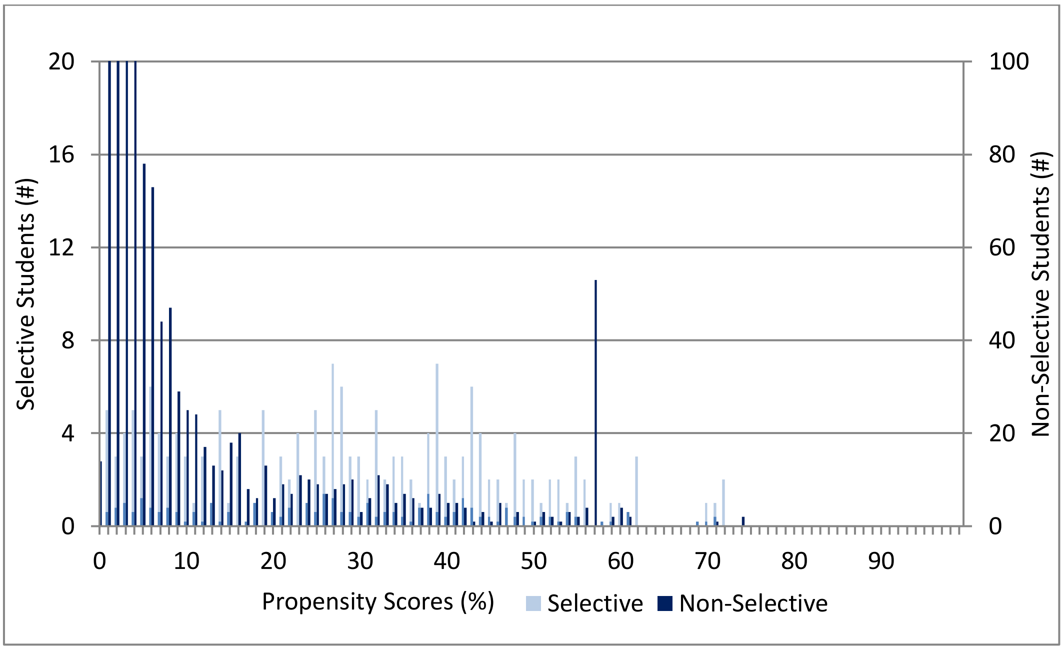
The inclusion of the interactions for the girls improved the Tjur goodness-of-fit from 0.190 to 0.208. It is calculated as the difference in predicted probability of students who attended selective schools and students who did not. For the boys the interactions had less effect, improving from 0.144 to 0.150. Likelihood ratio tests similarly indicated that the inclusion of interactions of SES and prior achievement resulted in statistically significant improvements in fits of the model for girls (p-value of 0.018) but not for boys (p-value of 0.065). More importantly, the selection model with the inclusion of the interactions did satisfy the balancing hypothesis for the girls.

To check for non-linearities in prior achievement, the selection models were also estimated with quadratic terms for the NAPLAN scores but the specifications failed the balancing hypothesis requirement (Appendix, Table 3.27 A). There was little difference in terms of fit for the girls although there was some improvement in fit for the boys.

In summary, for the preferred specifications I adopt the simplest selection models that satisfied the balancing hypothesis requirement. These were the basic selection model for the boys (column 3) and the basic model with interactions for SES and prior achievement for the girls (column 2).

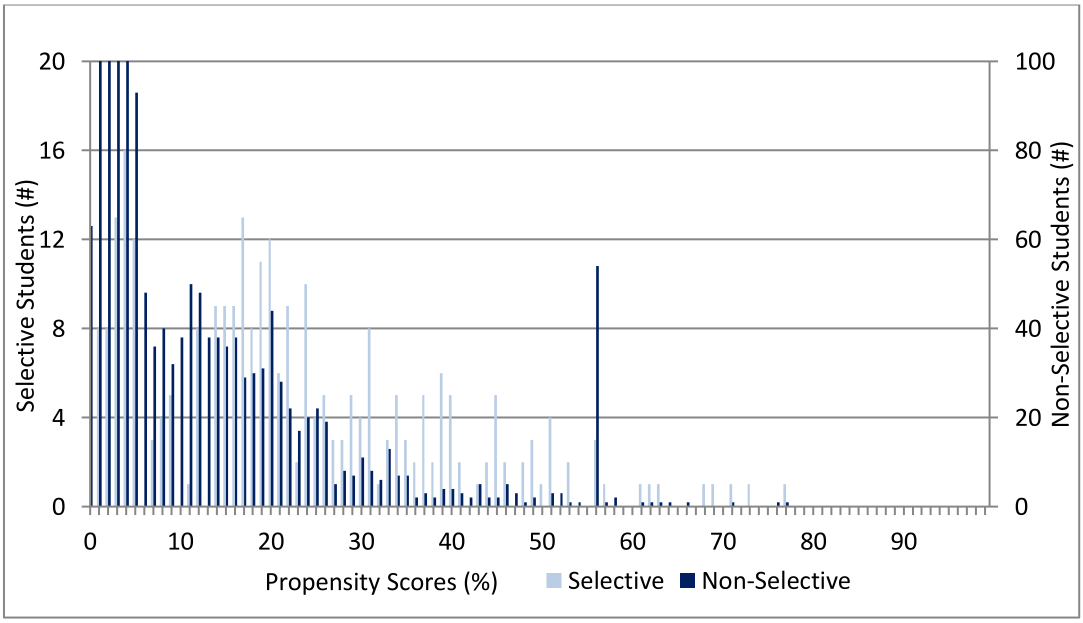
The extent of common support in the treated and untreated samples is presented next. Figures 3.3 and 3.4 show histograms of the propensity scores by selective attendance for girls and boys, respectively. The frequencies of each sample over 1 percentage point intervals in propensity scores are shown on separate vertical axes due to large differences in sample sizes. In both Figures 3.3 and 3.4, the selective sample is indicated by light blue bars measured on the left vertical axis and the non-selective sample is reflected by dark blue bars measured on the right vertical axis.

Figure 3.3: Histogram of Propensity Scores by Selective Attendance, Girls^



^ truncated non-selective counts of 358, 238, 175 and 104 at propensities from 1 to 4.

Figure 3.4: Histogram of Propensity Scores by Selective Attendance, Boys^



^ truncated non-selective counts of 593, 464, 233 and 137 at propensities from 1 to 4.

While the selective sample is more evenly distributed across the propensity scores, between 0 and 71 points for girls in Figure 3.3 and between 0 and 77 points for boys in Figure 3.4. The distribution of non-selective students is centered between 0 and 10 points, reaching counts of almost 600 for boys and 360 for girls in the lower ranges of propensity scores. At higher ranges of propensity scores, above 40 or 50 points, frequency counts for the non-selective sample were small at 5 observations of less, fewer than counts in the selective sample.

3.7.2.2. Balancing Hypothesis Tests

The main assumption of the propensity score matching approach requires that selection is independent of the covariates. This section presents the details and results from the balancing hypothesis tests (Becker and Ichino, 2002) for the preferred selection model specifications.

Firstly, the selective and non-selective students are allocated into subgroups based on their predicted probability of selective school attendance, obtained from the selection models. These subgroups ("blocks") are presented in the rows of Table 3.14 and Table 3.15, for girls and boys, respectively. The range of the predicted probabilities within each subgroup is shown in the second column, followed by the mean predicted probabilities for the treatment (T) and control group (C), as well as the difference (D) between the two and the associated statistical error; the statistical significance is indicated in the difference column (D), while the observation frequencies are presented in the remaining columns.

Table 3.14: Predicted Probabilities from Selection Model of Treatment and Control, Girls

						No. Observations		
Block	Range (%)	T	C	D	Error	NS	S	Total
1	[1.06, 3.56]	2.22	2.10	0.11	0.19	577	14	591
2	[3.58, 7.14]	5.67	5.11	0.56*	0.27	277	16	293
3	[7.21, 14.26]	10.96	9.76	1.20*	0.47	179	20	199
4	[14.3, 21.33]	18.40	17.25	1.15+	0.62	83	13	96
5	[21.58, 28.53]	25.73	25.00	0.73	0.48	63	27	90
6	[28.57, 42.8]	36.17	35.03	1.14	0.75	85	46	131
7	[43.0, 56.35]	48.88	50.33	-1.45	1.11	28	31	59
8	[57.59 71.06]	64.55	60.02	4.53	2.20	9	9	18
Total	[1.06, 71.06]	28.61	8.31	20.29***	0.99	1,301	176	1,477

* p < 0.05, ** p < 0.01, *** p < 0.001, T: Treatment, C: Control, D: Difference

Table 3.15: Predicted Probabilities from Selection Model of Treatment and Control, Boys

						No. Observations		
Block	Range (%)	T	C	D	Error	NS	S	Total
1	[1.67, 3.57]	2.60	2.48	0.12	0.11	1,165	21	1186
2	[3.57, 7.14]	4.70	4.83	-0.14	0.16	484	38	522
3	[7.18, 14.28]	11.44	11.05	0.38	0.38	325	31	356
4	[14.33, 28.45]	20.22	19.59	0.64	0.38	405	113	518
5	[28.61, 42.66]	34.71	34.03	0.68	0.72	90	50	140
6	[42.87, 54.72]	46.55	46.96	-0.41	1.00	27	18	45
7	[57.92, 69.64]	63.44	61.46	1.98	2.08	7	7	14
8	[72.91, 73.71]	73.71	73.31	0.40	.	2	1	3
Total	[1.67, 73.71]	21.38	8.13	13.24***	0.64	2,505	279	2,784

* p < 0.05, ** p < 0.01, *** p < 0.001, T: Treatment, C: Control, D: Difference

The balancing tests, shown in Table 3.16 and Table 3.17, are difference-in-means tests. The mean characteristics (prior achievement, socioeconomic and language background) of the control (C) and treatment (T) groups are presented within subgroups, along with the difference in their means (D). These tests were conducted at the 1% level of statistical significance.

From Table 3.16 for girls, the asterisk in the last column indicates that at lower propensity scores, in the row corresponding to block 2, there was a large difference of 18.25 percent in LBOTE between selective and non-selective students, statistically significant at a p-value of 0.05 but not at a p-value of 0.01. Similarly, there were differences between treatment and control groups of: 0.50 SD in Year 7 NAPLAN (column D in the Score panel) for block 8; and -0.63 in Std. SES for block 7, both of which were statistically significant at p-value of 0.05.

For boys the largest differences between selective and non-selective students were for SES, at lower propensity scores. For blocks 1 and 3 of column D in the Std. SES panel (in Table 3.17), the selective students (treatment) had 0.47 greater values in SES, and 0.46 lower values in SES, than their corresponding non-selective students (control), respectively.

Table 3.16: Tests for satisfying the Balancing Property from the Selection Model, Girls

	Score			Std. SES			LBOTE		
Block	T	C	D	T	C	D	T	C	D
1	1.64	1.66	-0.02 (0.06)	1.38	0.94	0.44 (0.24)	0.00	0.17	-0.17 (1.11)
2	1.95	1.95	0.00 (0.08)	1.02	0.85	0.17 (0.25)	31.25	13.00	18.25* (8.89)
3	1.96	1.89	0.07 (0.11)	1.23	0.94	0.29 (0.24)	50.00	54.19	-4.19 (11.81)
4	1.87	1.73	0.14 (0.13)	0.23	0.68	-0.45 (0.37)	92.31	91.57	0.74 (8.33)
5	1.90	1.93	-0.03 (0.09)	0.61	0.54	0.07 (0.26)	96.30	96.83	-0.53 (4.18)
6	2.13	2.13	0.01 (0.07)	0.47	0.49	-0.01 (0.24)	100.00	100.00	0.00 (0.00)
7	2.36	2.42	-0.06 (0.11)	0.49	1.11	-0.63*(0.29)	100.00	100.00	0.00 (0.00)
8	2.98	2.47	0.50* (0.21)	0.29	0.19	0.09 (0.67)	100.00	100.00	0.00 (0.00)

* p < 0.05, ** p < 0.01, *** p < 0.001, T: Treatment, C: Control, D: Difference

Table 3.17: Tests for satisfying the Balancing Property from the Selection Model, Boys

	Score			Std. SES			LBOTE		
Block	T	C	D	T	C	D	T	C	D
1	1.60	1.59	0.01 (0.05)	1.22	0.75	0.47* (0.19)	0.00	0.00	0.00 (0.00)
2	2.06	2.08	-0.02 (0.05)	1.49	1.21	0.28 (0.14)	0.00	0.21	-0.21 (0.74)
3	1.80	1.81	-0.00 (0.10)	0.24	0.69	-0.46* (0.21)	70.97	70.15	0.81 (8.62)
4	1.88	1.80	0.07 (0.04)	0.77	0.93	-0.16 (0.10)	97.35	96.54	0.80 (1.90)
5	2.36	2.40	-0.04 (0.05)	1.13	1.12	0.01 (0.16)	100.00	98.89	1.11 (1.49)
6	2.83	2.86	-0.03 (0.08)	1.49	1.34	0.15 (0.27)	100.00	100.00	0.00 (0.00)
7	3.35	3.31	0.04 (0.14)	0.95	0.51	0.45 (0.52)	100.00	100.00	0.00 (0.00)
8	3.86	3.86	0.00 (.)	1.69	1.54	0.15 (.)	100.00	100.00	0.00 (.)

* p < 0.05, ** p < 0.01, *** p < 0.001, T: Treatment, C: Control, D: Difference

3.7.2.3. Matching Results

This section presents the main results from the propensity score matching for girls and for boys in Table 3.18 and Table 3.19, respectively. Estimates from each of nearest neighbour, kernel and stratification matching methods are presented along the rows. Along the columns are the Year 12 outcomes and also frequencies of non-selective ("NS"), selective ("S") and total students corresponding to each method. The lower panel shows expected success rates for context and estimates of the selective school effect from the earlier probit regressions for comparison.

Table 3.18: Regression and Propensity Score Analysis, Girls

	Outcome *ATAR ≥ 95*	*ATAR ≥ 90*	*ATAR ≥ 85*	Frequencies NS	S	Total
Nearest Neighbour Matching	0.152** (.058)	0.096* (.041)	0.017 (.029)	151	178	329
Kernel Matching (bin width= 0.06)	0.179*** (.037) [b]	0.140*** (.029) [b]	0.071*** (.017) [b]	1,301	178	1,479
Stratification Matching	0.158*** (.04) [b]	0.126*** (.026) [b]	0.062*** (.019) [b]	1,303	178	1,479
Success Rate	46.6%	70.0%	82.3%			
Probit Regression	0.142*** (0.041)	0.183*** (0.049)	0.113* (0.045)	1,501	178	1,679

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

[b] indicates bootstrapped standard errors with 50 replications.

Table 3.18 shows increases in the success rates from selective school attendance were large for girls; almost all estimates were statistically significant, the exception being the estimate for ATAR ≥ 85 from nearest neighbourhood matching. In addition, the estimates from the probit regressions were supportive of the PSM estimates, being similar in effect size.

Estimates ranged from 15.2 ppt to 17.9 ppt for the outcome ATAR ≥ 95, an ATAR which would allow a student to enroll in, for example, a Bachelor of Commerce at the University of New South Wales, a course in high demand at one of the top ranked universities in Australia[81]; note that as the minimum ATAR requirements for certain courses are determined by demand, high ATAR entry requirements tend to reflect a course's popularity. For ATAR ≥ 90, estimates for

[81] The ATAR cut-off for entry to the Bachelor of Commerce at the University of New South Wales was 96.7 in 2011, the year in which the students completed high school.

the selective school effect ranged between 9.6 ppt and 14.0 ppt, while for ATAR ≥ 85 significant estimates ranged between 6.2 ppt and 7.1 ppt. The results suggested that the benefits were larger at the top end of the outcome distribution with effect sizes decreasing with lower achievement outcomes.

Under kernel matching, comparisons are made with all observations but comparisons are distance weighted using the predicted probabilities. For stratification matching, there are 8 subgroups for girls (section 3.7.2.2, Table 3.14) and 8 subgroups for boys (section 3.7.2.2, Table 3.15).

Of the different propensity score matching methods, generally kernel matching produced the largest effects, followed by stratification matching, and nearest neighbour matching, which produced the smallest estimates. The kernel matching method is potentially more reliable as it weights observations that are closer in proximity, in terms of likelihood of treatment, and it is based on more observations than is used in the nearest neighbour matching.

The results for the boys (Table 3.19) were much weaker than those for the girls. From both nearest neighbour and stratification matching, the results for selective and non-selective students were statistically indistinguishable, while for kernel matching the only statistically significant estimate was for ATAR ≥ 85, with an effect of 5.3 ppt. In contrast to the positive effects or no effects from the PSM results, the probit regression produced a negative effect of -6.2 ppt for the ATAR ≥ 95.

Table 3.19: Regression and Propensity Score Analysis, Boys

	Outcome			Frequencies		
	ATAR ≥ 95	*ATAR ≥ 90*	*ATAR ≥ 85*	NS	S	Total
Nearest Neighbour Matching	-0.065 (.047)	-0.026 (.040)	0.018 (.032)	261	273	534
Kernel Matching (bin width= 0.06)	-0.020 (.039) [b]	0.035 (.027) [b]	0.053** (.017) [b]	2,239	273	2,512
Stratification Matching	-0.058 (.033) [b]	-0.003 (.029) [b]	0.019 (.021) [b]	2,239	273	2,512
Success Rate	36.7%	58.2%	72.2%			
Probit Regression	-0.062* (0.029)	-0.024 (0.032)	0.015 (0.032)	2,417	273	2,690

* p < 0.05, ** p < 0.01, *** p < 0.001. Standard errors indicated in parentheses.
[b] indicates bootstrapped standard errors with 50 replications.

In summary, the results from propensity score matching showed that the girls attending selective schools increased their likelihood of attaining each of the Year 12 outcomes. The effect sizes were largest for ATAR ≥ 95, with estimates suggesting an increase in likelihood of over 15 ppt. For context, the expected success rate for ATAR ≥ 95 from probit regressions was 46.6%.

These results are positive for the girls in that there is evidence of gains across the three Year 12 outcomes, and gains are largest for the highest outcome of ATAR ≥ 95, meaning more selective students could conceivably attend courses and universities with greater demand. There were weak results for the boys, with limited evidence for selective boys improving their ATAR ≥ 85, which would still allow students to enter a wide range of courses; kernel matching produced statistically significant estimates of 5.3 ppt, off a base expected success rate of 72.2% for ATAR ≥ 85.

3.8. Robustness Checks

The robustness checks have the dual purposes of firstly addressing potential concerns that arose in the discussion of the data sample and the main analyses, and secondly, to provide complementary analyses. The first concern of two from the data sample was that the control group did not sufficiently resemble the treatment group due to the unrestricted manner of its construction to increase the sample sizes; the control group was drawn from the full state student population and only prior achievement was used to exclude students. For example, the control group includes students who reside in rural areas, which is debatable because almost all selective school students reside in metropolitan areas and overall student achievement is lower in rural areas.

The second concern related to the potential bias in the results from the missing ATARs (discussed in section 3.5.2 - Achievement Outcomes and Missing Data), whereby more non-selective students have missing ATARs than selective students. Descriptive statistics showed that in terms of prior achievement and socioeconomic background there was positive selection of students into missing values for the non-selective girls, however, this positive selection was not statistically different from that of selective girls. On this basis, that there was similar selection into missing data for both selective and non-selective students, students with missing ATARs were excluded from the main analyses.

For robustness checks, which are described in greater detail in the next section, I undertake sensitivity analyses that consist of varying the sample of non-selective students to identify better control groups, and secondly, I re-run the sensitivity analyses with imputed missing ATAR values instead of excluding them in the analyses. Finally, to complement the results from the propensity score matching, I undertake difference-in-difference analyses, which is a more direct approach, to produce local estimates within SES quartiles.

3.8.1. Methods

3.8.1.1. Varying Samples

The sample chosen for the main analyses balanced the concern of retaining a sufficiently large sample against obtaining suitable students for the control group, students who could have received offers to attend selective high schools strictly on the basis of entrance exam results. From the full Year 9 student population in the state, the sample was consequently limited to those students whose NAPLAN scores were in the top 85% of NAPLAN scores of each of the selective high schools[82].

To gauge the robustness of the estimates from the propensity score matching, I conduct sensitivity analyses by repeating the PSM analyses with the following samples:

(1) *Metro Sample* – Only students residing in metropolitan areas are included in the comparisons. For our case study this is a more realistic comparison as there are likely to be very few non-metropolitan students at selective high schools due to distance.

(2) *Non-Government Sample* – I reduce the clearly-in sample to students attending only non-government schools. Earlier comparisons of sector differences in clearly-in students (section 3.5.3) showed that selective students were more alike students at non-government schools than those at government schools in terms of prior achievement and socioeconomic background (but not language background).

(3) *LBOTE Sample* – The other significant difference of selective schools is the large representation of LBOTE students. For this sample, I limit the sample to students who attend schools for which 50% of more students are LBOTE.

[82] Recall that up to 5% of students at selective high schools were admitted on a discretionary basis. Also the NAPLAN was used as an approximation to the entrance exam results, which we do not have.

3.8.1.2. Imputing Missing ATARs

The main analyses excluded students without ATARs, noting the different rates of missing ATARs between selective and non-selective students (see section 3.5.2, Achievement Outcomes and Missing Data). Excluding students without ATARs is arguably adequate for reducing their impact on the results, conditional on students without ATARs resembling students with ATARs both selective and non-selective students. If, however, the students at non-selective schools with the highest academic ability have missing ATARs, along with the early school leavers, but the selective students with missing ATARs had weaker academic ability, then the estimates would be overstated.

Though descriptive statistics indicated that the potential bias would be limited, I test this assertion by imputing aggregated Year 12 scores from the full student population and calculate their ATAR equivalents. A reasonable goodness-of-fit (R-squared) of 0.45 was obtained in previous work (Houng and Justman, 2014) with administrative data, when that sample was limited to students with ATARs above 50. With this same sample, I adopt the regression model from Houng and Justman, with prior achievement from NAPLAN, sex, language and indigenous background and parental education and occupation as explanatory variables.

In summary, for this sensitivity analyses I include clearly-in students with missing ATARs to the main sample, assigning ATAR equivalents from the predicted aggregate Year 12 scores produced by the regression on the broader student sample. Propensity score matching is then applied on both the main samples and the metropolitan samples.

3.8.1.3. Difference-in-Difference

The appeal of the difference-in-difference method is that it is straightforward. The first difference compares pre- and post- outcomes, which are academic achievement measured in Year 9 and in Year 12. The difference-in-difference estimate is then a subtraction of the first difference of the control group from the first difference of the treatment group.

In mathematical notation, the difference-in-difference is (section 5.2 of Angrist and Pischke):

$$(S_{t+1}^{T} - S_{t}^{T}) - (S_{t+1}^{C} - S_{t}^{C})$$

Academic achievement (score) is denoted S_{t+1} in time period 2 and S_t in time period 1. The treatment group is denoted with superscript *T*, while the control group of students is denoted with superscript *C*.

For the academic achievement in Year 12, I again adopt the Year 12 outcomes of ATAR ≥ 95, ATAR ≥ 90, and ATAR ≥ 85. For Year 9, I define analogues to the Year 12 outcomes of: top 5% NAPLAN, top 10% NAPLAN, and top 15% NAPLAN.

Taking into account the earlier descriptive statistics of selection into treatment, I identify similar treatment and control groups in terms of socioeconomic status but not for language background, for which sample sizes are questionable for some subgroups. The difference-in-difference approach also has the benefit of producing local estimates by quartiles of SES, which are arguably easier to interpret than local estimates which are based on predicted probabilities, such as those produced from stratification matching.

3.8.2. Descriptive Statistics

Table 3.20 presents the number of students and schools in each of the robustness checks samples, broken down by selective attendance and sex. For comparison, the frequencies from the main sample, when students with missing ATARs were excluded are also shown. For boys and girls, the number of students and schools decrease moving from the main sample to each successive sample: metropolitan, non-government and LBOTE. As before, the samples for the boys were larger in size than for the girls reflecting the larger number of boys at the top end of the achievement distribution (refer to Sample Definition, section 3.3.2).

Table 3.20: Student Frequencies by Robustness Checks Samples^

Samples	Category	Girls	Schools	Boys	Schools
Main Sample	Non-Selective	1,303		2,239	
	Selective	178		273	
		1,479	360	2,512	381
Metropolitan	Non-Selective	1,044		1,809	
	Selective	178		273	
		1,222	228	2,082	228
Non-Government	Non-Selective	839		1495	
	Selective	178		273	
		1,017	164	1,768	162
LBOTE	Non-Selective	232		368	
	Selective	178		273	
		410	58	641	61

^ The samples exclude students with missing outcome values (i.e. they are not imputed).

3.8.3. Results

3.8.3.1. Varying the Samples

Table 3.21 and Table 3.22 present the estimates from the sensitivity analyses for girls and for boys, where the samples are labelled along the columns, including estimates from the main analyses for comparison. Along the rows are estimates from the probit regression and matching methods for each outcome.

Table 3.21: Regression and Propensity Score Analyses (varying sample), Girls^

Outcome	Method	Sample			
		Main Sample	Metro	Non-Gov.	LBOTE
ATAR ≥ 95	Regression	0.142***	0.141***	0.070	0.248***
	Nearest Neighbour	0.152**	0.132*	0.045	0.281***
	Kernel (bin width= 0.06)	0.179***	0.169***	0.104*	0.272***
	Stratification	0.158***	0.171***	0.091*	0.259***
ATAR ≥ 90	Regression	0.183***	0.168***	0.088	0.189***
	Nearest Neighbour	0.096*	0.110**	0.073*	0.169**
	Kernel (bin width= 0.06)	0.140***	0.129***	0.081**	0.147***
	Stratification	0.126***	0.125***	0.088*	0.138***
ATAR ≥ 85	Regression	0.113*	0.099*	0.027	0.123**
	Nearest Neighbour	0.017	0.065*	0.017	0.140**
	Kernel (bin width= 0.06)	0.071***	0.066**	0.024	0.090**
	Stratification	0.062***	0.061***	0.020	0.082***

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

^ The samples exclude students with missing outcome values (i.e. they are not imputed).

For the girls, the results from applying the propensity score matching analysis to variations of the samples (Table 3.21) were supportive of the main results. Estimates from the non-government sample were positive and similar in effect size but were smaller than that of the main results, while the results from the metropolitan sample were similar to that from the main results. Note though the absence of statistically significant estimates for ATAR ≥ 85 for the non-government sample.

The results from the LBOTE sample, where non-selective students were restricted to those that attended schools where LBOTE students made up at least 50% of the student population, were often stronger than those from the main results; all estimates were statistically significant and were slightly larger than the estimates from the main analysis. This suggests that the large estimates of selective school effects (from the main results, section 3.7.2.3) may be partially attributable to selective students outperforming non-selective students at LBOTE dominated schools, which are slightly more likely to be government schools (see Table 3.10).

Table 3.22: Regression and Propensity Score Analyses (varying sample), Boys^

Outcome	Method	Sample			
		Main Sample	Metro	Non-Gov.	LBOTE
ATAR ≥ 95	Regression	-0.062*	-0.066*	-0.117***	0.067
	Nearest Neighbour	-0.065	-0.046	-0.125	0.070
	Kernel (bin width= 0.06)	-0.020	-0.036	-0.077	0.062
	Stratification	-0.058	-0.073	-0.111	0.054
ATAR ≥ 90	Regression	-0.024	-0.022	-0.081*	0.099**
	Nearest Neighbour	-0.026	-0.026	-0.077	0.139**
	Kernel (bin width= 0.06)	0.035	0.020	-0.017	0.097**
	Stratification	-0.003	-0.018	-0.042	0.096**
ATAR ≥ 85	Regression	0.015	0.014	-0.037	0.069*
	Nearest Neighbour	0.018	0.027	-0.021	0.070
	Kernel (bin width= 0.06)	0.053**	0.039*	0.015	0.062*
	Stratification	0.019	0.008	-0.004	0.064*

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

^ The samples exclude students with missing outcome values (i.e. they are not imputed).

For the boys, the PSM results from the metropolitan sample were similar to those of the main results, with negative estimates for ATAR ≥ 95 and ATAR ≥ 90 and positive estimates for ATAR ≥ 85, and with only the estimate for ATAR ≥ 85 from kernel matching being statistically significant, at 3.9 ppt. For the non-government sample, the estimates appeared to be slightly more negative across all three outcomes but were all statistically indistinguishable from zero, aside from the regression estimates. The regression estimates, which do not account for selection, suggested there was a large negative selective school effect at -11.7 ppt and -8.1 ppt for ATAR ≥ 95 and ATAR ≥ 90, respectively.

The results from the LBOTE sample suggested that boys attending selective schools outperformed non-selective students who attended schools with a large proportion of LBOTE students (> 50%). The estimates are consistently positive for ATAR ≥ 95, ATAR ≥ 90 and ATAR ≥ 85, and were almost all statistically significant for the latter two outcomes. Estimates were large in magnitude at over 9.6 ppt for ATAR ≥ 90 and over 6.2 ppt for ATAR ≥ 85.

3.8.3.2. Imputing Missing ATARs - Results

Table 3.23 presents PSM estimates for both the Main Sample and the Metropolitan Sample when the missing ATARs are imputed as per Houng and Justman (2014). For both girls and boys, the results from imputing the missing ATARs were very similar to those from the main analyses, where students with missing ATARs were excluded from the sample. Estimates were slightly larger or smaller in effect size (less than 3ppt), while for some estimates statistical significance disappeared.

For the girls, the estimates with imputed ATARs were less than a few percentage points smaller than when students without ATARs were excluded from the PSM analyses (Table 3.18). For the boys, there was a slight increase in effect size for the only significant estimate from kernel matching for ATAR ≥ 85, while the estimates for ATAR ≥ 90 and ATAR ≥ 90 remained unchanged in statistical significance.

Table 3.23: Propensity Score Analyses with Imputed Missing ATARs

Outcome	Method	Girls		Boys	
		Main	Metro	Main	Metro
ATAR ≥ 95	Nearest Neighbour	0.077	0.065	-0.035	-0.076
	Kernel (bin width= 0.06)	0.166***	0.156***	-0.007	-0.025
	Stratification	0.145***	0.133***	-0.050	-0.060
ATAR ≥ 90	Nearest Neighbour	0.084*	0.073*	-0.006	-0.034
	Kernel (bin width= 0.06)	0.118***	0.110***	0.032	0.015
	Stratification	0.104***	0.097***	-0.012	-0.022
ATAR ≥ 85	Nearest Neighbour	0.026	0.047	0.036	0.002
	Kernel (bin width= 0.06)	0.061***	0.057***	0.044*	0.029
	Stratification	0.053**	0.050**	0.011	0.002

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

3.8.3.3. Difference-in-Difference Results

For the difference-in-difference analyses, I adopt the conservative sample of students residing in metropolitan areas. Table 3.24, which shows frequency counts by SES quartile (where the 1st quartile is the lowest), confirms there are sufficient numbers of observations in order to make sensible comparisons between selective (S) and non-selective students (NS). Note that observations generally increase with increasing SES quartile for both non-selective and selective students, particularly for the non-selective students.

Table 3.24: Frequency Counts by Selective Attendance, Sex and SES Quartile

	SES Quartile									
	1st		2nd		3rd		4th		Total	
Sex	NS	S	NS	S	NS	S	NS	S	NS	S
Girls	87	25	116	25	349	24	666	104	1,218	178
Boys	101	19	221	27	426	41	1,218	189	1,966	276
Total	188	44	337	52	775	65	1,884	293	3,184	454

NS: Non-Selective, S: Selective, ascending SES quartiles signify increases in SES.

The estimates from the difference-in-differences analysis are presented in Table 3.25 below. For the girls, I find increases in the success rates for ATAR ≥ 85 and for ATAR ≥ 90 for the 1st, 3rd and 4th SES quartiles; estimates range between 10.0 ppt and 46.6 ppt, and the effects are larger for ATAR ≥ 90 than for ATAR ≥ 85 and effects also increased with decreasing SES quartile.

Table 3.25: Difference-in-Difference Estimates by SES and Sex

Sex	Outcome	SES1	SES2	SES3	SES4
Girls	ATAR ≥ 95	17.3	0.1	-0.3	6.8
		(14.8)	(14.7)	(14.6)	(7.1)
	ATAR ≥ 90	46.6***	17.9	28.2**	19.6***
		(10.3)	(10.5)	(9.6)	(4.4)
	ATAR ≥ 85	34.5***	14.7	18.6*	10.0**
		(9.6)	(9.4)	(8.0)	(3.5)
Boys	ATAR ≥ 95	19.2	10.5	-17.5	-18.6***
		(16.8)	(13.6)	(11.0)	(5.4)
	ATAR ≥ 90	32.7**	29.3**	19.2*	5.8
		(12.1)	(10.1)	(8.1)	(3.7)
	ATAR ≥ 85	28.4*	31.0**	26.6***	5.2
		(11.2)	(9.8)	(7.3)	(3.1)

* p < 0.05, ** p < 0.01, *** p < 0.001, ascending SES quartiles signify increases in SES.

The difference-in-difference analyses also produced large positive effects for the boys, for the students from the first to third SES quartiles and only for the outcomes ATAR ≥ 85 and for ATAR ≥ 90. The effect sizes were around 30 ppt for students in both the 1^{st} and 2^{nd} SES quartiles and were 19.2 ppt for ATAR ≥ 90 and 26.6 ppt for ATAR ≥ 85 for students in the 3^{rd} SES quartile. The estimates were generally statistically indistinguishable from zero for the outcome ATAR ≥ 95, and likewise for the highest SES quartile, with the sole exception being that boys in the highest SES quartile performed 18.6 ppt worse relative to their non-selective counterparts.

The much larger magnitude of the estimates from the difference-in-difference analyses in Table 3.25 can be attributed to: comparisons stratified by SES quartile, not accounting for selection like in the matching analyses, and the simplified approach in comparing two different achievement outcomes.

Table 3.26 presents additional difference-in-difference estimates for a general effect, rather than by SES quartile. Aside from ATAR ≥ 95 for girls, the estimates are statistically significant for all outcomes, ranging between -12.6 ppt and 23.1 ppt, and are smaller in magnitude compared to the local estimates from Table 3.25 for subgroups by SES quartile.

Table 3.26: Difference-in-Difference Estimates by Sex

	Outcome		
Sex	ATAR ≥ 95	ATAR ≥ 90	ATAR ≥ 85
Girls	5.0	23.1***	14.0***
	(5.4)	(3.5)	(2.9)
Boys	-12.6**	12.9***	13.4***
	(4.4)	(3.1)	(2.7)

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The difference-in-difference estimates are similar but slightly larger than those produced from the probit regressions, which like the difference-in-difference comparisons do not account for observable selection into treatment. The average marginal effects from the probit regressions (in Table 3.12) ranged between 11.3 ppt and 18.3 ppt for girls and was only statistically significant for ATAR ≥ 95 at -6.2 ppt for boys. Note that the difference-in-difference estimates of both local and general effects were robust to additional controls of a continuous measure of SES and a LBOTE indicator.

3.9. Discussion

In this section I summarize the results from the main analyses and the robustness checks, discuss limitations of the research, and also interpret the results in the context of the literature.

3.9.1. Summary

The main results from the PSM analyses showed gains for girls attending selective schools, with increases in probability of over 15.2 ppt for ATAR ≥ 95, 9.6 ppt for ATAR ≥ 90, and 6.2 ppt for ATAR ≥ 85[83]. For the boys, only the kernel matching produced an estimate of 5.3 ppt for attaining an ATAR ≥ 85.

For robustness checks, I conduct sensitivity analyses of the main results by varying the samples to obtain more narrowly defined control groups, and re-run the analyses with imputed missing ATARs to determine the extent of potential bias from excluding students who completed the International Baccalaureate, or moved interstate. I also estimate local effects by SES quartile with difference-in-differences to complement the PSM analyses. In the sensitivity analyses of varying the samples, I restrict the pool of students to: 1) students who reside in metropolitan areas; 2) those at non-government schools; and 3) those at schools where at least 50% of the students are LBOTE.

For both boys and girls, the results from varying the samples supported those from the main analyses but were mostly smaller in magnitude for the metropolitan sample and the non-government sample. The LBOTE sample, however, resulted in larger estimates than that from the main analyses, and in addition, there was strong evidence of a positive effect for ATAR ≥ 90 for the boys, whereas in the main analyses the results were much weaker for this outcome.

Similar to varying the samples to obtain better control groups from the metropolitan and non-government samples, imputing the missing ATARs for the girls produced estimates that were smaller in magnitude than those from the main analyses. Estimates of the selective school effect for the girls were mostly statistically robust for all outcomes, with minimum significant effects of 5.3 ppt for ATAR ≥ 85, 7.3 ppt for ATAR ≥ 90 and 13.3 ppt for ATAR ≥ 95.

The difference-in-difference analyses indicated that there were positive effects from attending selective schools concentrated in the outcomes ATAR ≥ 85 and ATAR ≥ 90 for both boys and girls. One could interpret this as suggesting that selective schools provide a floor for the

---

[83] All estimates were statistically significant with the exception of nearest neighbour matching for ATAR ≥ 85, which produced an estimate of 0.017.

achievement outcomes of their students. Girls in the lowest SES quartile appeared to benefit the most with large positive and statistically significant estimates for ATAR ≥ 85 and ATAR ≥ 90, while boys in the highest SES quartile appeared to be negatively affected in attaining ATAR ≥ 95 from selective school attendance.

3.9.2. Limitations

The limitations of the research concerned finding students for the control group that most resembled those that enter selective high schools. The problem of finding a suitable counterfactual arises from the issue of unobservables, where students who express a preference to attend a selective high school may differ in ways that may lead to better outcomes. Without observing the participation of students in the entrance exam, the analyses cannot account for the unobservable differences such as potentially greater motivation.

The study relied on data from only one cohort of students and only a very limited number of selective schools (not specified for privacy reasons). The results are time sensitive for both the students in the cohort at the time and also the environments at the schools that are analysed. In addition, the small number of selective schools in the data sample means that the selective school effects can readily interpretable as individual school effects.

Another related data limitation is that NAPLAN is used to approximate and exclude the group of discretionary-intake students, who were even more likely to have been selected for positive unobservables. As a result, more students are excluded from the sample than would have been necessary with the entrance exam data to ensure the exclusion of these discretionary-intake students.

In relation to the methods, the main limitation was the constraint imposed by the balancing (ignorability of treatment) requirement from the propensity score matching, such that treatment and control outcomes are independent conditional on the pre-treatment characteristics. It is a strict requirement, and the outcomes and characteristics may not be fully independent for two reasons: the first is that in the most narrowly defined measures of socioeconomic background (parental education and occupation) there are clear differences between selective and non-selective students (section 3.4.2); and secondly, there are unobservables that differ between treatment and control groups, meaning that prior achievement, language and socioeconomic background may be insufficient to accurately model attendance of selective high schools. For the difference-in-difference analyses, taking differences of two different achievement measures, NAPLAN and ATARs, introduced measurement error.

In all, despite the limitations of the research, the large effect sizes produced suggest that it is possible that there are gains from attending selective high schools, though the actual effects are probably smaller than estimated from the main analyses.

3.9.3. Research Context

College Outcomes

This empirical investigation uncovered new evidence of benefits from attendance of selective high schools. The finding of positive selective school effects are consistent with Dobson and Skuja (2005), who found that students from selective schools performed worse than students from non-selective schools in the first year of university. It is possible that selective schools improve the academic outcomes of certain students in high school, but their students subsequently revert back towards their Year 9 baseline in a university environment. Nonetheless, positive selective school effects can be interpreted as successes for the individual, as Year 12 outcomes determine, to a large extent, university and course admission.

There are several mechanisms that could explain the existence of positive selective school effects, including: more academically inclined peers which encourages competition among students, better or more motivated teachers and staff, and also potentially access to more resources. Intangible benefits could include school history and culture, the success of alumni, as well as school reputation.

As to speculative reasons for the fall in academic performance at university, the subsequent drop-off could be attributed to the absence of the academically conducive high school environment, which could lead to difficulty in adjusting to a more self-reliant university environment. Other possibilities include: a reduction in effort from students, having attained their university and course entrance goals; or alternatively, the cumulative impact from the heavy study workload, expectations pressure, and the possibly longer commuting time associated with selective schools attendance. Note that long commuting times for some students arises from the limited number of selective schools in combination with the large metropolitan area from which the student intake is drawn.

Peer Effects and Self Concept

The large positive effects for the girls attending selective high schools are similar to results documented by Lavy et al. (2007), who find using within-student test score variation from grade 6 that girls benefit from stronger peers for their ninth grade results[84]. The stronger results for the girls compared to the boys suggest that girls benefit more from having higher achieving peers than boys.

Related research on the relative status of students in test scores affecting their self-concept could help explain the weaker relative results for the boys; a negative relationship has been established between individuals' scores and that of their peers, when measured by classroom rank (Murphy and Weinhardt, 2014), or in reference to the average scores of peers (Marsh and Hau, 2003). At selective schools the concentration of high achieving students means that the majority of students will experience large decreases in classroom rank upon attendance, and as a result, they may be more sensitive to their change in relative status. The results in this study imply that the negative effects from these changes may be larger for boys.

Student Background and Institutions

Studies on selective schools from the UK and the USA have found little effect from attendance of selective high schools on test scores or college outcomes. While there are differences in the socioeconomic background of the students in this study compared to the earlier studies, it is unlikely that these differences explain the limited results in the earlier research as the statistical techniques in all studies compare selective students with comparable non-selective students; in this study there is a high representation of LBOTE students in selective schools with comparatively high SES, which contrasts with the low socioeconomic background of students in the Boston and New York studies.

An advantage of our data over earlier work is that the data includes students at non-government schools, whom comprise around two thirds of the students in our data sample, while about 50% of attendees at selective high schools were previously from the non-government sector[85]. In Adulkadiroglu et al. (2014), and Dobbie and Fryer (2014), and Clark (2010), high school outcomes were limited to comparisons between students at selective and government high schools, potentially biasing estimates of selective school effects upwards because students who attended government high schools were typically from low or lower socioeconomic backgrounds.

[84] A 10% increase in students in the top 5% of students implies an improvement of 10% of the within-student standard deviation of grade 9 scores.
[85] Students attending non-government schools comprise around 40% of all students and around 50% of applicants to, and attendees of, selective schools. Source: author's calculations from NAPLAN data.

The educational outcome assessed in this study may however partially explain the stronger results as students may be more motivated to perform well as there is a greater incentive for them to do so and also because the other studies were not able to assess the analogous educational outcome. Australian university entrance results are a high stakes outcome in that university and course admissions are almost solely determined by these results and consequently have large implications for career paths and expected incomes. The analogous educational outcomes in the UK and the USA studies were not assessed due to lack of data, and also do not determine university outcomes to such a large extent, as is the case in Australia.

Methods

Aside from institutional factors and the better access to data, the absence of results from earlier work (Adulkadiroglu et al., 2014, Dobbie and Fryer, 2014 and Clark, 2010) could also be explained by their preferred econometric techniques. With regression discontinuity and instrumental variable methods, these previous studies estimate the effect of attendance on marginal students, by comparing students who were only just admitted against those that just missed out. Although this is a convincing method due to its resolution of the unobservables problem, the results from these comparisons may alter the treatment because marginal students are likely to suffer from a decrease in academic performance arising from diminished self-concept. That is, marginal students who are only just admitted are academically the weakest at the selective high schools, while the students who just missed out on admission are likely to be academically stronger students in the schools they attend, which would lead to the estimates of the selective school effect being understated.

The positive results from the system-wide allocation of students by academic performance found in the case studies of high schools in Trinidad and Tobago and Romania are also supportive evidence for there being gains in academic performance from entrance based admissions into high school. The idea of relative status or self-concept can also reconcile both the absence of previous findings for selective schools (due to technique) and the positive findings for system-wide allocation of students by academic performance. Under the system-wide allocation of students, at lower achievement levels it is unlikely that differences in marginal students from their peers are as stark as what occurs for applicants of selective high schools at the margins for receiving an offer.

3.10. Conclusion

From a state cohort of Year 9 students from an anonymous Australian state, I estimate the effect of attending selective high schools on students' university entrance results. With the challenge of finding a group of students who most resemble those that enter selective schools, I limit non-selective students to those with high prior achievement. For estimation, I favour Propensity Score Matching over largely equivalent nonparametric regression approaches for its intuitiveness; I model attendance of selective high schools with prior achievement and language and socioeconomic background, and subsequently compare selective and non-selective students with similar probabilities of attendance.

In contrast to earlier work, which concentrated on assessing the effect of attendance on marginal students against government school students, the full student population in the data allows for the estimation of a general selective school effect. The analyses may, however, be considered exploratory due to concerns that the control groups are not similar enough to the students who attend selective high schools. Specifically, like with many of the earlier studies, the probable traits of additional motivation and aspiration in the selective students are unobserved which is likely to mean that the results are overestimates. There are also limitations in sample size from having only one cohort of students and only a limited number of selective schools.

From the matching analyses, I find positive effects of attendance for attaining successful Year 12 outcomes from the PSM analyses. For selective girls, I find large increases in probability of attainment with minimum significant effects of 6.2 ppt, 9.6 ppt, and 15.2 ppt, for ATAR ≥ 85, ATAR ≥ 90, and ATAR ≥ 95, respectively. I also find evidence of increases in probability of around 5 ppt for attaining ATAR ≥ 85 for the boys.

The robustness checks, which comprised of varying the samples and also imputing missing ATARs, supported the main results for the girls, and also the effects for ATAR ≥ 85 for the boys. The difference-in-difference analyses by socioeconomic quartile suggested that gains were large for girls and boys for attaining ATAR ≥ 85 and ATAR ≥ 90, and that there was a negative effect for boys in the highest socioeconomic quartile in attaining ATAR ≥ 95.

The results from this chapter offer tentative support for the existence of positive peer effects arising from the concentration of high achieving students at selective schools. In this chapter I find positive results for selective school attendance, but the comparisons do not account for the probable higher levels of aspiration of students attending the selective schools. Girls appear to benefit more from having high achieving peers, with one possible explanation being that boys are more sensitive to changes in relative status from selective school attendance that potentially influences students' self-concept.

The research on relative status and student achievement could help explain the absence of results from earlier work due to a methodological preference for regression discontinuity approaches that concentrate on analyses of marginal students. That is, comparing the outcomes of the academically weakest students at selective schools against academically stronger students at other schools, who do not differ on the observables, could lead to understating the effect of selective school attendance.

Regarding extensions to this research, entrance exam results would help identify students with a preference to attend selective schools, which should improve the comparability of control and treatment groups. In addition, with entrance exam results one could verify the consistency of the absence of selective school effects for marginal students from a regression discontinuity approach with the identification of significant effects for all selective students from propensity score matching. The inclusion of subsequent cohorts would provide greater assurance that the results are not particular to one cohort and would also allow for comparisons within more narrowly defined subgroups, such as by language background, without encountering small sample problems.

Appendix

Table 3.27 (A): Selection Model with Quadratic Terms for Year 9 Scores

	Girls		Boys	
	(1)	(2)	(3)	(4)
Std. NUM	0.143***	0.101*	0.171***	0.152***
	(0.038)	(0.046)	(0.033)	(0.043)
Std. NUM ^2	-0.015*	-0.010	-0.022***	-0.019**
	(0.007)	(0.008)	(0.006)	(0.007)
Std. RDG	0.071*	0.046	0.139***	0.128***
	(0.032)	(0.038)	(0.032)	(0.039)
Std. RDG ^2	-0.012	-0.009	-0.031***	-0.029**
	(0.008)	(0.008)	(0.009)	(0.009)
LBOTE	0.145***	0.141***	0.137***	0.135***
	(0.014)	(0.014)	(0.011)	(0.011)
SES	0.001	0.005	0.006	0.006
	(0.004)	(0.009)	(0.004)	(0.007)
SES X Score categories	N	Y	N	Y
Tjur	0.173	0.186	0.144	0.146
N				
Selective	1,738	1,738	2,697	2,697
Non-Selective	189	189	279	279
Total	1,927	1,927	2,976	2,976

Chapter 4. A Comparison of Two Methods Using the Exam Sample

4.1. Introduction

This chapter also assesses whether academically selective schools add to the achievement of their students, beyond what they would achieve in non-selective schools[86]. The research contributes to the literature by estimating the selective school effect in the Australian context with more sophisticated statistical methods than previous studies and explains how test score performance reflects the incentives provided by different measures of achievement.

This second set of analyses follows the cohort of Year 7 students from an anonymized Australian state in 2007. We estimate the selective school effect via two methods: propensity score matching, which compares students of similar background and prior achievement, and regression discontinuity (RD), which compares marginal selective and non-selective students based on performance on the entrance exam.

As part 2 of the main analyses, this chapter takes advantage of new information from the admissions process. For the matching approach, we identify an improved comparison group consisting of applicants to the selective high schools who did not subsequently attend the schools. These non-selective students, who applied to the selective schools and sat the entrance exam, are likely to share unobservable characteristics with the selective students, including their educational aspirations.

As a quasi-experiment, the RD approach compares students who just got into the selective schools, and those who just missed out. It accounts for the unobservable differences between selective and non-selective students by exploiting the admissions thresholds in the entrance exam, which are arbitrarily determined by the number of positions available at the schools. The RD approach is arguably more convincing than the matching, but is subject to greater limitation from small sample sizes and estimates a local effect rather than a general effect.

Previewing the results, the selective school effect appears to be small. We find estimates of approximately 2 percentile points in university entrance ranks from the matching approach, but the effect may be overstated. By making comparisons based on tests taken before the admissions process, the estimates do not account for the positive selection between students who go on to attend selective schools and students who decline or do not receive offers to attend selective schools. There is also evidence, with the same caveat of positive bias, that the

[86] Chapter 4 is an iteration of a research project conducted in 2016 and is joint work with my supervisor Associate Professor Chris Ryan.

selective schools provide a floor for Year 12 achievement; estimates are between 6 ppt and 13 ppt increases in probability of attaining ranks in the top 10% and 15% of students.

In contrast, the regression discontinuity estimates are not significant, suggesting that marginal selective students may not experience the same gains in academic performance that the majority of selective students do. One possible mechanism for differences between local and general effects is that marginal selective students are more negatively affected by entering academically competitive schools due to experiencing large decreases in their relative academic status. The other limitation from the RD approach is that the small sample sizes arising from the reliance on one cohort of students could also be the explanation for the lack of statistically significant estimates.

Refer to Chapter 3: Literature Review (section 3.2) for the relevant literature for this chapter. Chapter 4 is structured as follows. Section 2 presents the background information of the selective high schools and establishes the extent of academic selection at the selective schools. A description of the data is presented in Section 3. Section 4 presents the matching approach, with subsections for: sample definition, descriptive statistics, methodology, the selection model, and then the results. Section 5 is on the regression discontinuity approach, where we describe the methodology, establish its suitability and present the results. In Section 6, we discuss the results from each of the two approaches, and the influence of unobserved educational aspirations in explaining differences between the estimates from each approach. Section 6 also discusses the limitations of the case study, and our findings are summarized in Section 7, the conclusion.

4.2. Background

4.2.1. Selective High Schools and the Admissions Process

This case study follows a state cohort of students in Year 7 in 2009, through to the end of high school, and specifically the subset of students who participate in the admissions process to enter a set of selective high schools. The selective schools were attended by less than 2% of the student population. The entrance exam for admission into the selective high schools is held annually in the year preceding entry. Results from the entrance exam determined the vast majority of places, with the balance of students comprising of: applicants who were exceptions to a quota rule designed to limit the impact of selective schools on other schools; and applicants from a disadvantaged background who performed well on the exam.

4.2.2. Academic Selection

The level of academic competitiveness for attending selective schools is illustrated in Figure 4.1, which presents histograms of Year 7 standardized test scores for the general population, in the left panels, and the students who attend the selective schools, in the right panels; one selective school is excluded for being in its first year of operation (discussed later in section 4.4.1.1, the Data Sample, the Selective Schools). Histograms for boys and girls are presented in the top and bottom panels, respectively. The test scores are the average of Reading and Numeracy, which are each standardized to have means of zero and standard deviations of one.

Figure 4.1: Prior Achievement for General Population and Selective Students^

^ based on Year 7 standardized test scores. Red and green vertical lines indicate minimum and mean values at selective schools respectively: for girls is -0.19 and 1.71; and -0.05 and 1.71 for boys.

For boys and girls separately, the minimum and mean Year 7 test scores for students at selective schools are indicated by the red and green vertical lines. The minimum selective student for girls is at the 42nd percentile of all students, and corresponding value for boys is the 48th percentile. The means are far higher, placed in the top 5% of all students, at almost the 96th percentile, with the same value for boys and girls. At the most academically competitive selective school, the minimum score (0.58) was at the 72nd percentile of all students.

Note that more boys were admitted to selective schools, consistent with the greater proportion of boys among students with higher levels of achievement on the entrance exam. [87]

4.3. Data Description

4.3.1. Standardized Tests and Year 12 results

The data consists of nationally standardized tests, “National Assessment Program – Literacy and Numeracy” (NAPLAN), from Years 7 and 9 from 2009 and 2011, and Year 12 results from 2014. To represent prior achievement, we use Numeracy and Reading from Year 7 NAPLAN, and a combined variable that is an average of the two. For ease of interpretation, the Numeracy and Reading were first standardized with mean zero and one standard deviation. Tests scores from both domains follow a normal distribution as shown in Figure 4.1 earlier.

We adopt the same Year 12 achievement outcomes as in Chapter 3, which were defined to provide meaningful measures of success for the high-achievers: ATAR ≥ 95, ATAR ≥ 90, and ATAR ≥ 85; effectively scores within the top 5%, 10%, and 15% of results, respectively. Australian Tertiary Admission Ranks (ATAR) values are percentile ranks derived from a combination of the results from subjects completed by students in Year 12.

Refer to Chapter 3: Data Description (section 3.4.1) for further details of the standardized tests and the Year 12 assessment.

4.3.2. Demographic Information

The demographic information of the student population from the anonymized Australian state is shown in Table 4.1. There were 67,687 students aged 14 to 16 in Year 9, of whom 42,193 (63.4%) appear in the Year 12 data. Details of the fuzzy matching involved in combining the administrative and standardized data with the data from the selective schools admissions process are included in the Appendix, A.2.

Table 4.1: Demographic Information of Student Population

	YR7, 2009	YR9, 2011
Number of students	66,509	67,687
Male (%)	51.5	52.4
Aged 14 / 15 / 16 in 2011 (%)	17.6/76.8/5.6	17.7/76.3/6.0
Language Background Other Than English (%)	25.0	23.4
Aboriginal and Torres Straits Islanders (%)	1.3	1.3

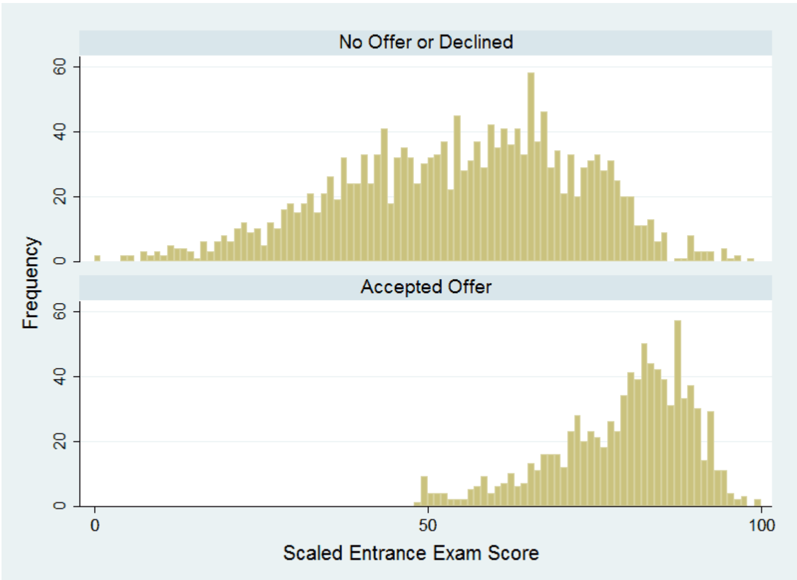
[87] From the sample of applicants, the percentage male, as a local weighted average, increased from 50% to around 60% over the full range of exam results (see Figure 4.18 A in Appendix A.1, which shows a locally weighted regression of the percentage male over percentile ranks of entrance exam results).

4.3.3. The Entrance Exam

The entrance exam data included information relating to the school preferences of applicants, their equity status – relating to one of the admission intakes, and their exam scores and outcomes, which were comprised of whether and what types of offers were accepted.

Figure 4.2 presents the histograms of the entrance exam scores by the outcomes of the admissions process. It presents a composite measure of the tests undertaken, which we scaled from zero to a hundred. It is clear that the distribution of entrance scores for those who attend the selective schools lies substantially to the right of those who do not attend the schools. There are two main groups who do not attend – those whose test results gave them a rank in the distribution too low to be offered a place, and those whose test results were such that they received offers, but they declined them. [88]

Figure 4.2: Histograms of Entrance Exam Scores by Selective Attendance



[88] A small number of students also missed out on receiving offers due to quotas intended to limit the loss of students from other schools.

4.4. Propensity Score Matching

4.4.1. The Data Sample

For the analysis our challenge is to find a group of students comparable to those who attended selective high schools. Given that admission is based on academic performance, the most comparable students to those at selective high schools are students with similar prior achievement. With access to standardized test scores and background characteristics of students from the general population (a state cohort), we identify two sets of students for our initial comparisons:

a) the selective potential sample ("SP"), who are the set of students with Year 7 achievement above the minimum level observed among those who subsequently attended the selective high schools[89]; and

b) the "exam sample", who are a subset of the selective potential sample, and have the added condition that they also sat the entrance exam.

For later descriptive purposes, when we review differences between the comparison groups, we label the non-selective students among the selective potential sample "SP" and the non-selective students among the exam sample "NS"; the selective students are labeled "S".

From these two groups of non-selective students we opt to use the sample of students who sit the entrance exam, reasoning that the students who sit the entrance exam but do not attend the selective schools are likely to most resemble the students who attend selective schools. We confirm that this is the case in the descriptive statistics (section 4.4.2), where we present student characteristics for the non-selective students, in the selective potential and exam samples. Further, sitting the test reveals a high level of academic ambition and motivation among these students, which we think deals to some extent with otherwise unobserved factors that are problematic in using matching estimators.

Within the sample of students who sit the entrance exam, we estimate a pooled selective school effect. In the first instance, we exclude a small number of students admitted to selective high schools on a discretionary basis, entailing an additional written application and interview process, due to concerns they have been selected for other characteristics that are not reflected in the data.

---

[89] With one selective school excluded, discussed in the next subsection, Selective Schools.

Students admitted through the discretionary intake are a subset of the students not admitted despite attaining sufficiently high results on the entrance exam due to a quota rule intended to protect the loss of students from other schools. Including these positively selected students is problematic for estimating the overall effect from selective school attendance due to the anticipated positive selection bias from the unobserved factors that led to their selection. We also excluded students from a disadvantaged socioeconomic background admitted through another intake, for the lack of sample size.

In summary, we define the main data sample as students who sit the entrance exam who have prior achievement of at least that of the minimum level observed among those who subsequently attended selective high schools: the exam sample. We choose to exclude the disadvantaged intake and exclude the discretionary-intake, whose inclusion is likely to positively bias the results (discussed above), from the pool of students with high prior achievement for our base case analysis. We also exclude students who attend one of the selective schools for the reason that they entered the school in its first year of operation as a new selective school (discussed next). The discretionary intake students are discussed further in the Appendix, in A.3.

4.4.1.1. The Selective Schools

In the main analyses, we estimate the selective school effect at the aggregate level, pooling the selective schools. In order to present reliable estimates, we exclude one of the selective schools from the analyses as it is a new school for which the intake of students we observe is the first of its cohorts to reach Year 12; there are several aspects about this cohort of students that differ from those observed for the other selective schools.

First, few students who sat the select entry test identified this school as their first preference school to attend, meaning that relatively few of those who finished up attending it had placed it first. Second, the prior achievement and social background profile of its first intake differed substantially from other selective school applicants. In particular, the minimum entrance exam result required to receive an offer was much lower than the other schools.

Lastly, when the minimum NAPLAN score at the excluded selective school was used to define the selective potential sample, the covariate balancing tests[90] of the intake and the comparison group who sat the test were not passed for that sample, which we ran as a robustness check. Failing the balancing tests is further confirmation that the students at the

[90] From the matching approach, refer to Methodology, section 4.4.3.

excluded selective school are unlike the other selective students. For all these reasons, we decided to exclude this school from the analyses that follows.

4.4.2. Descriptive Statistics

In this section we review the prior achievement, student characteristics and achievement outcomes of the two groups of comparable students that we defined in section 4.4.1, the Data Sample.

4.4.2.1. Prior Achievement

For the purposes of defining the data sample we derive a NAPLAN variable that is the average of Numeracy and Reading, which are themselves standardized with mean zero and standard deviation of one from the whole student population. Table 4.2 shows the mean and standard deviation in NAPLAN scores of selective students who were admitted to selective school strictly on the basis of their entrance exam results. That is, it excludes the discretionary intake and disadvantaged intake students. The table also shows the average NAPLAN score for non-selective students from both the selective potential ("SP") and exam samples ("NS").

Table 4.2: Standardized Year 7 NAPLAN Scores by Data Sample

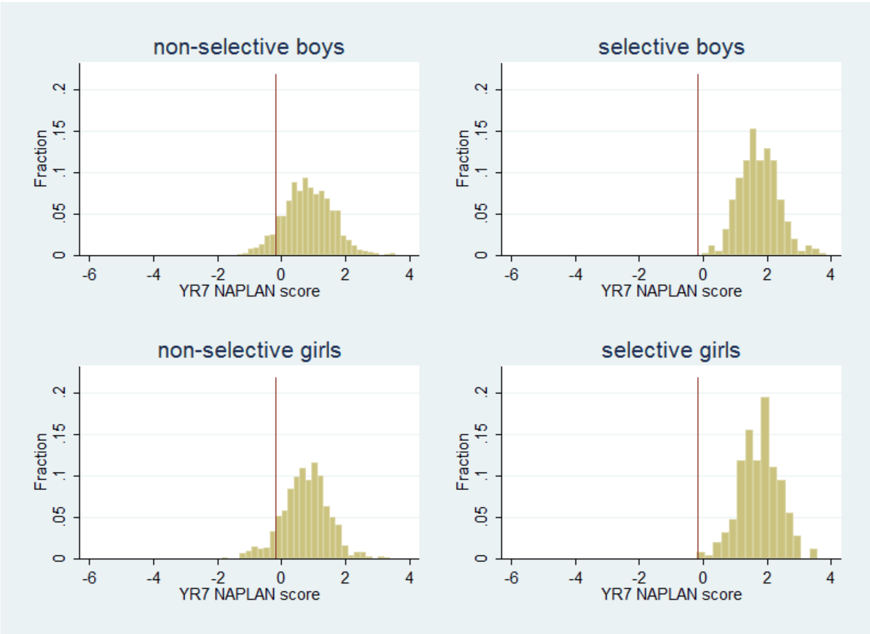
	Mean	SD	Missing (%)	Count
Selective	1.76	0.63	0.7	592
Non-Selective				
Sat Exam (NS)	0.94	0.64	0.5	1,550
Selective Potential (SP)	0.60	0.59	1.6	34,231

It is apparent that the NS students, who sat the exam but did not attend selective schools, had lower NAPLAN scores than selective students. The mean NAPLAN score for the non-selective exam sample is 0.8 standard deviations below that of the selective sample (0.94 compared with 1.76), whereas the selective potential sample is more than one full standard deviation lower in NAPLAN score (difference of 1.16). The difference among applicants whose first preference was the most competitive selective school between non-selective (sat exam) and selective samples was smaller at 0.55 (not shown).

Non-selective students who sit the exam are nonetheless closer to the selective school students in terms of Year 7 NAPLAN than the broader school population. This is perhaps unsurprising as only around 4% of the student population sat the admissions test, and in addition, it is probable that the students who do are informed by their perceived likelihood of getting into the selective schools.

The suitability of the exam sample as a comparison group is demonstrated visually in the histograms in Figure 4.3 below. The histograms of Year 7 NAPLAN scores are shown by selective school attendance and by sex. The red vertical lines indicate the lowest minimum cut-off score among the selective schools, which are used to limit the sample of both selective and non-selective students. Reassuringly, the numbers of students to the left of the red vertical line for the non-selective students are relatively few (left panels) – meaning that the non-selective students in the exam example sample are fairly comparable to the group of selective students.

Figure 4.3: Histograms of YR7 NAPLAN Scores for the Exam Sample^



^ The red lines show the minimum Year 7 NAPLAN scores at the selective schools, at -0.19.

4.4.2.2. Student Characteristics

Table 4.3 presents average characteristics for the selective students and the two comparison groups of non-selective students (defined in the Data Sample, section 4.4.1)[91]. With all students having NAPLAN scores above the minimum at the selective schools, the groups consist of those who attended the selective schools ("S") (other than the omitted school), those who did not attend the selective schools ("SP"), and those who sat the entrance test but did not attend a selective school ("NS"). The rows contain information of the language background and socio-economic background[92] in quartiles of these groups, as well as sector of the school attended by students in each group.

Table 4.3: Characteristics of Selective and Non-Selective Student groups

	Selective (S)	Non-Selective	
		Sat Exam (NS)	Selective Potential (SP)
N	592	1,520	32,446
Male (%)	57.4	51.4	47.0
LBOTE (%)	80.1	75.8	22.5
SES Category (%)			
Quartile 1	11.3	14.9	16.0
Quartile 2	13.0	17.0	21.4
Quartile 3	27.0	26.8	28.3
Quartile 4	48.6	41.2	34.2
YR7 Sector (%)			
Government	54.9	54.5	46.6
YR9 Sector (%)			
Government	100	51.8	46.1

^ above minimum score among students at Selective Schools (excludes one school).

[91] In the analyses we assess the girls and boys separately. The student characteristics for those samples are included in the Appendix A.4, in Table 4.10 (A).

[92] SES quartiles are based on a socioeconomic index derived from parental education and occupation information using principal components analysis (see Chapter 3, section 3.4.2).

The socio-economic backgrounds of the two groups who sit the select entry test are much more alike than either group is with the group of students who do not sit the test. The broad definition of "selective potential" students created from the general population meant that the SP non-selective students sample size was also very large, comprising almost half of all students (32,446). This is even after excluding the new selective school, and taking the minimum NAPLAN score from the selective school with the second lowest admissions threshold to define the samples.

In terms of language background, selective school students are overwhelmingly from a foreign language background; over 75% of both the selective school students and those who sat the test but did not attend were from foreign language backgrounds, compared with less than a quarter of the group who did not sit the test. Selective school students were also from a higher socio-economic background, with greater representation in the top quartile of the socio-economic distribution[93], of over 40%. There were also a slightly greater proportion of males in the selective (57.4%) and non-selective exam (51.4%) samples, compared with non-selective students from the selective potential sample (47.0%).

Since an important part of the research design for this project is to find the right comparison group for the selective school students, it is clear that an important first step can be made by focusing on the group who sat the test as the relevant comparison group. From Table 4.3, their observed socio-demographic characteristics were already like those of the group who attended the school in terms of language background and social background. We posit that their motivation levels are similar, since they were prepared to sit the test, and it is likely that any important unobserved characteristics are also more like those of the group who attend selective schools.

[93] The SES index is derived from parental education and occupation using principal components analysis and has a mean 0 with a standard deviation of 1 for Year 7 in 2009 and also for Year 9 in 2011.

4.4.2.3. Achievement Outcomes

Table 4.4 presents mean ATARs, the derived binary ATAR outcomes, and the percentage of missing values for the selective sample and the two non-selective samples[94]. The difference in prior achievement between selective students and non-selective students from both the exam (NS) and selective potential (SP) samples is also apparent for the Year 12 achievement outcomes. While there remain large differences in ATAR ranks between selective students and non-selective students who sat the exam, there are even larger differences between selective students and non-selective students with selective potential.

Table 4.4: Achievement Outcomes of Selective and Non-Selective Students

	ATAR	% Missing	ATAR ≥ 95 (%)	ATAR ≥ 90 (%)	ATAR ≥ 85 (%)
Selective	92.6	5.7	51.2	69.8	81.1
Non-Selective					
Sat Exam (NS)	83.9	8.6	23.2	39.4	52.4
Selective Potential (SP)	71.1	19.3	7.5	15.5	23.4

For example, for students without missing ATARS (first column), the mean ATAR value for selective students is 92.6 as compared with 83.9 for NS students, and 71.1 for SP students. The differences in the derived achievement outcomes were more marked, with only 7.5% of general population non-selective students attaining an ATAR >= 95, whereas around half of the selective school students did so.

Note the percentage of missing ATARs (second column) is fairly similar between the selective and non-selective students who sat the entrance exam, at 5.7% and 8.6%, respectively. Students have missing ATAR values for various reasons, including: completing the qualification the following year, leaving high school early, moving interstate, or completing alternative qualifications, such as the International Baccalaureate. There is potential bias to the extent that there is differential selection into missing ATAR values between the selective and non-selective students, but the magnitude of the bias is relatively small with a difference of 2.8 ppt, when compared with the size of the statistically significant matching estimates from the results (section 4.4.5, later), which are over 6.0 ppt.

[94] In the analyses we assess the girls and boys separately. The achievement outcomes for those samples are included in Appendix A.4, in Table 4.11 (A).

4.4.3. Methodology

We apply the same propensity score matching methodology as from Chapter 3, where comparisons are made between selective and non-selective students based on their likelihood of treatment (selective school attendance). Formally, the propensity score is defined by Rosenbaum and Rubin (1983) as the probability of treatment given **x**, the function p(**x**):

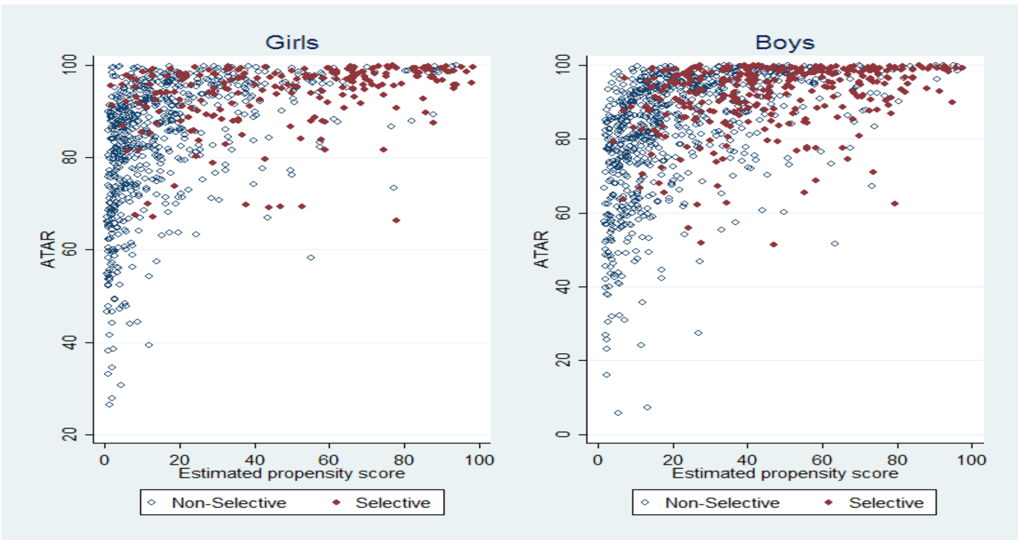
$$p(\mathbf{x}) = \Pr(D = 1|\mathbf{x})$$

where D = {0, 1}, and **x** are the pre-treatment characteristics (Becker and Ichino, 2002). Propensity score matching requires the *ignorability of treatment* assumption, such that treatment and outcomes are independent, conditional on **x.**

To satisfy the balancing requirement that selection is independent of covariates, we follow Becker and Ichino (2002) and test for various specifications that the characteristics of treatment and control groups are not different from each other within a given number of subgroups. Refer to Chapter 3: Methodology section 3.6 for further details of the propensity score matching and the balancing hypothesis that selection is independent of covariates.

Scatterplots of joint ATAR and propensity scores for girls and boys separately appear in Figure 4.4 below. Selective students are indicated by red dots, while non-selective students are indicated by hollow blue circles. The selective dots are clustered above ATARs of 80, with less variance than the non-selective dots, which have a large representation below 20 in propensity score, with ATARs across the whole range of 20 to a 100.

Figure 4.4: Scatterplot of ATARs and Propensity Scores by Sex & Selective School Attendance^



^ From selection model specification in Table 4.5.

4.4.3.1. Prior Achievement

To measure prior achievement in the matching approach, we adopt Year 7 NAPLAN over the alternatives of Year 9 NAPLAN, or the entrance exam, due to both the timing and interpretation of the achievement tests. The Year 7 NAPLAN tests precede the commencement of the selective schools by two years, and precede the admissions process by one year, meaning that the tests are likely not to reflect any influence from attending the selective schools, thus making it suitable as the basis for comparison of later achievement outcomes between selective and non-selective students.

By the same reasoning, we rule out Year 9 NAPLAN as students sit the tests after having attended the selective schools for several months, which could bias the estimates from the immediate influence of selective schools on the achievement of their students. There would be a downwards bias from an expected positive selective school effect.

Although like Year 7 NAPLAN the entrance exam is sat by applicants before the selective schools start, when applicants are in Year 8, the entrance exam cannot be used as prior achievement for the matching comparisons because the stronger applicants generally accept their offers to the selective schools. This results in an uneven proportion of non-selective and selective students at given levels of performance in the entrance exam, which also means that the balancing hypothesis cannot be satisfied due to differences in characteristics between groups for the same probability of selective school attendance.

The different achievement tests, however, pose the question of how to interpret what they measure. Our approach is informed by Sternberg (2007), who suggested that achievement tests are differently qualitatively due to their purpose. The NAPLAN tests, as nationally standardized tests, appear to emphasize accumulated curriculum knowledge due to their accountability function of benchmarking student and school performances. In contrast, the entrance exam assessment is designed to differentiate between high intellectual ability students, appearing closer in form to intelligence tests which have been described as measuring achievement that individuals should have accomplished several years earlier (Sternberg 2007, citing Anastasi and Urbina, 1979).

We secondly distinguish between the assessments by the incentives that they offer to the students, consistent with Duckworth (2016), who defined achievement as the output of skill and effort, where skill itself is a function of effort and talent. This is a departure from the conventional economic interpretations of education, as a cumulative process in which prior achievement is an input (Todd and Wolpin, 2003), or as a product of cognitive and noncognitive skills (Heckman and Rubinstein 2001, Cunha and Heckman 2008), both of which do not account for incentives in test assessments.

When considered by the incentives offered by the tests, NAPLAN is a low stakes outcome for most students such that it is likely that students exert much less effort on these tests, which allows for a clearer interpretation of the matching estimates. The achievement gain produced by the matching approach is a fair comparison in that the university entrance results are a high stakes outcome and the Year 7 NAPLAN are a low stakes outcome for both selective and non-selective students. The qualifier to the interpretation is that some (potentially large) part of the expected gain likely reflects positive unobservables in the selective students.

In comparison, the potentially large incentive of attending a selective school suggests that there is variable effort in taking the entrance exam by applicants to selective schools, conditional on their motivations[95]. Using the entrance exam for prior achievement, if it had satisfied the balancing requirement of the matching approach, consequently suffers from the likely comparison of students who were trying against some of those who were not, leading to a negatively biased selective school effect.

The interpretation of the selective school effect (as the value-added) changes from using the entrance exam in place of Year 7 NAPLAN as prior achievement. Due to variable effort from applicants, a large part of the selective school effect would be more accurately described as the difference in remaining improvement in achievement between selective and non-selective students, with a small part of the effect being attributable to the difference in schools. In other words, by making comparisons based on the entrance exam, the difference in achievement gains between selective and non-selective students would already exclude the gain in academic performance that occurred before the applicants sat the entrance exam.

4.4.4. The Selection Model

The earlier descriptive statistics showed that students at selective schools were more likely to be from a foreign language background, have higher prior achievement scores, and were more likely to be from a higher socio-economic background. Modelling selection into treatment, we estimate a logistic regression on selective school attendance with standardized Numeracy and Reading included as linear terms, language background as an indicator variable, and socio-economic background as a continuous variable, derived from parental education and

[95] We explore reports of additional effort and resources on the part of selective school applicants in Educational Aspiration (section 4.6.4).

occupation. The average marginal effects for these variables are presented in Table 4.5 for each of the exam and selective potential samples. The results are presented separately for boys and girls.

Table 4.5: Logistic Regressions of Selection into Treatment by Sex

	Selective Potential		Sat Exam	
	Boys	Girls	Boys	Girls
	(1)	(2)	(3)	(4)
Std. Numeracy	0.0200***	0.0186***	0.1531***	0.1555***
	(0.0016)	(0.0014)	(0.0141)	(0.0135)
Std. Reading	0.0044*	0.0073***	0.1253***	0.1269***
	(0.0020)	(0.0015)	(0.0183)	(0.0163)
LBOTE	0.0452***	0.0346***	0.0226	0.0754*
	(0.0034)	(0.0030)	(0.0320)	(0.0308)
Std. SES	-0.0022	0.0063*	0.0049	0.0206
	(0.0033)	(0.0026)	(0.0131)	(0.0117)
Government			0.0164	
			(0.0264)	
SES x Score categories	Y	Y	N	N
Number of blocks	14	12	6	8
Tjur	0.171	0.222	0.221	0.324
N				
Selective	328	224	328	224
Non-Selective	11,673	12,115	702	675
Total	12,001	12,339	1030	899

* p < 0.05, ** p < 0.01, *** p < 0.001.

Note: tests scores are from Year 7 NAPLAN and the samples are from a different from a cohort than from the data sample in Chapter 3. Interactions between quartiles of SES and three categories of Year 7 NAPLAN were included in the selective potential samples to satisfy the balancing requirement.

Recall the differences in prior achievement and background characteristics, presented in Table 4.2 and Table 4.3. The selective school students were different from non-selective students with selective potential or who sat the exam in terms of their prior achievement, but they were not different from non-selective students who sat the exam in terms of their language background or SES. In regressions distinguishing selective school students against the two groups, we should expect achievement, language background and SES to distinguish those at selective schools compared to the selective potential students, while only achievement should distinguish selective school students from non-selective students who sat the exam.

This is largely what we see in Table 4.5. Year 7 Numeracy and Reading were statistically significant for both samples, but the effect sizes were much larger for the exam sample (attending a selective school is a much more common event among the group who sat the test compared to those with selective potential: 27% compared to 2%). For the boys in the exam sample, Numeracy had a stronger effect than Reading on selection into treatment (selective attendance); a one standard deviation increase in standardized Numeracy translates to a 15.2 ppt increase in the probability of attendance, while the equivalent increase for standardized Reading is 11.8 ppt.

When we move from the selective potential sample to the exam sample, the ability of language background to distinguish selective school students from others largely disappears for boys, decreasing from 4.5 ppt to 2.3 ppt, but increases in magnitude for girls, increasing from 3.5 ppt to 7.5 ppt. Language background is not significant for boys in the exam sample, but remains significant for girls at the 5% level.

Socio-economic background was mostly unimportant in size and statistical significance, although having higher values in socio-economic background marginally improves the likelihood of attendance for girls in the selective potential sample; a one standard deviation increase translates to less than 1% increase in probability.

The differences in socio-economic background between selective and non-selective students must largely be picked up via the relationship between socio-economic status and prior achievement in the logit regression equation; to satisfy the balancing requirement, interactions between quartiles of socio-economic background and three categories of Year 7 NAPLAN were included in the selective potential samples. A Year 7 government sector indicator was also included for the samples for the boys in the exam sample; the inclusion of sector did not satisfy the balancing tests[96] for the other samples. The Tjur goodness-of-fit values, which are differences in predicted probability between treatment and control, were high for the exam samples, and also indicated greater goodness of fit for the girls' specifications, reaching 0.32 for the exam sample.

[96] Refer to Chapter 3, sections 3.7.2.2 and 3.6.1.1, for descriptions of the balancing hypothesis and balancing tests.

4.4.5. Results

This section presents the main results from the propensity score matching for the girls and boys, from the pooled sample of included selective schools. These results relate to the exam sample, as we found that the non-selective students who sat the exam most resembled students attending selective schools (in the Descriptive Statistics, section 4.4.2).

Figure 4.5: Lowess ATAR and Density of Propensity Scores by Selective and Sex

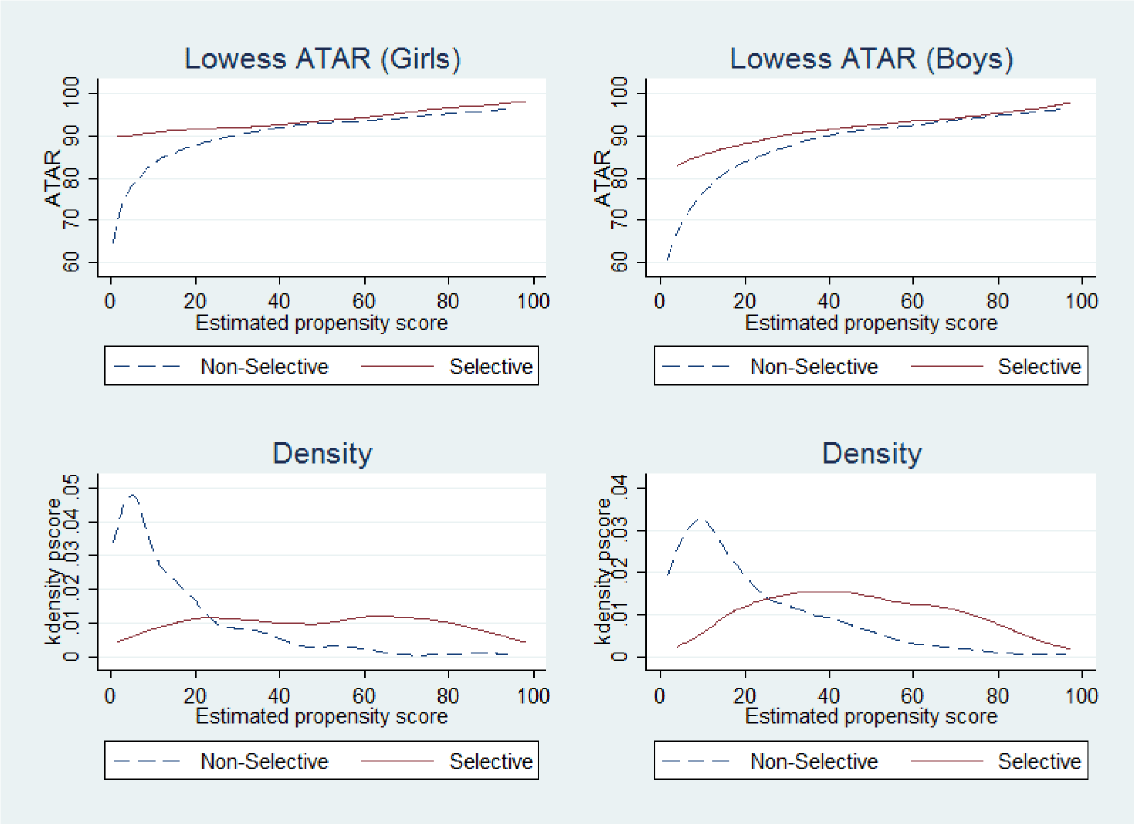


Figure 4.5 presents lowess plots of ATAR ranks against the propensity scores, for girls and boys depending on their selective school attendance status in the upper panels, and the densities of the propensity scores from the selection model in section 4.4.4, the Selection Model, in the lower panels. Lowess, derived from ‘locally weighted scatterplot smoother’, is a smoothing technique applied to scatterplots to help visualise the relationship between two variables, and is also a non-parametric regression model (Fox and Weisberg, 2010); Locally weighted regressions are applied within bandwidths over the range of values to produce polynomial fits of the data, where the local weighting refers to horizontal distance of sample observations from the observation for which the fit is being estimated (Cleveland, 1979).

The graphs in the top panels provide guidance in identifying which selective students are likely to outperform the non-selective students; it shows that the non-selective students with lower propensity scores, those below 40 percent, have lower average ATAR ranks than selective students with the same propensity scores. Above that point, there is little difference in localized average ATAR ranks of the two groups for either sex. In other words, it appears that the students who are least likely to attend selective schools appear to gain the most. These are the students who get into the schools but whose prior achievement is lower than their peers, as is implied by the selection model.

The densities indicate that the distribution for selective school students is relatively uniform, while that for the group who do not attend is quite bunched at low values of the propensity score. That is because the typical prior achievement levels of the group who did not attend selective schools are lower than those who do. The panels on the left correspond to the girls and the panels on the right are for the boys.

The average treatment effects for boys and for girls are detailed in Table 4.6 and Table 4.7, respectively. Estimates from each of the nearest neighbour, kernel and stratification matching methods are presented along the rows. Along the columns are the Year 12 outcomes and also frequencies of non-selective (“NS”), selective (“S”) and total students used in each method. The sample frequencies are included in Table 4.12 (A), Appendix A.4.

Table 4.6 shows increases in the success rates from attending selective schools were significant for boys, with the most consistent results for ATAR and ATAR >= 85; selective students benefited by 1.6 to 2.3 ATAR percentile points depending on matching type, and were more likely to attain ATAR >= 85 by between 5.6 ppt and 9.4 ppt.

Table 4.6: Propensity Score Analysis, Boys

	Outcomes			
	ATAR	ATAR ≥ 95	ATAR ≥ 90	ATAR ≥ 85
Nearest Neighbour Matching	2.281* (1.091)	0.116* (.052)	0.088* (.052)	0.094* (.045)
Kernel Matching (bin width= 0.06)	1.944** (.699 [b])	0.051 (.042 [b])	0.045 (.029 [b])	0.065* (.034 [b])
Stratification Matching	1.647* (.718 [b])	0.041 (.039 [b])	0.040 (.036 [b])	0.056* (.03 [b])
Expected Success (Non-Selective)	83.6	0.325	0.520	0.638

* p < 0.05, ** p < 0.01, *** p < 0.001. Standard errors indicated in parentheses.

[b] indicates bootstrapped standard errors with 50 replications.

Among the different propensity score matching methods, nearest neighbour matching produced the largest effects, and its estimates for all achievement outcomes were statistically significant. The nearest neighbour matching was followed by kernel and stratification matching, with increasingly smaller estimates, and non-significant results for ATAR >= 95 and ATAR >= 90. However, the latter matching types of kernel and stratification matching methods are potentially more reliable as they utilise a greater number of observations in estimation than in nearest neighbour matching.

Table 4.7: Propensity Score Analysis, Girls

	Outcomes			
	ATAR	ATAR ≥ 95	ATAR ≥ 90	ATAR ≥ 85
Nearest Neighbour Matching	1.021 (1.369)	-0.040 (.072)	0.080 (.065)	0.058 (.057)
Kernel Matching (bin width= 0.06)	2.202** (.931 [b])	0.056 (.05 [b])	0.135** (.045 [b])	0.091** (.037 [b])
Stratification Matching	2.065** (.774)	0.048 (.044)	0.132*** (.034)	0.090** (.035)
Expected Success (Non-Selective)	84.3	0.319	0.487	0.640

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

[b] indicates bootstrapped standard errors with 50 replications. Without sector in selection equation.

On the whole, the results for girls (Table 4.7) were stronger than those for boys. None of the estimates from the nearest neighbour matching were statistically significant but this could be because there are fewer control observations. From both kernel and stratification matching, the propensity score matching produced effects of around 9 ppt in attaining ATAR >= 85, and effects of around 13 ppt in attaining ATAR >= 90; girls attending selective schools had higher ATARs by 2.1 to 2.2 percentile points.

In summary, the results from propensity score matching showed that the girls increased their likelihood of attaining each of the Year 12 outcomes from attending selective schools, with the exception of ATAR >= 95. The effect sizes were largest for ATAR ≥ 90, with estimates suggesting an increase in likelihood of around 13 ppt. For context, the expected success rates for ATAR ≥ 90 from probit regressions was 48.7% for the base, that of non-selective students.

These results are positive for the girls in that there is evidence of gains across three of four Year 12 outcomes, and gains are largest for the highest outcome of ATAR ≥ 90, meaning selective students were more likely to be able to attend courses and universities with greater demand. The results for the boys were not as strong as that for the girls, but there was evidence for boys improving their ATAR ≥ 85 and ATARs in general from attending selective schools, which would still allow students to enter a wide range of courses; the matching produced statistically significant estimates of between 6 ppt and 9 ppt for ATAR >= 85, off a base expected success rate of 63.8% for non-selective students.[97]

For both sexes, the differences are larger for the more attainable Year 12 outcomes of ATAR ≥ 90 and ATAR ≥ 85 (except for the nearest neighbour matching estimates for ATAR ≥ 90 for boys), suggesting that the gains are larger for those with lower propensity scores, which roughly corresponds to those with lower achievement levels who were admitted to the selective schools. [98]

4.5. Regression Discontinuity

We discuss the methodology and review the suitability of the data for the regression discontinuity approach, describing the samples, the mean student characteristics at the cut-offs and the relationship between bandwidth and sample sizes, before presenting the results.

4.5.1. Methodology

The regression discontinuity approach relies on the assignment of treatment from a rating variable that is continuous and measured before treatment. At the same time, treatment is randomly determined by a cut-point or cut-off, and because treatment is random the intuition of the approach is that the characteristics of participants just below and above the cut-point are the same aside from the treatment. The regression discontinuity approach consequently

[97] As would be expected from an interpretation that applying to the selective schools reflects educational aspiration, much larger effects are estimated if the comparisons are made against the selective potential sample via matching. Simple unweighted regression estimation on the same sample also points to positive selective school effects of about the magnitude estimated via matching (which does not account for performance on the entrance exam).

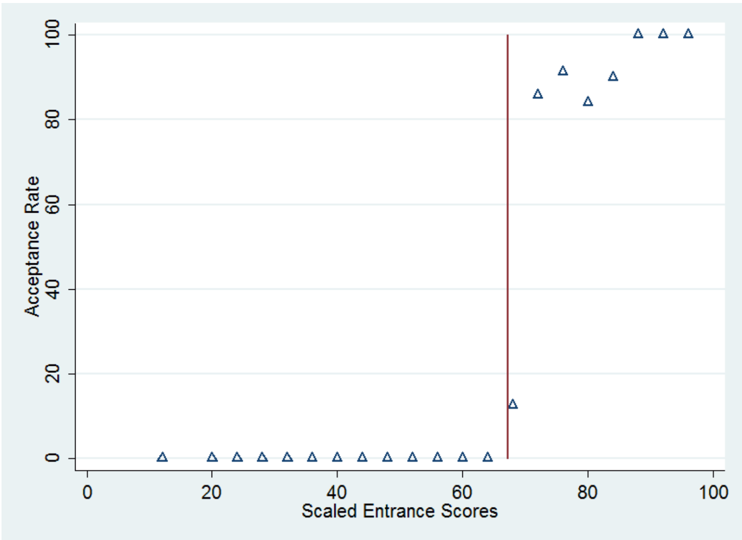
[98] Regression estimates also pointed to positive effects from attendance of around 3 ATAR points for those students who gained entry via the disadvantaged background intake, consistent with the estimated effects being larger for those least likely to get into the selective schools. Students admitted through the discretionary intake were found to receive similar effects to other students in the selective schools.

produces a local treatment effect as comparisons are made on marginal students; students who just get in compared with students who just miss out.

In this institutional context, the entrance exam results determine the treatment of attendance at selective schools, and the cut-offs are determined arbitrarily by the size of the schools but are nevertheless high. The cut-offs are the minimum scores obtained by students admitted through the main intake, which are influenced by the number of positions at each school.

Figure 4.6 shows the acceptance rates over intervals of scaled entrance exam scores for applicants who listed one of the selective schools as their first preference, with the cut-off indicated by the red vertical line; students admitted via the discretionary or disadvantaged background intakes are excluded. The acceptance rates are similar to that of the other selective schools, which have different cut-offs.

Figure 4.6: Acceptance Rates and Entrance Exam Scores



The cut-off is indicated by the red vertical line, with a minimum scaled score at the selective school of 67.3. Rates calculated over intervals of 4 points.

The variability in acceptance rates to the right of the threshold suggests the existence of a fuzzy discontinuity rather than a strict discontinuity, where all the offers are accepted by students. Above the cut-off, to the right of the vertical red line, less than 20% of students who attain sufficiently high results on the entrance exam, and have ranked the school as their first preference, decline the offers. Note that the appearance of the acceptance rates in Figure 4.6 is due to the size of the intervals (of 4 points) used to show how student characteristics and acceptance rates vary with entrance scores. The observation below an acceptance rate of 20, but not zero, is explained by students falling within the interval that includes the cut-off.

In the analyses, the range of points around the cut-off used to determine the data sample, defined as the “bandwidth”, is varied to show that the RD estimates are not sensitive to this bandwidth. Another variation to the approach is to apply different weightings associated with distance from the cut-offs.

4.5.1.1. Non-compliers and Instrumental Variables

To account for the students who decline offers, defined as *non-compliers* in the literature (from Jacob et al. 2012), we apply two-stage least squares (2SLS) using the rating variable, the entrance exam scores (r_{i}). This divides the difference in outcome from the discontinuity by the difference in treatment from the discontinuity. The specification of the 2SLS is:

$$S_i = \alpha_1 + \gamma D_i + f_1(r_i) + \varepsilon_i$$

$$Y_i = \alpha_2 + \beta S_i + f_2(r_i) + \mu_i$$

S is selective school attendance, *D* is an indicator variable above the cut-point, while *f* is the different functional form for stages 1 and 2. The remaining variables are constants (α) and error terms (ε and μ) that are assumed to be identically and independently distributed.

For continuous outcome variables we apply linear regression to both stages, and for binary outcome variables (ATAR >= 95, ATAR>= 90 and ATAR>= 85) we apply linear and probit regressions for stages 1 and 2, respectively. For all outcomes, we estimate cubic and quadratic functional forms of the rating variable (r) in both stages, and report the cubic estimates as it has no bearing on the results. Similarly, adding an interaction between the indicator variable and the rating variable, $\mu D_i * f_1(r_i)$, in the first stage, or replacing the first stage with probit regressions, did not change the results. Estimates were slightly more positive in the former, and slightly more negative in the latter, but both remained statistically insignificant.

4.5.2. Regression Discontinuity Samples

Offers were allocated from entrance exam results to applicants in order of highest demand to lowest demand selective school, with the cut-offs for each of the schools differing according to the demand for the respective schools and the size of the schools. Excluding schools with the highest demand, for which all admitted students had the schools as their first preference, precedence was given to entrance exam score over school preferences, meaning that some applicants accepted second-preference offers.

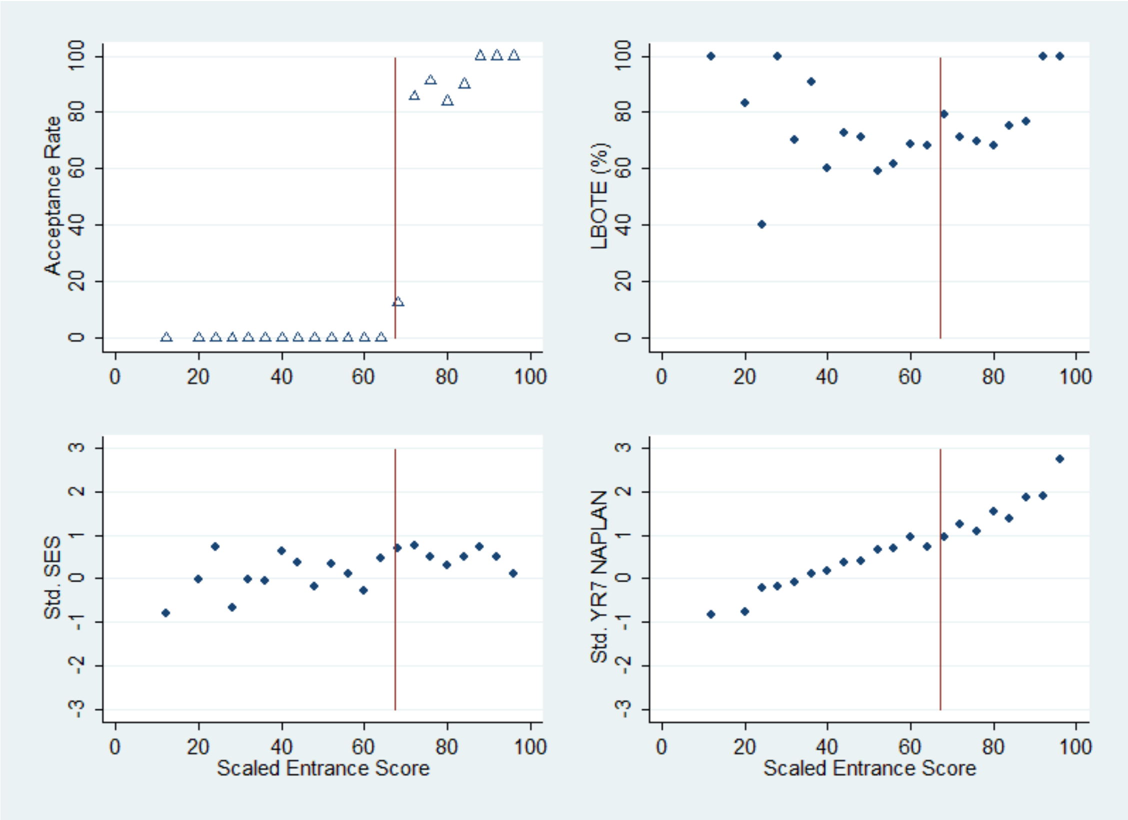
For the analyses, we exclude students who accepted offers for selective schools that were their second preferences because the acceptance rates for students who received offers for second preferences were much lower. The lower acceptance rates imply that there are unobserved influencing factors or characteristics that differ between applicants who accepted or rejected their second preference offers, which is likely to lead to biased estimates from comparing the two groups. Because preferences were much weaker for certain schools, this reduced the sample size for the pooled analyses.

As with the matching approach, we excluded the new selective school from the analyses, due to the small number of applicants expressing the school as a first preference, and also because of the lower rates of acceptances from students receiving offers. Specifically, from the few applicants choosing the school as their first preference, the significant proportion who decline offers to attend the omitted school introduces doubt that the marginal students attending and not attending are essentially the same.

4.5.3. Student Characteristics

The regression discontinuity approach relies on random variation of assignment such that the characteristics of the students are the same in the vicinity of the cut-point. For applicants expressing as their first preference one of the selective schools, Figure 4.7 presents the acceptance rates (from Figure 4.6), and the mean student characteristics within bins of scaled entrance scores for the selective schools: the percentage with foreign language background, the socio-economic index of parental education and occupation, and Year 7 NAPLAN scores.

Figure 4.7: Student Characteristics by Scaled Entrance Scores



The minimum scaled score at the selective school was 67.3. Bin-size of 4 points.

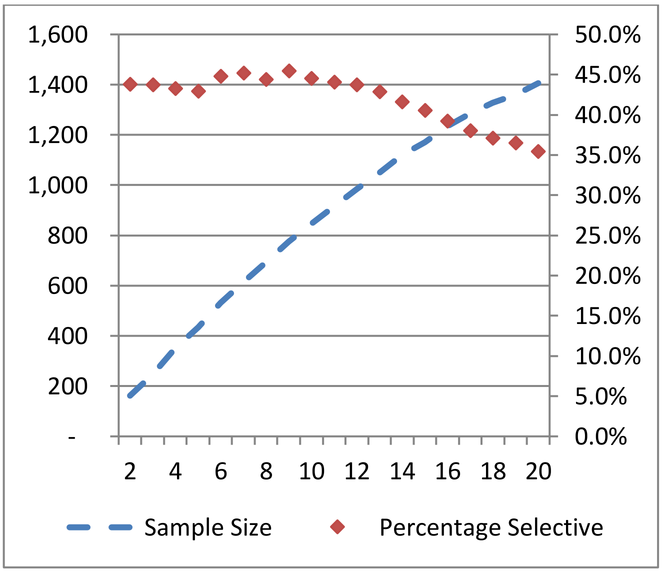
Although slightly noisier in the behaviour of student characteristics against the scaled entrance scores, the example selective school is reflective of all the selective schools in the pooled analyses; the language and socio-economic background and prior achievement of students behave in a continuous manner in the vicinity of the cut-offs, and the visualisations support the assertion that the cut-offs were arbitarily determined and that students are essentially the same on either side of the cut-offs. From the mean student characteristics, it appears that the data is suitable for the regression discontinuity analyses.

4.5.4. Bandwidths and Frequencies

In the comparisons between treatment and control groups, the bandwidth or range chosen may influence the results as this determines the number of observations surrounding the cut-offs on either side of the threshold. A trade-off exists between the increases in the sample size from a larger bandwidth, which is balanced against the weakening of the assumption that the students on either side have the same mean characteristics. This is most apparent, for example, in the trend of increasing Year 7 NAPLAN score with increasing entrance exam (Figure 4.7).

Figure 4.8 below shows how the sample size increases with increasing bandwidth, as measured in entrance scores, for the pooled selective schools sample. For example, the sample size, represented by the dashed blue line, increases from 162 observations to 435 on the left vertical axis, moving from a bandwidth of 2 to 5. The red diamonds, measured on the right vertical axis, indicate the percentage of students attending selective schools within the bandwidth dependent samples; the percentage of selective students hovers around 45% before steadily decreasing with increasing bandwidth.

Figure 4.8: Frequencies by Bandwidth and Percentage Selective



As a point of reference, when the bandwidth approaches 20 (entrance exam) points, the sample includes almost all the students who attend two of the selective schools. The sharp increases in acceptance rates (from Figure 4.6) fall within a bandwidth of around 5 points, while a bandwidth over 10 points appears not to be very meaningful from the perspective of identifying the marginal students; those marginally attending and those just missing the cut-off. The precise bandwidth chosen for the analyses can be an important decision but, previewing the results, in our study the results are not sensitive to the bandwidth chosen.

4.5.5. Results

The estimates from the fuzzy regression discontinuity analyses of the pooled selective schools sample were not statistically significant. These estimates are presented in a visual format in Figures 4.9 to 4.15 to show how they vary with bandwidth choice, on the horizontal axis; the blue dots are the estimates (left vertical axis) and the dashed purple lines represent the sample size (right vertical axis). Note that the standard errors represented by the bars narrow with increasing bandwidth as sample sizes increase. For statistically significant results, the tops or bottoms of 95% confidence intervals, which are almost twice (1.96) the size of the error bars shown, would need to be completely below or above zero respectively. These estimates are also included in Appendix A.5 in Table 4.13 (A) for reference.

For the ATAR outcome (Figure 4.9), the estimates are positive and large at less 4.0 and 3.2 ppt at a bandwidth of 2 and 3 entrance exam points, but the estimates are not statistically significant due to the magnitude of the error bars and small sample sizes. With increasing bandwidth, between 4 and 14 points, estimates of the selective school effect on ATAR fall below zero, indicating that the overall effect is closer to zero. Eventually, the estimates move upwards with increasing bandwidth, to 2 ATAR points at a bandwidth of 16, as is expected with an increasing differential between levels of prior achievement in the marginal selective and marginal non-selective students (see positive relationship between Year 7 NAPLAN and entrance scores in Figure 4.7).

Figure 4.9: ATAR Estimates by Bandwidth

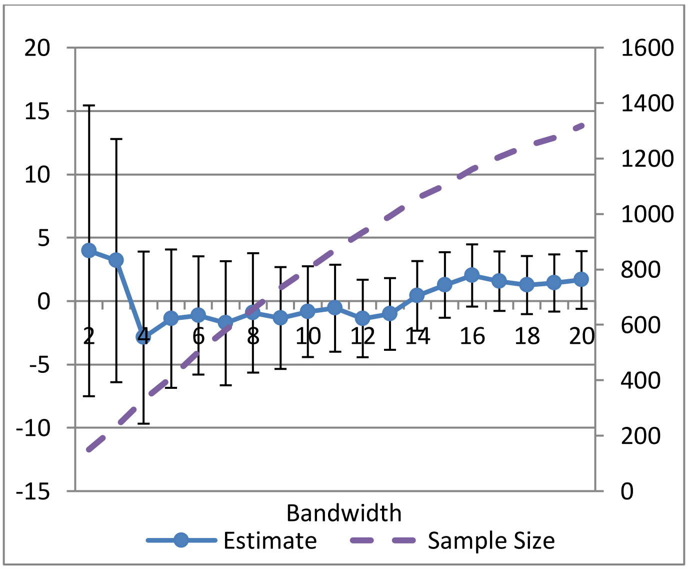


Figure 4.10: ATAR ≥ 95 by b/width

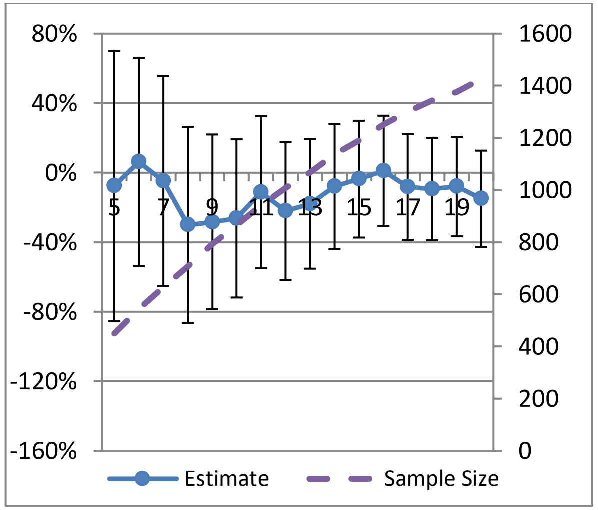


Figure 4.11: ATAR ≥ 90 by b/width

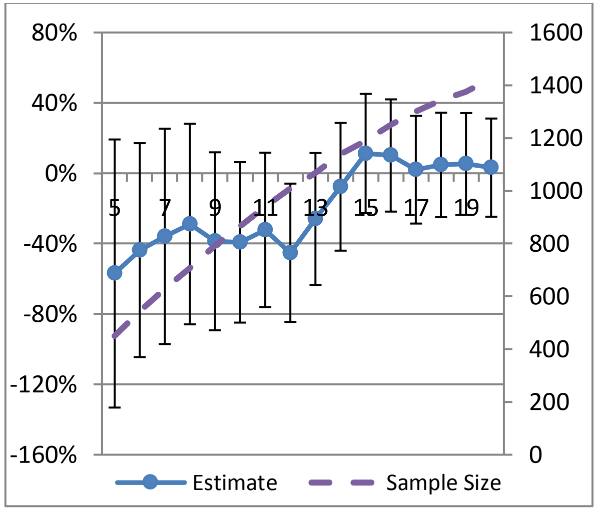


Figure 4.12: ATAR ≥ 85 by b/width

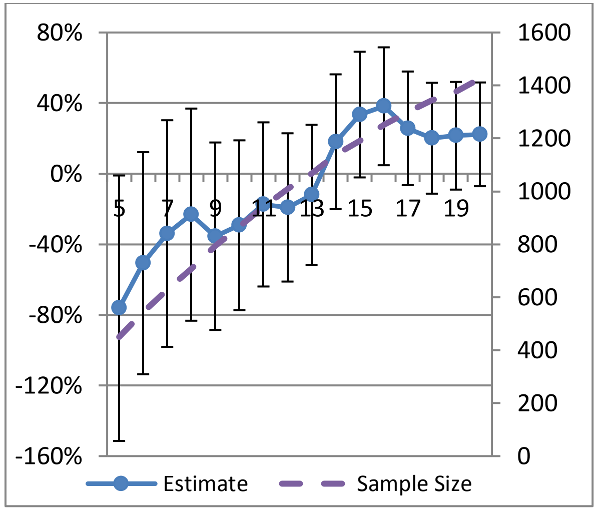


Figure 4.13: Mathematics by b/width

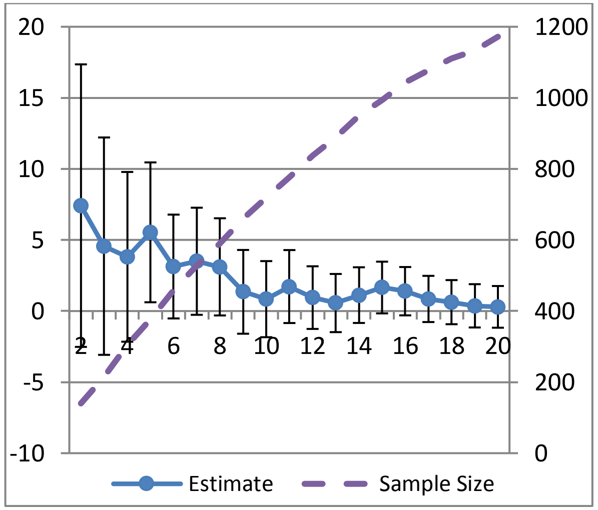
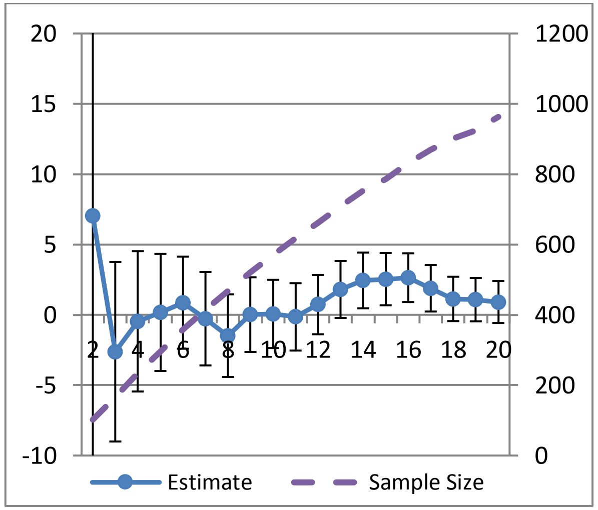
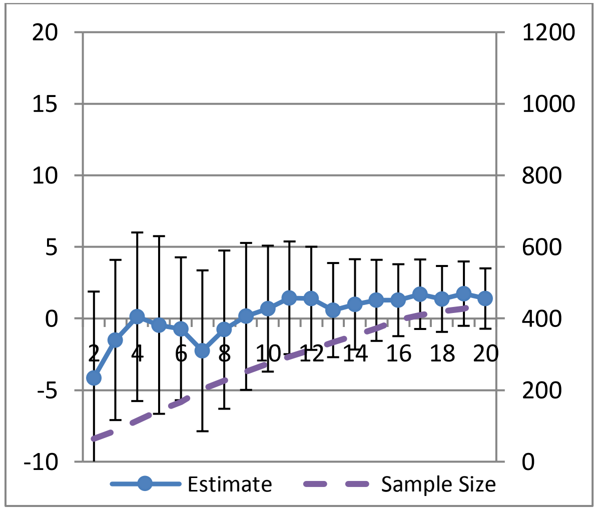


Figure 4.14: English by b/width ^



^ Truncated error bar for first estimate (bandwidth = 2) ranges from -11.4 to 25.5.

Figure 4.15: Physics by b/width ^



^ Truncated error bar for first estimate (bandwidth = 2) ranges from -10.2 to 1.9.

The estimates for the binary outcomes of ATAR ≥ 95, ATAR ≥ 90 and ATAR ≥ 85 are mostly negative and large in magnitude but not statistically significant with very large error bars (Figure 4.10 to Figure 4.12). The estimates for ATAR >= 95 (Figure 4.10) for the bandwidth range between 5 and 7 are more neutral, ranging between -7.7 ppt and 6.2 ppt, and are the strongest indication that the selective school effect is not negative, noting still the large standard errors. Note the same vertical and horizonal scales across the binary outcomes for Figure 4.10 to Figure 4.12, and across continuous subject outcomes for Figure 4.13 to Figure 4.15 (discussed next).

The standard errors for the binary outcomes (Figure 4.10, Figure 4.11, and Figure 4.12) are much larger than for continuous outcomes, such as ATAR (Figure 4.9), due to probit regressions in the second stage of the instrumental variables which are less efficient than the linear regressions. This is also the reason, in combination with the small samples, for the lack of estimates for the binary outcomes for bandwidths less than 5 points.

The results for Mathematics, English, and Physics, as continuous outcomes, are shown in Figure 4.13, Figure 4.14, and Figure 4.15 for reference. These Year 12 subjects contribute to students' final university entrance results[99], and were taken by the greatest number of selective students. Note subject scores have a maximum of 50. Estimates across the three subjects were also not statistically significant, and the pattern in estimates for English and Physics were similar to those relating to ATAR and the ATAR derived outcomes, being mostly negative at smaller bandwidths before trending upwards at larger bandwidths. The pattern in estimates for Mathematics (Figure 4.13) was different with mostly positive estimates at smaller bandwidths, including a maximum estimate of 7.4 at bandwidth 2, before fluctuating around zero from bandwidths above 9.

[99] This is intentionally vague in order to retain anonymity of the state and identities of the selective schools.

4.6. Discussion

4.6.1. Matching

The results from the matching approach (from section 4.4.5) suggested there were small effects of selective school attendance on achievement at the end of high school. The analyses produced estimates of 2 percentile points in ATAR, and estimates of between 6 ppt and 13 ppt increases in probability of attaining ATARs in the top 10% and 15% of students. The estimates are of significant magnitude for high-achieving students, and suggest that the schools provide a floor for Year 12 achievement, but they are overstated as they do not account for positive unobservables, such as is reflected in the results from the entrance exam (discussed further below).

Two points in ATAR at the top of the distribution may be the difference in whether students get into their preferred course at their desired higher education institution, so even effects of this size may have substantial implications for individuals. Estimates of around 2 ATAR points are also similar to value-added estimates of New South Wales selective schools for Year 12 outcomes (Lu and Rickard 2014). Between Year 9 NAPLAN and measured Year 12 outcomes, the selective school effect was just 0.04 of a standard deviation. This translates into around 0.7-0.8 of one ATAR point (0.04 x 28.9 – the SD of the uniform ATAR distribution).

An advantage of the matching approach is that it helps with conceptualizing the selection issues, of which there are two forms; first, students express a preference for attending the schools, and second, they are sorted into the schools on the basis of their performances on an academic entrance exam. We identified the comparison group from their behaviour of sitting the entrance exam, and excluded students whose achievement on Year 7 NAPLAN was below that of the minimum level at the selective schools. For the second form of selection, we identified selective students from this restricted sample as more likely having higher levels of earlier achievement, and a foreign language background for girls (Selection Model, section 4.4.4).

A limitation from the matching approach was that it was not able to account for both prior achievement in Year 7 NAPLAN and achievement on the entrance exam due to the balancing requirement (see Methodology, section 4.4.3); there was not equal representation between selective and non-selective groups, against the predicted probability of attendance, because the former outperformed the latter on the entrance exam for similar levels of prior achievement. Including Year 7 NAPLAN in the selection model does, however, increase the ease of interpretation in that Year 7 NAPLAN provides the same incentive to both selective and non-selective students as a low stakes outcome.[100]

We find matching estimates of 2 percentile points in ATAR from attending the selective schools, but they may be overstated for partially reflecting pre-existing differences between the selective and non-selective students, such as additional motivation or latent academic ability. That is, the selective students might view admission to the selective schools as a high stakes outcome, from their outperformance over non-selective students with similar levels of prior achievement on the entrance exam. We investigate this assertion further, in Educational Aspiration (section 4.6.4); asking whether applicants have different levels of motivation.

4.6.2. Regression Discontinuity

The results from the regression discontinuity analyses were not statistically significant, which is consistent with the 2014 US studies (Adulkadiroglu et al. and Dobbie and Fryer), though in our case this could be attributed to the small sample of marginal students taken from one cohort. In isolation the regression discontinuity results are inconclusive, but there is information from both the negative direction of the estimates for the outcomes of ATAR >= 85 and ATAR >= 90.

Although regression discontinuity is a convincing method due to its resolution of the unobservables problem, there are differences in the characteristics of students for which the local effect is relevant as compared with the characteristics of selective students in general. Most notably, the marginal selective students are among the academically weakest students at the selective schools, giving rise to the possibility of a decrease in their academic performance arising from diminished self-concept. Further, these marginal selective students are compared with students who just missed out on admission that are likely to be academically stronger students in the schools they attend, which could also help explain the absence of positive effects that we found in the matching analyses.

Another possible explanation for the lack of effect is that the high levels of educational aspiration of both marginal selective and marginal non-selective students means that they are less influenced by their school environment[101]. An interpretation of applicants to selective

[100] Replacing Year 7 NAPLAN with entrance exam in the selection model, potentially leads to unfair comparisons between applicants with differing levels of motivation. In addition, students who attain sufficiently high scores and decline offers appear to have less incentive to try on the exam because they are less invested in the outcome (admission to the selective schools).

[101] Note that we think that their educational aspiration is also relevant to the matching results, such that they are positively biased for not having accounted for the additional motivation of selective students over non-selective students on average, rather than locally just for marginal students, as is the case here.

schools as being driven and conscientious is supported by anecdotal evidence, including media reports of the coaching of applicants, and is also consistent with a large survey of applicants and their parents (Braithwaite and Kensell, 1992), which indicated that the most important reasons for applying were related to educational aspiration.

In a similar application to this study, Angrist and Rokkanen (2015) extend the RD approach using matching to estimate achievement effects away from the cut-offs for the Boston selective schools; they replace the entrance exam with propensity scores of demographic control variables and standardized test scores of students prior to sitting the entrance exam. Consistent with our matching and RD results, they find similar achievement effects for students away from the cut-offs.

Overall, the local effect that applies to marginal students is non-existent or at least weaker than the positive effect from the matching analyses**,** and can be attributable partially to the impact of relative status in the comparisons of marginal students, but mostly from the similarity in the characteristics of successful and unsuccessful applicants to selective schools. This refers to their similar foreign language background, their academic performance, and the reportedly high levels of both educational aspiration and coaching in preparation for the entrance exam.

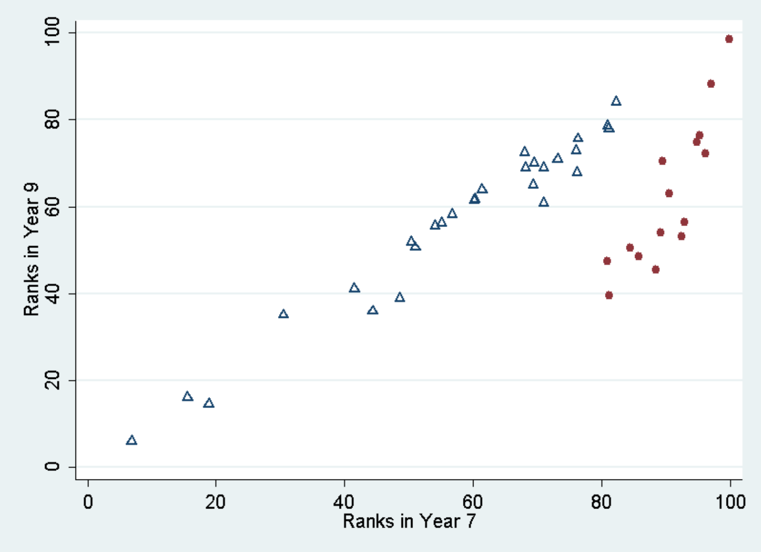
4.6.3. Rank Effects & Self-Concept

The relevance of rank effects to the selective schools stems from the potentially large changes in academic status that students experience when they enter the selective schools, by virtue of sorting into an academically competitive environment. It is possible that a negative effect on achievement from diminished academic self-concept, i.e. students' views of their own ability as informed by others, could offset potential benefits to later achievement anticipated by students from attending the selective schools. The importance of relative status in academic achievement was first documented by Marsh and Parker (1984), who found an inverse relationship between the achievement levels of schools and students' academic self-concept[102].

[102] More recently, Murphy and Weinhardt (2014) documented the influence of "rank effects" on later achievement using within-school ranks to reflect relative academic status and implied self-concept (see Chapter 3: Literature Review, section 2.2).

Figure 4.16 shows the extent of the change in relative status implied by changes in ranks between Year 7 and Year 9 from attending one of the selective schools, as an example[103]. It is a scatterplot of achievement ranks in Year 7 and in Year 9 for applicants of one of the selective schools; ranks are measured as within-school ranks of averaged standardized Numeracy and standardized Reading from Year 7. Non-selective and selective students are represented by blue triangles and red dots respectively.

Figure 4.16: Local Ranks in Year 7 against Local Ranks in Year 9 by Selective Attendance^



^ Relates to applicants for one selective school only. Ranks are based on results from Year 7 NAPLAN and discretionary selective students are excluded. Red dots and blue triangles represent selective and non-selective students, respectively, and each point is an aggregation of students over small intervals of Year 7 local ranks. Applicants who decline offers not shown.

There is a clear linear relationship between ranks in Year 7, along the horizontal axis, and ranks in Year 9, along the vertical axis, for the non-selective students. This hypothetical line for non-selective students ends at around 80 percentile points in Year 7 ranks, where observations of the selective students begin, but with a different trajectory. Note the difference in Year 9 ranks between non-selective and selective students near this point, where the Year 7 ranks overlap, which is as large as 40 percentile points. This difference in ranks would be even larger but for the exclusion from the plot of selective students with lower achievement, who were admitted through the secondary discretionary and disadvantage background intakes (for consistency

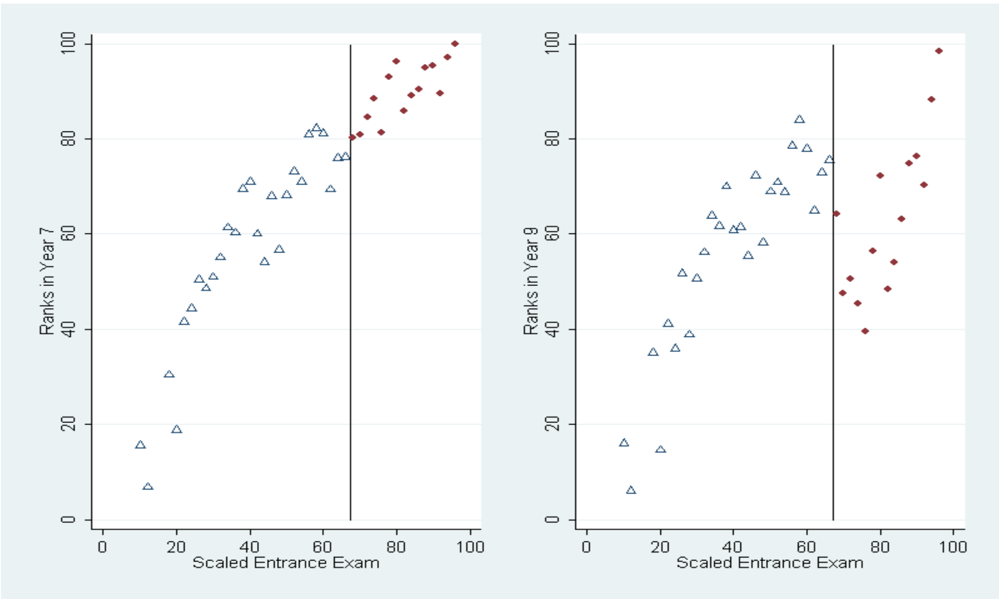
[103] This is the same selective school as was presented in comparing the student characteristics at the entrance exam cut-off for the regression discontinuity analyses (section 4.5.3). See Figure 4.7 which corresponds to Figure 4.17, discussed next.

with the analyses); these secondary intake students explain the absence of observations between 0 and 40 in ranks in Year 9.

The potential influence of rank effects at selective schools is also evident in the regression discontinuity approach, which compares marginal selective students, who are likely to be among the academically weakest at the selective schools, with marginal non-selective students, who are likely to be academically stronger than others in the school they attend.

Figure 4.17 shows again the extent to which there is a change in relative status from attending a selective school, but against performances on the entrance exam. The left panel is a plot of Year 7 ranks and the right panel is a plot of Year 9 ranks, which together show the change that is experienced by marginal selective, compared with marginal non-selective students. Based on their Year 7 ranks, the selective students follow a similar ascending pattern of within-school ranks with increasing entrance exam results. Like in Figure 4.16 before, there is a decrease in ranks in Year 9 for the selective students, with a difference between marginal selective students and marginal non-selective students, which is between 20 and 40 percentile ranks, depending on weighting.

Figure 4.17: Local Ranks in Year 7 and Local Ranks in Year 9 against the Entrance Exam^



^ Ranks in Year 7 are shown in the left panel, and ranks in Year 9 are shown in the right panel. Ranks are based on Year 7 NAPLAN for both year levels. Blue triangles and red dots represent non-selective and selective students respectively, and each observation represents students aggregated over small intervals of the entrance exam, as per Figure 4.16.

An indirect consequence of sorting into academically competitive selective schools is that many students experience a sharp increase in peer quality, which results in downgrades in their own relative academic status, and could negatively affect their later achievement. That is, marginal selective students are likely to be among the academically weakest at the selective schools.

While the offsetting influence of negative rank effects can help explain how the selective school effect might be lower than expected, the overall influence from rank effects for selective students may be less problematic as they are high-achievers and because they enter the schools at a relatively later stage of schooling. Research on academic self-concept and academic selection across schools indicates that the negative effects on self-concept from having high-achievement peers were weaker for high-achieving students (Trautwein et al. 2009), while countries with earlier explicit school-level tracking produced larger negative effects on academic self-concept (Salchegger 2016), which is not the case here.

4.6.4. Educational Aspiration

We sought to ascertain the extent of unobserved educational aspiration in selective school applicants relative to the general population, and between selective and non-selective students, via exploratory analyses and also a synthesis of the relevant research and empirical facts.

An earlier study of the surveyed reasons for applying to selective high schools by students and their parents strongly supports the interpretation that selective students have greater levels of educational aspiration. Braithwaite and Kensell (1992) found that academic emphasis, including the perceived advantage for future careers, and the reputation of the schools were the most important factors for accepting offers to selective high schools; the reputation of the schools referred to perceived social, cultural and sporting reputations as well as academic standing, while reputation effects were much weaker for newly created selective high schools.

The high representation of students from a foreign language background at selective schools suggests that for these students their ambitions are closely related to their immigrant background, especially given the positive association documented between educational attainment and foreign language background[104]. It could be argued that the selection of students in the selective schools parallels the successful positive selection that is designed in immigration policy. E.g. Miller (1999) noted that Australian immigration policy allowed for the positive selection of individuals for adapting to the labour market.

[104] Gemici, et al. (2014) and Homel et al (2012), refer to section 3.2.3, Chapter 3: Literature Review.

Linking the selective schools to historical immigration patterns, a significant proportion of migrants to Australia were of Asian background for the time-period relevant to this case study[105]. It is possible then that the positive associations between educational attainment and foreign language background are inter-related with the well-documented outperformance in academic achievement by students of Asian background (see for example, Jerrim 2015[106], Hsin and Xie 2014[107], and Dandy and Nettelbeck 2002). The research attributes the academic outperformance mostly to educational and occupational aspiration and cultural factors, providing guidance for how we interpret the exploratory analyses presented next. There are indications that the relationship between achievement and geography is more pronounced for the subgroup of students from East and South-East Asia.

Research from the UK supports the interpretation that for many selective students their high levels of education aspiration are closely related to their immigrant backgrounds, and that they can be attributed to cultural attitudes. Abbas (2007) interviewed students and their parents from South Asian backgrounds with reference to the selective schools, noting that South Asian parents are especially encouraging of their children in education. Abbas found that social class was the strongest factor in gaining entry; middle class families were highly motivated and possessed the requisite economic, cultural and social capital, while some working-class parents possessed strong attitudes towards selective education, regardless of resources.

Finally, there is also anecdotal evidence of the additional effort exerted by applicants to selective schools. Private tutoring for preparation for selective entry tests in Australia is endemic, if the popular press is to be believed[108]. Another phenomenon that has been observed within the schools (anecdotally) is that many students who make it into the schools find their academic performances in the school diminished. Both phenomena point to

---

[105] Individuals from Asia (from South-East, North-East, Southern and Central Asia) comprised 55.3% of permanent settlers in 2016, up from 21.1% in 1996, 3401.0 Overseas Arrivals and Departures, Dec 2016, Australia Bureau of Statistics and Asian Immigration, Current Issues Brief 16 1996-97, Parliamentary Library.

[106] Jerrim (2015) documents the outperformance on the Programme for International Student Assessment (PISA) of students from an East-Asian background in Australia, explaining most of the variance in performance with multiple factors relating to subjective norms, instrumental motivation, attitudes to school and future aspirations.

[107] Relying on teacher-evaluations of student behaviours and attitudes, Hsin and Xie (2014) explained the outperformance of Asian-Americans on academic achievement by their exertion of greater academic effort and not to cognitive abilities or socio-demographics.

[108] See for example "Testing times: selective schools and tiger parents", Anna Broinowski, Sydney Morning Herald, 01/24/2015 http://www.smh.com.au/good-weekend/testing-times-selective-schools-and-tiger-parents-20150108-12kecw.html, accessed 28 June 2017. We note that tutoring services for the specific purpose of attaining entry to the selective schools were easily found.

applicants of selective schools of being conscientious, having high educational aspirations, as well as having access to educational and family resources (in terms of time and financially).

4.6.4.1. Exploratory Analyses

We undertake additional exploratory analyses which support the interpretation of selective students having greater levels of educational aspiration than non-selective students who sat the entrance exam. First, analyses of achievement growth showed that selective students outperformed non-selective students on the entrance exam, within deciles of prior achievement (see Appendix A.6). Second, value-added regressions of ATAR showed that the inclusion of the entrance exam improved the predictive power for selective students but not for non-selective students (see Appendix A.7). We interpret this to suggest that selective students treated both ATAR and the entrance exam as high stakes outcomes.

The analyses of achievement growth consisted of calculating differences in relative performance between selective and non-selective groups, after students from the exam sample were assigned into deciles based on Year 7 NAPLAN[109]. The difference in relative performance between selective and non-selective students was largest for the entrance exam, and smaller for Year 9 NAPLAN and generally even smaller for ATAR. The outperformance by selective students was higher at lower deciles of Year 7 NAPLAN, which is to be expected due to the admission mechanism, where offers are conditional on entrance exam performance; differences were as large as 50 percentile ranks for the entrance exam and first decile of Year 7 NAPLAN.

We estimate regressions of ATAR separately for selective and non-selective students, while varying the specification to include the different measures of prior achievement. When Year 7 NAPLAN was replaced by the entrance exam as prior achievement in regressions of ATAR, the R-squares improved by 0.09, from 0.18 to 0.28, for selective students, but were improved only marginally, by 0.02, for non-selective students; only basic student characteristics were included in the regressions.

Due to the incentive structure of the admissions process where offers to the selective schools are tied to performance on the entrance exam, the entrance exam provides an example of when effort appears to be variable among applicants[110]. From this, and the greater predictive

[109] Differences are measured in terms of percentile ranks. This is an approximation to Betebenner's Student Growth Percentiles (2009), developed in Houng and Justman (2013).

[110] See discussion of the incentives provided by assessments in section 4.4.2.1, Prior Achievement.

power of the entrance exam for the selective students of ATAR, we conclude that the entrance exam can be considered a high stakes outcome for selective students in the same manner as the university entrance results.

From the analyses of both achievement growth and the value-added regressions, we interpret much of the difference in results between the entrance exam and Year 7 NAPLAN of selective students, compared with their non-selective counterparts, to reflect greater short-term effort and motivation that is consistent with having greater levels of educational aspiration.

4.6.5. Limitations

We note that the results from our study are subject to certain limitations. First, the study relied on only one cohort of students, and replicating the analyses across several cohorts would assuage concerns relating to the generalisability of the results. Our results are, however, consistent in both magnitude and direction with those of other studies on selective high schools and on related education topics (Chapter 3: Literature Review, section 3.2.1).

A related concern is that our selective school effect is an aggregate of a relatively small set of schools, meaning that our results are influenced by the circumstances of specific schools. As such, we interpret the estimates in its appropriate context of comparing the characteristics of the selective students in relation to their differences from the student population, discussed in Academic Selection (section 4.2.2) and Descriptive Statistics (section 4.4.2); the prior achievement of selective students in this study approach nearly two standard deviations above the student population. We excluded one selective school for which applicants and students were noticeably different from the other selective schools (detailed in section 4.4.1.1 Selective Schools).

As to other data limitations, we do not have information on students with missing ATAR values; as noted in the descriptive statistics of the achievement outcomes (section 4.4.2.3) it is possible that the estimates are subject to some small positive bias from the positive selection of non-selective students into an alternative high school qualification such as the International Baccalaureate.

There are implications from the instruments we use to reflect academic achievement. All of Year 7 and 9 NAPLAN, the entrance exam and the Year 12 assessments (ATAR and ATAR derived outcomes), would appear to suffer from ceiling effects in that they are poor at discriminating achievement for students at the top end of the distribution[111], who are also few in numbers. Although it may be the case that ceiling effects apply in all assessments, this lack of precision applies to both selective and non-selective sets of student, and do not receive a large weighting, meaning that there should be little bias on the estimates.

More importantly, there is a question of interpretation of what the achievement tests measure. We interpreted the achievement tests as reflecting a combination of ability and effort associated with the incentives provided by the test (see Prior Achievement, section 4.4.3.1). For instance, the university entrance results can be thought of as reflecting fully expressed potential in achievement performance in individuals, as a multiplicative of effort and ability, while for applicants who tried their best on the entrance exam, the performance differential between the entrance exam and Year 7 NAPLAN could arguably be described as previously unexpressed latent ability or achievement potential.

The methodological limitation of the research is the existence of unobservable characteristics of applicants to selective schools, and those of students who are admitted to selective schools, which was also incidentally the key challenge of the research. This concern was highlighted by the initial background information that the demand for selective school attendance was high (e.g. from the percentage of successful applicants, roughly 33% in our study), and media reports that applicants have been coached and exert greater effort in preparation for the entrance exam. The strong preferences of applicants are also supported by earlier studies, including the surveyed views of selective students and their parents.

We sought to address the concern of positive selection in applicants by first identifying the sample of students who sat the entrance exam. Additional exploratory analyses, however, suggested that the educational aspirations of applicants who were accepted into the selective schools were much higher than those of the unsuccessful applicants. The positive effects from the matching were consistent with that expected from the use of low stakes standardized test scores, which reflected academic ability but did not account for educational aspiration, so we anticipate a large part of the effect on ATAR to be positive selection in the form of increased effort.

While applying the fuzzy regression discontinuity method with the entrance exam appeared to account for positive unobservables (i.e. access to resources and increased effort on the part of

[111] The difficulty in measurement when students' academic performances are at a high baseline has been noted before for selective schools, in the context of value-added analyses (Hunt and Merrotsy, 2010).

selective school applicants), as both marginal selective and marginal non-selective were identical in characteristics by design, the small sample sizes meant that the standard errors were very large, leading to estimates which were not statistically significant; the estimates were generally not positive, with the exception of those for Mathematics (Figure 6.8) and for only two point estimates with large standard errors, for ATAR (Figure 6.4).

Related studies on peer achievement and self-concept indicated that the RD estimates may understate the general effect because relative academic status has been shown to be positively associated with later achievement, and marginal selective students are likely to be among the academically weakest at the selective schools, while marginal non-selective students are likely to be academically stronger students in the school they attend.

We reviewed the extent that students experienced decreases in local achievement ranks from attending the selective schools (in Rank Effects and Self-Concept, section 4.6.3), finding that differences in within-school ranks between successful and unsuccessful applicants with the same selective school preference were as large as 40 percentile points. From other studies, the influence on later achievement from diminished self-concept for selective students in our case study was likely to be small as they were high-achievers and because they experienced downgrades in ranks in the later stages of schooling.

The comparisons in both matching and regression discontinuity approaches were affected, to a minor extent, by the small subset of students who obtain sufficiently high scores to receive offers but choose to decline them. The applicants who decline offers appear to differ in a meaningful way, making up a small percentage of students who receive offers[112]; there are indications that they perform even better than the selective students, after controlling for NAPLAN and background characteristics. [113]

Finally, estimating value-added effects is an exercise in abstraction, one which involves the measurement of academic achievement, and relies on an interpretation of the results in its institutional context. As noted by Clark and Del Bono (2016), the effect size is dependent on

[112] See Figure 4.6 from Regression Discontinuity Methodology (section 4.5.1), which shows the acceptance rates for one of the selective schools. From graphs of these acceptance rates, less than 20% of applicants declined offers, depending on selective school preference and conditional on entrance exam score.

[113] As implied by the value-added analyses of ATAR (Appendix A.7), when both measures of prior achievement are included in addition to student background characteristics for the exam sample, which produces a negative estimate for selective attendance.

the counterfactual school environment, which may tilt the advantage towards selective schools in different areas and for certain subgroups of students; such as, for example, the larger effects that were found for disadvantaged students with lower quality outside options attending a selective high school (from the North Carolina case study, Shi 2017). While we emphasize the importance of university entrance results on subsequent university and course enrolments, we acknowledge the narrow focus of the study in ignoring other outcomes from schooling. [114]

4.7. Conclusion

This study asks whether selective schools improve the Year 12 achievement of their students, beyond what they would achieve in non-selective schools. We follow a cohort of students through high school from an anonymized Australian state, comparing the end-of-high school results of students who attend selective schools with other students who sat the entrance exam. We estimate the selective school effect via two methods: propensity score matching, which compares students of similar background and prior achievement, and regression discontinuity (RD), which compares marginal selective and non-selective students on the basis of the entrance exam.

Our results point to a small effect in terms of university entrance ranks, consistent with previous studies in similar high-income countries. We find no statistically significant estimates from the RD, which could possibly be attributed to the small sample sizes from the reliance on only one cohort of students. Although we find estimates of 2 percentile points from the matching, they appear overstated for not having accounted for additional positive selection between selective and non-selective students with similar levels of prior achievement, beyond that which exists from applying to the schools. There is also evidence, with the same caveat, that the selective schools provide a floor for Year 12 achievement; estimates are between 6 ppt and 13 ppt increases in probability of attaining ranks in the top 10% and 15% of all students.

A small selective school effect is consistent with higher levels of educational aspiration of both selective students and non-selective applicants who performed well on the entrance exam. The exploratory analyses indicated that selective students performed better on the entrance

[114] We note again the path of research inquiry pursued by Clark and Del Bono (2016) on longer-run outcomes. In addition, the poor performance of selective students for their first year grades in university remains an open question (from Dobson and Skuja, 2005).

exam than non-selective applicants who performed similarly on test scores taken before the admissions process, meaning the matching estimates may reflect pre-existing differences between selective and non-selective students, such as additional motivation or latent academic ability. In support, value-added regressions suggested that both ATAR and the entrance exam were high stakes outcomes for selective students, but not for non-selective students.

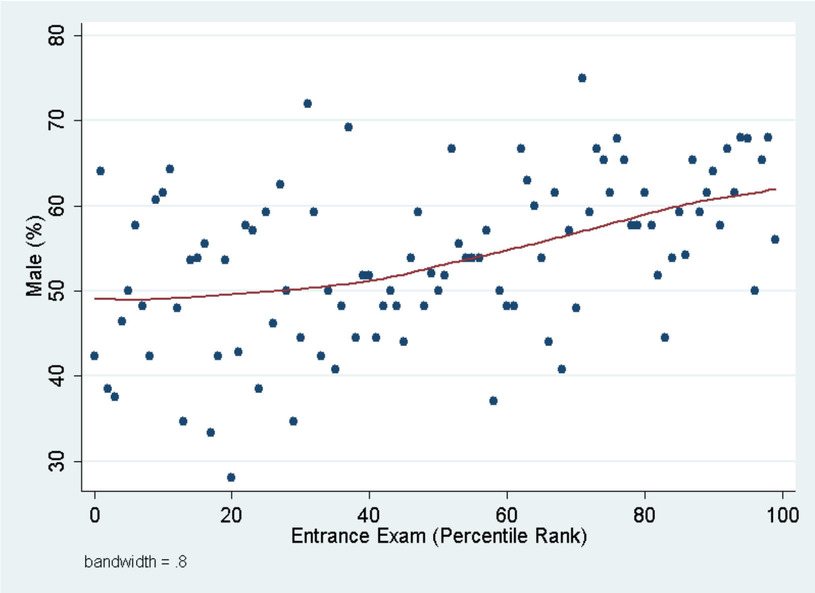
A large survey of applicants also showed that the most important reasons for accepting offers to selective schools were for their academic emphasis, including perceived advantage for their future careers, and reputational benefits (Kensell and Braithwaite, 1992). In addition, applicants to selective schools were disproportionately from immigrant and socio-economicly advantaged backgrounds, the former of which has been found to be correlated with higher levels of high school completion and educational aspiration. Reports that applicants have been coached and exert great effort in preparation for the entrance exam provide further indication of the positive selection in the applicants.

Overall, it is perhaps not so surprising that the effects of selective schools on ATARs might be relatively small for students who are among the most driven and motivated, having been positively selected from a group of students who have already expressed these positive character traits, by applying to the selective schools. The findings are consistent with the interpretation that the selective schools' policies encourage the educational and career aspirations of both their students as well as applicants who attended other schools.

Appendix

A.1. Academic Selection

Figure 4.18 (A): Locally Weighted Regression of Percentage Male & Entrance Exam



Produced using lowess command in STATA. Each dot represents at least 26 students along percentile ranks of the entrance exam. We do not distinguish between selective and non-selective students.

A.2. The Matched Data

Adding the Entrance Exam Data

To allow for comparisons of education outcomes and control for prior achievement, we match the entrance exam data to the NAPLAN and Year 12 data. Due to the absence of a unique identifier between the datasets there is some loss of observations, as described below.

Table 4.8 (A): Fuzzy Matching between the Entrance Exam and NAPLAN/Year 12 datasets

	N	(%)	Cum.
(1) Exact match	2,432	85.2	85.2
(2) Exact match with cleansed data*	37	1.3	86.5
(3) Fuzzy match with cleansed first name, surname and birthdate	15	0.5	87.0
(4) Fuzzy match with cleansed surname and birthdate	202	7.1	94.1
(5) Fuzzy match with cleansed first name and birthdate	10	0.4	94.4
(6) Fuzzy match with cleansed first name and surname	45	1.6	96.0
(7) Unmatched	115	4.0	100.0
	2,856	100	

* removing spaces, non-alphabetic characters and brackets

Fuzzy matching, which matches observations where the spelling in names are very close, was applied to students' first names, surnames and birthdates[115]. This increased the exact matching of 85.2% to 94.1% (from Table 4.8 A). We opted to use the data from match types (1) to (4), where the most lenient form of matching is a fuzzy match between cleansed surname and birth date. Dropping 170 observations (including 15 selective school students), there are 2,686 observations in the resulting sample.

Note that 4 students (1 selective school student) in the entrance exam data were not in the Year 9 NAPLAN dataset with a restricted age range of 14 to 16, leaving 2,682 observations.

[115] This step was undertaken by the State Education Department before we received the data for anonymization.

A.3. Data Sample - Further Details

Secondary Intake Students

From the sample of students who sat the entrance exam, we first exclude students who were not admitted strictly on the basis of their entrance exam results (Table 4.9 A). These are students admitted through a discretionary intake or a disadvantaged background intake. For the discretionary intake we exclude these students to reduce the possibility of positive bias in the event that these students were selected for favourable characteristics that are not reflected in the data.

We distinguish between students in the disadvantaged background intake who attained sufficiently high entrance exam scores and those who did not. Excluding the latter group of 56 and also the 46 students admitted in the discretionary stream, the remaining pool of students in the matched sample is 2,580.

Table 4.9 (A): Selective Students by Intake

		N
Matched Sample		2,682
less Discretionary Intake	46	2,636
less Disadvantaged Background Intake (below min)	56	2,580

A.4. Matching Approach

Table 4.10 (A): Characteristics of Non-Selective ("NS") and Selective ("S") Students^

	Boys		Girls	
	NS	S	NS	S
N	781	340	739	252
LBOTE (%)	76.3	77.1	75.2	84.1
SES Category (%)				
Quartile 1	14.6	10.3	15.3	12.7
Quartile 2	15.9	13.8	18.3	11.9
Quartile 3	27.3	26.5	26.4	27.8
Quartile 4	42.3	49.4	40.1	47.6
YR7 Sector (%)				
Government	58.5	58.2	50.3	50.4
YR9 Sector (%)				
Government	54.7		48.8	
Year 7 NAPLAN Results				
Std. NUM				
Mean	1.4	2.3	1.0	2.0
SD	0.9	0.9	0.8	0.9
%Missing	0.8	0.6	0.3	0.8
Std. RDG				
Mean	0.6	1.2	0.8	1.5
SD	0.7	0.7	0.7	0.7
%Missing	0.4	0.9	0.1	0.0
Score				
Mean	1.0	1.8	0.9	1.8
SD	0.6	0.6	0.6	0.6

^ above minimum score among students at Selective Schools. Excludes one school.

Table 4.11 (A): Achievement Outcomes of Selective and Non-Selective Students

	ATAR	% Missing	ATAR ≥ 95	ATAR ≥ 90	ATAR ≥ 85
Selective Boys	91.8	2.1	49.1	68.2	81.5
Non-Selective	83.6	9.0	23.9	40.8	52.0
Selective Girls	93.8	10.7	54.0	71.8	80.6
Non-Selective	84.2	8.3	22.3	37.9	52.9

Table 4.12 (A): Matching Sample Frequencies

	Boys			Girls		
	NS	S	Total	NS	S	Total
Nearest Neighbour	189	324	513	111	224	335
Kernel	686	324	1,010	642	224	866
Stratification	686	324	1,010	642	224	866

NS: Non-Selective, S: Selective

A.5. Fuzzy RD Estimates

Table 4.13 (A): Fuzzy RD Estimates of Selective School Effect by Bandwidth^

	Bandwidth								
	2	3	4	5	6	7	8	9	10
ATAR	4.0	3.2	-2.9	-1.4	-1.1	-1.8	-0.9	-1.3	-0.8
	(11.5)	(9.6)	(6.8)	(5.5)	(4.7)	(4.9)	(4.7)	(4.0)	(3.6)
	150	232	330	411	502	582	656	734	801
Maths	7.4	4.6	3.8	5.5	3.1	3.5	3.1	1.4	0.8
	(9.9)	(7.6)	(6.0)	(4.9)	(3.7)	(3.8)	(3.4)	(2.9)	(2.7)
	140	215	304	376	457	529	590	660	719
English	7.0	-2.6	-0.5	0.2	0.9	-0.3	-1.5	0.0	0.1
	(43.3)	(9.9)	(9.3)	(8.3)	(7.3)	(6.2)	(4.3)	(5.2)	(4.8)
	102	165	234	296	359	415	468	520	570
ATAR ≥ 95				0.02	0.18	0.09	-0.22	-0.18	-0.22
				(1.57)	(1.44)	(1.30)	(0.94)	(0.88)	(0.70)
				450	544	628	707	793	865
ATAR ≥ 90				-0.57	-0.44	-0.36	-0.29	-0.39	-0.39
				(0.92)	(0.76)	(0.84)	(0.83)	(0.60)	(0.50)
				450	544	628	707	793	865
ATAR ≥ 85				-0.76	-0.51	-0.34	-0.23	-0.35	-0.29
				(0.71)	(0.73)	(0.92)	(0.95)	(0.69)	(0.65)
				450	544	628	707	793	865

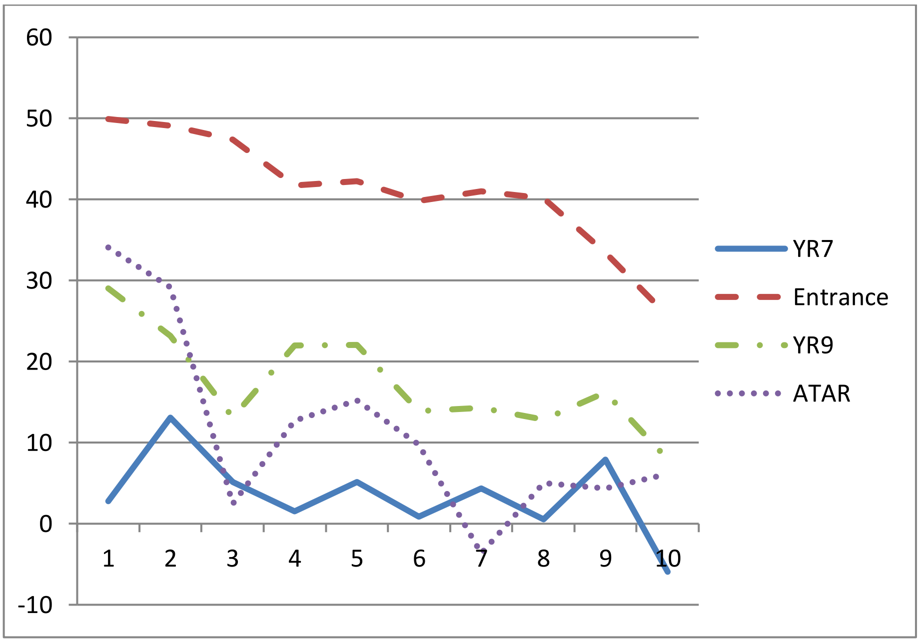
^ estimates, standard errors, and number of observations for each achievement outcome.

A.6. Student Growth Percentiles

To compare the four achievement measures over a small sample we make a coarse approximation of student growth percentiles analyses[116] (Betebenner, 2009), an approach developed in Houng and Justman (2013), the advantage of which is that we can compare across different achievement measures, by assessing relative rather than absolute performance. First, we assign students from the exam sample into deciles based on Year 7 NAPLAN. Then, within each of these deciles, we calculate the percentile rank for each of the achievement measures. Finally, we compare the percentile ranks between groups by decile.

Figure 4.19 plots the differences in percentile ranks between selective and non-selective groups on the verticial axis, by decile on the horizontal axis. Each line represents a different achievement measure: the thick blue line is Year 7 NAPLAN, the red dashed line is the entrance exam, the dash-dot is Year 9 NAPLAN, and the small dotted line is ATAR. The differences between selective and non-selective groups in Year 7 NAPLAN are unsurprisingly the smallest, considering that the deciles are based on it; the thick blue line fluctuates mostly between 0 and 10 percentile ranks.

Figure 4.19 (A): Differences between Selective and Non-Selective by Deciles of Year 7 NAPLAN^



^ There are at least 5 observations of selective students from the 2nd decile and 10 observations from the 3rd decile. Differences in percentile ranks between selective and non-selective students as calculated within deciles of Year 7 NAPLAN.

[116] The approach measures relative performance for students at similar levels of achievement in the same way that height and weight growth charts provide an indication of expected changes.

The differences in ranks for the entrance exam are of primary interest for the reason that they can help with interpreting the results from the matching approach. There is a large difference of almost 50 percentile ranks in the lowest decile (1st), which persists with a gradual decline until the 8th decile, from which there is a sharper decline from 8th through 10th deciles. As we are inclined to interpret Year 7 NAPLAN as reflecting academic ability when the stakes are low, it is surprising that the gap between selective and non-selective students remains as large as 25.6 percentile points for the 10th decile. This can plausibly be interpreted as additional effort reflecting educational aspiration as a response to the incentive of attending the selective schools.

The higher levels of outperformance of selective students over non-selective students on the entrance exam for the lower deciles is expected due to the admission mechanism, where offers are conditional on entrance exam performance. This pattern is consistent with the lowess plots of ATAR against the propensity scores (Figure 4.5, section 4.4.5), where the largest gains were attained by students with the least probability of attendance.

Analyses of student growth in achievement showed that the selective students outperformed non-selective students on the entrance exam above that which is expected from prior achievement[117]. The patterns in the gap between selective and non-selective students were similar for Year 7 NAPLAN, but were much smaller in magnitude for Year 9 NAPLAN and generally even smaller for ATAR; the exceptions were due to the first two deciles, where there were 5 observations of selective students or fewer.

[117] Differences in absolute performance by decile of Year7 NAPLAN were consistent with the above analyses. Similarly, replacing Year 7 NAPLAN with propensity scores from the matching method, to account for the background characteristics of students, produced similar results.

A.7. Value-Added Regressions

This appendix investigates with value-added regressions the idea that selective students have high levels of educational aspiration from their outperformance on the entrance exam compared with the non-selective students. We estimate regressions of ATAR separately for selective and non-selective students, while varying the specification to include the different measures of prior achievement: using Year 7 NAPLAN first; the entrance second; and lastly, together. For the other explanatory variables we include the basic student characteristics of sex, language and socio-economic background.

Table 4.14 (A) presents results from regressions of ATAR for non-selective students in columns 1 to 3 and selective students in columns 4 to 6. For the non-selective students, we reduced the sample to those who sat the entrance exam and received scores above the minimum at selective schools; the goodness of fit is higher using the results from the standardized tests than for the entrance exam score; R-square of 0.180 for Year 7 NAPLAN compared with 0.158. There is a slight increase to 0.206 with the inclusion of both achievement measures.

Table 4.14 (A): ATAR Regressions by Selective Attendance while varying Prior Achievement^

	Non-Selective			Selective		
	(1)	(2)	(3)	(4)	(5)	(6)
Entrance Exam	0.41***		0.24***	0.53***		0.46***
	(0.04)		(0.04)	(0.03)		(0.05)
YR7 NAPLAN		7.45***	5.28***		6.03***	1.55*
		(0.58)	(0.69)		(0.53)	(0.68)
Male	-2.29***	-2.41***	-2.60***	-2.99***	-2.12**	-2.86***
	(0.69)	(0.68)	(0.67)	(0.70)	(0.75)	(0.71)
Std. SES	1.45***	1.50***	1.31***	0.53	1.21***	0.51
	(0.34)	(0.34)	(0.33)	(0.35)	(0.37)	(0.35)
LBOTE	3.00***	4.13***	3.65***	1.69	2.40*	1.77*
	(0.82)	(0.82)	(0.81)	(0.87)	(0.93)	(0.88)
Constant	58.62***	76.18***	63.64***	48.07***	79.46***	51.19***
	(2.38)	(1.07)	(2.42)	(2.84)	(1.31)	(3.18)
N	992	985	985	721	713	713
R-squared	0.158	0.180	0.206	0.275	0.184	0.279

^ Entrance exam scores above min at selective schools, includes excl. selective school.

For the selective students the relationship between prior achievement variable and goodness of fit is reversed for selective students (columns 4 and 5): the entrance exam variable pushes the goodness of fit up to 0.275, which is much higher than for the non-selective students, and when Year 7 NAPLAN is included in the regression instead, the goodness of fit is similar to that for the non-selective sample. When both measures of prior achievement are included in the regression for selective students, the goodness of fit is improved slightly, at 0.279, compared to the entrance exam scores by itself.

For both non-selective and selective groups, the same general relationship in goodness-of-fit and prior achievement holds when the selective school in its first year of operation is excluded from the sample[118]; using the entrance exam produced a lower r-square than when NAPLAN was used for non-selective students and the opposite was true for selective students but the increase in predictive power from the entrance exam was smaller, at around 0.03 instead of 0.09. When the non-selective and selective students were pooled, the goodness-of-fit from ATAR regressions was similar to that of the selective students, being higher for the use of entrance exam than for NAPLAN[119].

The value-added analyses, which consisted of regressions of ATAR while varying the measures of prior achievement, showed that the entrance exam had greater predictive power for selective students than for non-selective students, which is consistent with an interpretation that selective students view the entrance exam as a high stakes outcome. We found that the entrance exam variable has greater predictive power for selective students in terms of their subsequent performance in Year 12, as indicated by higher values in R-square (larger by 0.09, 0.28 compared with 0.18), whereas the inclusion of the entrance exam makes only a small contribution to the goodness of fit for non-selective students (difference of 0.02).

[118] There was both weaker demand from applicants, and lower academic competitiveness for attendance, at the new selective school (see subsection 5.1.1, The Data Sample, The Selective Schools).

[119] We also run regressions for a combined sample of all applicants restricted to students with Year 7 NAPLAN greater than the minimum at selective schools, without the excluded school (not shown); the R-square value is higher with the inclusion of the entrance exam variable, at 0.304, compared with 0.279 with Year 7 NAPLAN, and is highest at 0.329 when both are included. The negative coefficient of -2.4 for selective attendance when we use the entrance exam to reflect prior achievement is consistent with what we found from the exploratory analyses of achievement by selective attendance; selective students outperform non-selective students in the entrance exam in general, and there is a subsequent regression-to-the mean pattern for Year 9 NAPLAN and for ATAR.

Chapter 5. The Effect of Inferred Changes in Rank on Academic Achievement

5.1. Introduction

The effects of comparisons with others have been well documented in individual-oriented performance domains characterized by competition, including: education, sports, and in organizational behaviour[120]. Individuals are thought to be influenced by comparisons they make with others through their self-concept, their belief and confidence in their own ability, which in turn affects their later performances or achievement.

In research on student achievement, Marsh and Parker (1984) introduced the 'Big-Fish-Little-Pond' effect (BFLPE), finding that similar ability students in high-achievement schools had lower self-concepts than those in low-achievement schools. This effect, also known as a group frame-of-reference or contrast effect, has been posited to explain the positive influence of within-school rank position for achievement, which contributes to research on peer effects.

In this chapter, I estimate the effect on students' achievement on standardized tests from changes in their local ranks, extending Murphy and Weinhardt (2014) who first estimated the effect of local percentile ranks on later achievement. Rather than actual changes in ranks which are measured from new test scores in the later environment, I calculate the later rank based on previous test scores and refer to the change as inferred or perceived changes in rank. This approach avoids the problem of interdependency between rank and achievement in the later time period, where the test scores already incorporate part of the response in students from receiving new peers.

I hypothesise that the changes to students' local rank affect their later achievement, as it informs students of their academic ability, and can also be an input into their decision regarding time and effort allocation. The changes in local rank can be large due to academic selection from both the formal and natural sorting of students across schools, which is driven by student demand. For example, many high achievement students experience large downgrades in relative status when moving from a general school to a highly selective school.

The study follows students transitioning from primary to secondary school, with standardized test scores in Year 5 and Year 7, from four state cohorts of government sector students from

[120] For example, Gill et al. (2015) show in a workplace study that effort provision is linked to rank-order feedback with a university lab experiment, while in competitive swimming Leung and Denton-Schneider (2017) show that swimmers who marginally qualify for heats perform worse than those who marginally miss out.

Victoria, Australia. I adopt an estimation approach of school-by-cohort fixed effect regression, which relies on idiosyncratic variation in achievement levels of the peer groups in the two different contexts, and estimate effects separately for Numeracy and Reading.

Although it is possible that a greater proportion of high-achievement students leave the government sector after leaving primary school, the use of ranks as a relative measure of performance alleviates concerns of potential bias in combination with sufficient variation in peer achievement between school environments.

In greater detail, an inferred change in rank is a student's rank of Year 5 test scores among new peers in secondary school minus their previous rank of Year 5 test scores among peers in primary school. As primary schools are generally much smaller in size then secondary schools, there is some small overlap in peers with students likely attending local schools at both primary and secondary levels. Changes in rank then reflect changes in the achievement quality of peers such that: a decline in rank denotes a move to a better peer quality school, and a rise denotes the attendance of a worse peer quality school, when peer quality is determined by achievement.

Previewing the results, I find that that large inferred increases in rank had a significant negative effect for Numeracy, of up to -0.065 SD (top decile of change), while there were only small negative effects for the top and bottom deciles for Reading (-0.02). That is, when students experienced large somewhat unexpected increases in local ranks, and entered environments with lower peer achievement, they performed worse on achievement tests. Conversely, large perceived decreases in local rank appeared only to be moderately negative for Numeracy, while the study found only small negative effects for large inferred increases or decreases in rank for Reading.

While a positive effect from an increase in ranks would be anticipated from interpreting local ranks as reflecting relative academic status, I suggest that the negative estimates for inferred increases in rank can be explained by a reduced allocation of effort on the part of students who are striving for efficiency. By relative status, I mean status in a performance sense, which is arguably social status to the extent that it provides instrumental value to others (an argument made by Anderson, Hildreth and Howland, 2015).

The results provide new evidence of self-referencing behaviour, where individuals make comparisons with their past performances (which could be relative performances, like local ranks), over short time horizons. For instance, moderate decreases in local rank may prompt students to try harder to maintain their previous standing or self-referenced benchmark. The behaviour of reduced effort allocation from students who experienced large perceived increases in rank makes sense in a context where the new information was unexpected and attributable to random variation.

This chapter is structured as follows. In section 2, I review the relevant literature, relating self-concept and rank effects to peer effects and discussing the importance of relative status in academic achievement. Section 3 provides the intuition for the approach in estimating effects from local ranks subject to random variation, and discusses the possible expected effect from inferred changes in rank. In section 4, I describe the regression specifications, building up from Murphy and Weinhardt's regressions of local rank effects. Section 5 reports the descriptive statistics, and consists of four subsections. The section leads with a description of the data, followed by demographic characteristics, local ranks, and the data sample. The results and discussion are presented in sections 6 and 7 respectively, and the robustness analyses, limitations and conclusion are in sections 8, 9 and 10.

5.2. Relevant Literature

5.2.1. Self-Concept, Peer Effects, and Rank Effects

By estimating the effect of changes in local rank on student achievement, this study draws from the literature on self-concept and frame-of-reference effects from educational psychology, and can be positioned alongside other peer effects studies on academic achievement. As noted earlier, Marsh and Parker (1984) introduced the 'Big-Fish-Little-Pond' effect (BFLPE), whereby a student's self-concepts, belief or confidence in their own ability, is influenced by the achievement level of their peer group. The effect is also described as a group frame-of-reference or contrast effect.

Within educational psychology, in addition to the effects from comparisons within local groups, another observed phenomenon is a comparison by students of their own achievement between subject domains: an internal/external frame-of-reference (Chiu, 2012; Shalvelson, Hubner and Stanton, 1976). That is, individuals tend to believe that they are better in one domain over another, despite small differences in performance. Specifically, the correlation between subject performances is higher than the correlation between subject specific self-concepts.

Counter to the negative effect on individuals' self-concept from having high-achieving peers, peer effects are generally conceived of as a conforming influence, whereby individuals' behaviour or outcome is positively associated with that of the group. For example, in the study of school choice, parents exhibit preferences consistent with the existence of positive peer effects, such that exposure to high-achievement classmates is expected to benefit students' achievement (Hastings and Weinstein, 2008).

Peer effects in education have generally been found to be modest in size from linear-in-means models, where peer achievement level is reflected by the average of the group (Sacerdote, 2011). There is, however, evidence of larger peer effects when heterogeneous effects in achievement are considered. For example, Gibbons and Telhaj (2008) found that low achievers were harmed by high achievers, whereas Lavy, Paserman, and Schlosser (2007) found that high achieving students benefit from the presence of other high achieving students, but high achievers did not help average students. Peer effects relating to the racial composition of peers and for social outcomes (e.g. drinking, drug use and criminal behaviour), which are larger than peer effects for academic outcomes, have also been documented (see Sacerdote, 2011).

The influence of peers on performance has even been observed in situations with subtle cues, such as the effect from nearby individuals; classroom seating has been shown to influence achievement (Hong and Lee, 2007)[121], while peer effects, which operate like rank effects in education, have also been observed among competitive swimmers (Yamane and Hayashi, 2015)[122].

The Big-Fish-Little-Pond effect has the most relevance in education contexts relating to academic selection within or across schools, such as gifted and talented programs as an example of the former (Marsh et al 1995), or selective high schools as an example of the latter (Marsh et al, 2007). More generally, the effect has been demonstrated in comparisons from international standardized testing (Marsh and Hau, 2003).

More recent studies have taken advantage of rich administrative datasets to apply more complex techniques which show the importance of rank position for achievement (e.g. Murphy and Weinhardt (2014), Tincani (2015), and Elsner and Isphoring, 2017)**.** In a similar approach to Lavy et al. (2012), who documented peer effects in achievement, as measured by group means, Murphy and Weinhardt (2014) used school-by-cohort fixed effects to estimate positive rank effects on later achievement from cohorts of students transitioning from primary to secondary school.

[121] Peer effects were interpreted from the influence of adjacent students from fixed seating classrooms in college in South Korea. Effects were larger for below-average students and those at the top end.
[122] There is a performance boost from having slower peers behind them and decrease in times when swimmers have a faster peer ahead of them. Observability of peers is demonstrated through the comparison of backstroke and freestyle.

The same frame-of-reference effect from high school rank was shown by Elsner and Isphoring (2017) to improve own perceived intelligence and confer higher expectations about their future career. With exogenous variation from distance to earthquake shocks in Chile, which affected the ability of students to study, Tincani (2015) showed that there were heterogeneous effects in achievement by low, middle and high ability from variance in peer achievement.

Jonsson and Mood (2008) found somewhat disconcertingly that having high achieving peers in high school reduced students' aspirations for attending university, albeit by a small amount. The influence of self-concept through ordinal achievement rank has also been extended by Elsner and Isphording (2015) to explain the likelihood of students engaging in risky behaviours, including smoking, drinking, having unprotected sex and engaging in physical fights.

5.2.2. Relative Status

The significance of comparisons to others by individuals for domain-specific performance can be situated among goal-directed behaviour explained by the motive of a desire of status, (Anderson, Hildreth and Howland, 2015)[123]. Anderson, Hildreth and Howland define status as respect and voluntary deference based on the perceived ability by others for individuals to facilitate their goal accomplishment, which they call: instrumental social value. Stated plainly, an individual's status is determined by others' perceptions of their usefulness.

The instrumental social value interpretation of status supports the assertion that competence in performance-related domains implies more utility and allows individuals to attain more respect. In support, Tay and Diener (2011) found that being social respected was more important than being socially accepted by others for measures of subjective well-being. Furthermore, the social value of status is context dependent, with competence being valued more highly for individual endeavours compared to when members worked interdependently[124]. For example, competition between individuals within university research groups drives members towards acquiring technical knowledge, which from research on rivalry can be interpreted as a form of status competition.

Anderson et al. recognize the importance of local environment for ascertaining the significance of status, as opposed to focusing on absolute measures or globally defined measures like

[123] Anderson et al. posit that the desire for status is a fundamental motive which is not derivative of other motives, and is universal across cultures. They also suggest that the desire for status induces goal-directed behaviour and that status has long term consequences, which they identify as psychological adjustment, health and subjective well-being.
[124] Shepard (1954), Kilduff et al. (2012), and Fragale (2006), cited in Anderson, et al. (2015).

income inequality at the country level. They note studies showing the relationship between organizational rank and morale, greater levels of life satisfaction among individuals with higher incomes within geographic areas, and similar studies showing that subjective well-being is particularly affected for individuals who uniquely experience unemployment within their social circle.

Within Anderson et al.'s framework, relative status appears to operate at two levels, with the striving for academic success conferring general social status as one interpretation, while comparisons with others by individuals based on academic achievement, being performance within a domain taking place as part of a competitive process, as another interpretation.

The translation of effort from short-term comparisons to longer term specialization and occupational outcomes appears to be mediated by perceptions of ability and future expectations like that found by Elsner and Isphording (2017); high school ranks in achievement improved the own-perceived intelligence and career expectations of students.

5.3. Intuition

5.3.1. Random Variation in Local Ranks

The intuition for the approach builds on the idea that individuals are informed by relative differences to others from comparisons within local groups. In terms of self-concept, this is the BFLPE or group frame-of-reference effect applied to the education context, as described by Marsh (1987). For convenience, I define relative comparisons to mean comparisons individuals make with others within the local context.

The relative-age advantage that leads to 'redshirting' behaviour both in education and in sports, where parents delay their child's entry to kindergarten and hence schooling, can be readily interpreted as motivated by the potential boost in self-concept received from being a big fish in a small pond. The behaviour is posited to confer an advantage for children over their peers, which can be academic, social or physical.

The importance of relative comparisons for individuals has broader applicability than for performance-related domains. In the family context the birth-order of individuals has been found to influence their personalities, which was most clearly shown by Sulloway (1996) for researchers and their receptivity to new scientific ideas; first-borns tend to be more conservative and reject revolutionary ideas unlike later-borns[125]. Kagan (2013) also suggests

[125] They are also more likely to be dominant, aggressive, ambitious, jealous, anxious about status and emotionally intense (Koch, 1954, 1955, and Brim, 1958, cited in Sulloway, 1966).

that sibling rivalry leads to specialisation in different areas: e.g. if one child is inclined towards sports, then the other might pursue musical interests.

Kagan (2013) provides examples for the size of the community with a similar line of reasoning; he makes the point that individuals growing up in small towns are likely to be over-represented among successful professionals and scientists (cosmologists) despite individuals having access to greater resources in cities[126]. Similarly, in this chapter there is natural variation in the size of schools that students attend which results in different relative comparisons, and therefore allows for the estimation of its impact[127]. I am able to additionally identify the impact of changes in relative differences when students move from primary to secondary school, which is typically a move from a smaller to larger school, as measured by the number of students.

5.3.2. Expected Effect from a Change in Local Rank

Under an interpretation where academic achievement can be thought of as a performance domain characterized by competition and relative status, it is likely that students are sensitive to changes in local ranks to the extent that the ranks reflect competence. When the students enter secondary school, the change in school environment is likely to prompt a re-evaluation in (achievement) status, which was one of the observed behaviours noted by Anderson, Hildreth and Howland (2015).

In their intent to establish the desire for status as a fundamental motive, Anderson et al. hypothesise that individuals would vigilantly monitor their status. In support of this idea, they cite studies which show that people pay attention to trivial or subtle indicators of status such as office décor, negligible differences in clothing, emotions such as pride in achievement, while another level of support for the importance of perceiving status is that individuals display a high degree of accuracy in identifying each other’s status.

The a priori effect on achievement from a change in relative status, which is implied by a change in local rank, is unclear. Although a positive relationship has been established between local ranks and later achievement within the research on self-concept and the big-fish-little-pond effect, it is not obvious that retaining or gaining high local ranks in a new school environment will help improve achievement, particularly when motivation is factored in. One can imagine certain individuals performing better with increased expectations from ranking

[126] Size of community, Chapter 4: The Family and Beyond, The Human Spark.
[127] The variation in school size is largely determined by the residential choices of individuals and reflects a natural sorting process (Schelling, 1969).

highly and others performing worse from increased expectations. In a workplace setting experiment, Gill et al. (2015) found that high and low ranked participants exerted greater effort[128]. It seems probable that the optimal level of peer achievement, which the local ranks measure indirectly, should be neither too high or too low for each student.

Separate from the optimal peer level of achievement, there is another effect from the random variation in peer achievement which is dependent on school choice and geography. The positive rank effect established by Murphy and Weinhardt suggests that it is beneficial to maximize local ranks, such that it is also preferable to obtain increases in rank, where possible. This is consistent with the 'redshirting' behavior of parents holding their children back from school entry to obtain a relative-age advantage. However, the tendency for students to positively sort into schools on an academic basis, whether through tracking into ability classes or into high-achievement schools, implies that there are other important considerations as students advance through school.

The research by Dweck (2000) on achievement motivation and Duckworth (2016) on grit suggests that overcoming adversity and increased competition may be beneficial for attaining longer-term outcomes, given a growth mind-set towards learning[129]. In the short-run, which this study concentrates on, a case can be made for detrimental effects from artificially high or low ranks caused by random variation in the schools that students choose to attend. Changes in either direction would probably prompt a re-evaluation of their relative status in students.

Lower local ranks than reflected in students' global ranks could lead to lower academic self-concept and school anxiety, as is documented in Becker, Neumann, et al. (2014), when they compared early-entry and regular students who attended an academically selective school in Germany. On the other hand, higher relative status from increased ranks may unintentionally result in students reducing their motivation or aspiration, having seemingly performed well, and perhaps above their own expectations.

Another relevant study is Becker and Neumann (2016), who investigated how the BFLPE, the positive effect on academic self-concept from having low-achieving peers, changed across primary and secondary school contexts. Extending Becker, Neumann, et al. (2014), they have three data points, one point before the transition to secondary school, and two points after:

[128] With reference to relative-performance in the workplace Gill et al. (2015) conduct an experiment to obtain an estimate of the rank response function, how effort provision responds to rank-order feedback. They find a U-shaped function, which they characterize as first-place loving and last-place loathing.
[129] Dweck distinguishes between a fixed and growth mindset in students, finding that they are related to helpless or master-oriented patterns of behaviour, respectively. See section 2.3.5 on Achievement Motivation from Chapter 2: Literature Review.

directly after the change and one year later. Becker and Neumann find statistical evidence that the BFLPE on academic self-concept persists after the transition to the new school, but that the effect fades and a new self-concept is established in the new school environment.

5.4. Method

The objective is to estimate the effect of changes in local rank for individuals on later achievement, between two time periods. The time periods are two years apart reflecting the year levels in which the standardized tests were conducted. Test scores for each of the subject-domains Numeracy and Reading are used as both the outcomes of interest and measures of prior achievement. First, I set up a baseline regression specification by building up from the previously used conventional models. To assist with intuition if necessary, a simplified overview of the steps is presented in Appendix A.1 that describes the positive and negative associations between the key variables.

5.4.1. Rank Effects

5.4.1.1. Measuring Local Ranks

Local ranks are calculated as within-school percentile ranks, which range from 0 to 100. The percentile ranks ($rank_{ij}$) are calculated, explicitly as a function of achievement ($f_j(A_i)$), as follows:

$$rank_{ij} = f_j(A_i) = 100 \times \frac{R_i - 1}{N_j - 1}$$

where R_i is the rank of achievement (A_i) for student *i* in school *j*, and N_j is the number students in school *j*. The ranks of achievement are ascending with increasing achievement, with 1 for the lowest rank and N_j for the highest.

A shortened notation for ranks (without subscripts for individual and school) is used throughout the chapter: r_{t} , while superscripts *l* and *g* are introduced to indicate local and global ranks. i.e. $r^{\mathrm{g}}{}_{\mathrm{t}}$ and $r^{\mathrm{l}}{}_{\mathrm{t}}$. Global ranks are presented for context in the descriptive statistics in Local Ranks, section 5.5.3. The global ranks are calculated the same way as the local ranks but over the whole student population within each year-level, who are students attending Victorian government sector schools.

5.4.1.2. Regression Model

The conventional approach for estimating the influence of self-concept from relative status, referred to as a rank effect or big-fish-little-pond effect, is to create a measure within the same time period as the achievement outcome of interest (e.g. Elsner and Isphording, 2017 and Marsh and Hau, 2003). Murphy and Weinhardt (2014) extend the conventional approach to calculate local ranks from achievement in the previous time period to estimate their effect on later achievement. They posit that the ranks inform students about their relative ability which in turn affect their beliefs about costs in terms of effort.

The following regression specifications formalize the empirical approach by including individual background characteristics and school level indicators. There are two specifications (that follow the simplified versions presented in Appendix A.1): one for the rank effects **(1)**, and one for the effect from changes in rank between time periods, the main specification **(2)**, which is introduced in the next section (section 5.4.2.2). The main specification focuses on the impact of changes to local ranks between time periods **t** and **t + 1**.

$$A_{ik}(\boldsymbol{t}+\boldsymbol{1}) = a + \beta_1 \boldsymbol{X}_i(\boldsymbol{t}) + \beta_2 A_{ij}(\boldsymbol{t}) + \beta_3 r_{ij}^{l}(\boldsymbol{t}) + p_j + s_k + c_v + \varepsilon_i \quad \textbf{(1)}$$

For the outcome variable, I estimate value-added regressions of achievement (***A***) in time period **t+1** on local rank (from period **t**) and secondary school (**s**) and *cohort* fixed effects (**c**) and primary school random effects (**p**). The assumption of random effects for primary schools is appropriate as students tend to attend local schools such that the distribution of primary schools in achievement (t+1) resembles that of a normal distribution. In support of this assertion, Figure 5.1 in section 5.5.2 (later) shows that the mean achievement (t) in primary school appears normally distributed. Explanatory variables are prior achievement, individual and socioeconomic and language background characteristics (***X***).

Note that secondary school effects are included to address the possibility that the estimates of inferred change in rank are picking up some form of specific individual school effects. This suggests that students who did not experience increases in rank do better on later achievement due to sorting rather than from the change in ranks. That is, there is a positive sorting of students based on their prior achievement into schools with greater achievement value-add. The trade-off from including the secondary school effects is that the estimates of inferred ranks may be biased downwards due to the correlation of the former with local ranks in secondary school.

5.4.2. Changes in Local Ranks

5.4.2.1. Measuring Inferred Local Ranks

Next, I introduce the time notation to compare local ranks between two time periods, and drop the subscript *i*. I denote subscripts *t* and *t+1,* for periods **t** and **t+1**, respectively. Relative status is then reflected by percentile ranks, in periods **t** and **t+1***,* of:

$$rank_{jt} = f_j(A_t)$$

$$rank_{k(t+1)} = f_k(A_{t+1})$$

with the possibility that students attend different schools *j* and *k,* in **t** and **t+1**. As before they are both functions of achievement, and the local ranks are calculated within the time periods the standardized tests are taken, as convention.

Now, I extend the previous research to measure the later achievement (in period **t+1**) as a function of earlier tests results (from period **t**), with an asterisk superscript to indicate this idiosyncrasy. I define the percentile rank for student *i* in school *k* (**t+1**) as follows:

$$rank_{k(t+1)^*} = f_k(A_t)$$

where the percentile rank is calculated from the students who attend the new school *k* in period **t+1**, but achievement is from period **t**.

Lastly, we arrive at the variable of interest, the change in ranks between periods **t** and **t+1**. I drop the school subscript (*j* or *k);* they are assumed. I define the change in ranks as a subtraction of local rank (*t*) from local rank (*t+1**) below.

$$\Delta rank_{t,t+1^*} = rank_{t+1^*} - rank_t$$

In the regression specifications, presented in the next section, local ranks are shown in more concise notation as r^{l}, with changes in local ranks represented by $\Delta r^{l}_{t,(t+1)^*}$.

5.4.2.2. Regression Model

The rank effects specification (from section 5.4.1.2) is presented below in (1), along with the main specification in (2), which includes the inferred change in ranks. The inferred change in rank (Δr^{l}_{ijk}) is defined as the difference between previous and later ranks (from the previous section):

$$A_{ik}(\boldsymbol{t+1}) = a + \beta_1 \boldsymbol{X}_i(\boldsymbol{t}) + \beta_2 A_{ij}(\boldsymbol{t}) + \beta_3 r^{l}_{ij}(\boldsymbol{t}) + p_j + s_k + c_v + \varepsilon_i \quad \textbf{(1)}$$

$$A_{ik}(\boldsymbol{t+1}) = a + \beta_1 \boldsymbol{X}_i(\boldsymbol{t}) + \beta_2 A_{ij}(\boldsymbol{t}) + \beta_3 r^{l}_{ij}(\boldsymbol{t}) + \beta_4 \Delta r^{l}_{ijk}(\boldsymbol{t+1}^{*}) + p_j + s_k + c_v + \varepsilon_i \quad \textbf{(2)}$$

For notation, the letters in parentheses indicate the time period in which achievement was measured (t, t+1 or t+1*), while the subscripts denote the group of students (*j* or *k*) from the school attended.

5.4.3. Specifications

In the previous sections I outlined the specification for the main analyses of inferred change in rank, building up from a baseline specification from Murphy and Weinhardt (2014), which includes a local rank variable. In the results, I first present a simplified version of the main specification, specification 2 from section 5.4.2.2, to introduce the relevant variables, and also review the changes in goodness of fit arising from changes in school effects, or the inclusion of local rank. I note also the change in test score coefficient from the inclusion of local rank, which likely reflects the interdependency between achievement and rank. i.e. higher achieving students will mechanically have higher local ranks.

For the remainder of the results, the majority of time is spent assessing the functional forms of the key variables of interest: local rank, inferred change in local rank, and Year 5 test scores. The functional form of local rank was also considered by Murphy and Weinhardt, who found that the relationship was largely linear, while it is quite possible that that there is a stronger relationship between ranks and later achievement for higher ranked and lower ranked students, as was found in Gill et al.'s (2015) workplace study.

In summary, I propose the following regression analyses:

1) A simplified version of the main specification
2) Varying forms of local rank in baseline specification (Murphy and Weinhardt)
3) Results from main specification
    - including varying forms of local rank
    - including non-linear cubic form for prior achievement
4) Results from interaction variables between local rank and inferred change in rank.

All regressions are estimated with clustered standard errors for primary school.

5.5. Descriptive Statistics

This section consists of a description of the administrative dataset, an overview of the demographic characteristics at both individual and school level, an illustrative example of the variation in local ranks and inferred local ranks, a review of the data sample, and lastly an exploratory analysis, visualizing the key variables of interest.

5.5.1. Data Description

The data consists of administrative data and standardized test results from the "National Assessment Program – Literacy and Numeracy" (NAPLAN) for four matched cohorts from two year levels which are two years apart: Grade 5 and Year 7; spanning the years 2008 to 2013. The median ages of the students in the cohorts are 11 and 13 in Year 5 and Year 7, respectively. In addition to test results for Numeracy and Reading, and indicators for missing values, there is information relating to students' socioeconomic (SES) and language background, as well as whether they have an Indigenous background (ATSI[130]). From the parental education and occupation information, I derive an index of socioeconomic status (SES) using principal components analysis within each year-level and year[131].

5.5.2. Demographic Characteristics

There were 1,393 primary and 448 secondary schools for a state population of 265,254 and 226,526 students respectively, attending Victorian government schools between 2008 and 2013.

From Table 5.1, 51.9% of students in Year 5 were male, 24.8% were from a foreign language background, and 1.7% of students were from an indigenous background. The SES values are standardized normal with mean zero and a standard deviation of 1 within each year level. Around 8.9% and 8.7% of students had missing values for Numeracy and for Reading test scores, with a combined percentage of 7.6% for both.

[130] Aboriginal or Torres Strait Islander.
[131] This process is documented in Chapter 3, section 3.4.2.

Table 5.1: Student Level Demographic Characteristics (2008 - 2013)

	Year 5	Year 7
Male (%)	51.9	52.7
Foreign Language Background (%)	24.8	25.6
ATSI (%)	1.7	2.0
Age	10.5 (0.4)	12.5 (0.4)
Std. SES	0.0 (1.0)	0.0 (1.0)
NAPLAN Scores		
Mean Numeracy	495.2 (71.7)	539.1 (71.3)
Mean Reading	499.8 (76.2)	539.1 (67.8)
Missing Values (%) – Numeracy	8.9	9.3
Missing Values (%) – Reading	8.7	9.1
Missing Values (%) – Both	7.6	7.0
Observations (Students)	265,254	226,526

There were slight increases in percentage male, foreign language background and ATSI, in Year 7, at 52.7%, 25.6%, and 2.0%, respectively. The percentage of missing test scores was also slightly increased at 9.3% for missing Numeracy, 9.1% for missing Reading, and 7.0% for both Reading and Numeracy missing.

Table 5.2 presents average and standard values for school level characteristics in Year 5 and year 7; rows presenting median values, taken from the median school, are labelled (median). For reference, school level characteristics weighted by student numbers are included in Appendix A.2, Table 5.32 (A). Students usually move from smaller primary schools to larger secondary schools, with an average number of students of 34 in Year 5 and 95 in Year 7.

Table 5.2: School Level Characteristics^

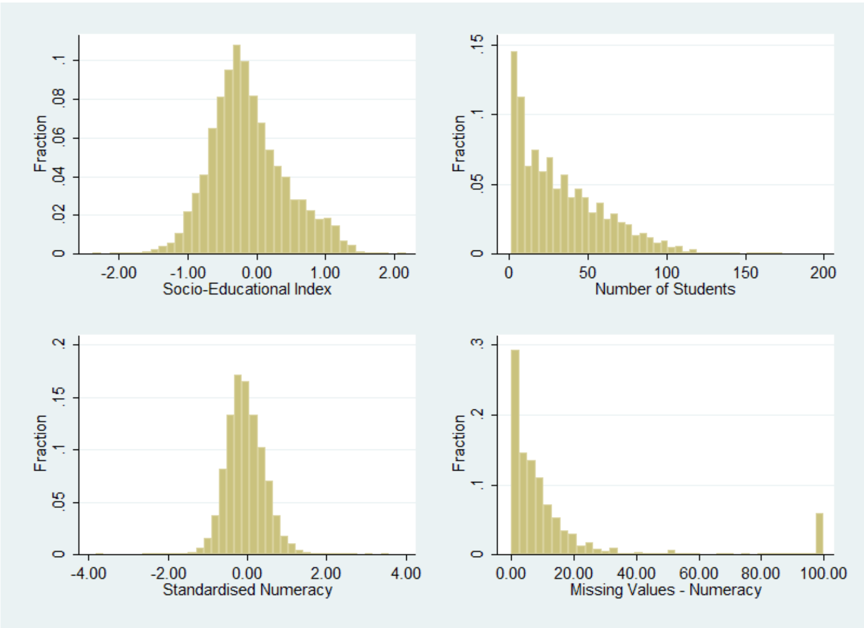
	Year 5	Year 7
Number of Students	34 (28)	94 (84)
Male (%)	52.9 (17.0)	56.0 (16.5)
Foreign Language Background (%)	19.2 (24.1)	22.5 (25.5)
(median)	8.7	10.6
ATSI (%)	2.4 (6.9)	3.0 (9.1)
Age	10.5 (0.2)	12.5 (0.2)
Std. SES	-0.11 (0.55)	-0.10 (0.50)
NAPLAN Scores		
Std. Numeracy	-0.04 (0.50)	-0.10 (0.46)
Std. Reading	-0.04 (0.48)	-0.09 (0.47)
Missing Values – Numeracy (%)	13.7 (23.7)	27.4 (37.4)
(median)	6.5	9.1
Missing Values – Reading (%)	13.4 (23.7)	27.2 (37.5)
(median)	6.2	8.8
Observations (Schools)	7,873	2,406

^ Means and standard deviations in parentheses, or medians where indicated

The negative test score values of -0.04 for both Numeracy and Reading suggests that there is a degree of sorting with more schools with a greater concentration of low achievement students, while the negative average SES values of -0.11 for Year 5 and -0.10 for Year 7 also show that there is a greater concentration of students with lower SES across schools. The median primary school had only 8.7% of students with a foreign language background in Year 5, and 10.6% in Year 7, which also showed that there was a tendency for students with a foreign language background to attend the same schools.

The histograms in Figure 5.1 show that both the number of students and the percentage of missing test score values in Year 5 were asymmetric in distribution, with the latter clustering towards zero with also a small percentage at 100.0. The same is observed in Year 7 (not shown here).

Figure 5.1: Selected School Level Characteristics for Year 5, 2008 to 2013



For reference, Table 5.3 reports the Year 5 values of standardized SES, Numeracy and Reading in Year 7; these variables were defined over the student population in Year 5 with means of 0 and SDs of 1. This provides some indication of the selection from students who leave the government sector to attend non-government secondary schools, and indicates in addition, the extent to which there is sorting of students into schools in Year 7. The former is not problematic for estimating the change in ranks given that the ranks are calculated within school in both year levels, so long as there is some variation in peer achievement and SES when students transition from primary to secondary school. Overall, the level of sorting from both forms of selection between Year 5 to Year 7, across sectors and across schools, does not appear very large.

Table 5.3: Primary School SES, Numeracy and Reading (Year 5) in Secondary School (Year 7)

Year 5	Student level (Year 7)	N	School level (Year 7)	N
Std. SES	-0.16 (0.95)	128,727	-0.22 (0.50)	1,649
Std. Numeracy	-0.08 (0.98)	117,171	-0.23 (0.52)	1,345
Std. Reading	-0.08 (0.98)	117,482	-0.22 (0.52)	1,344

Table 5.3 presents mean values of SES, Numeracy and Reading taken from Year 5 for students and schools in Year 7. At the student level, the mean values of SES, Numeracy and Reading were slightly negative, at -0.16 for SES, and -0.08 for both Numeracy and Reading, with standard deviations near one. This suggests that there is some negative selection of students at the end of primary school, with some students from advantaged socioeconomic backgrounds and with higher levels of achievement, moving to non-government sector secondary schools. Overall, these differences in average levels of SES and achievement between year levels do not appear particularly large, although selection is slightly higher for SES.

The negative sorting across schools by both SES and achievement was more apparent than at the student level, where the mean Year 5 values in SES, Numeracy, and Reading were remarkably close, at -0.22 for both SES and Reading, and -0.23 for Numeracy. These school level means of Year 5 values reflect the selection across secondary schools, in addition to the negative selection across education sectors between primary and secondary school.

The mean Year 5 SES in Year 7, at -0.22, is slightly larger in magnitude than the mean Year 5 SES at the student level, -0.16. This difference of -0.06 was less than the difference between school level and student level SES within year-levels, which were -0.11 for Year 5 and -0.10 for Year 7, from Table 5.1 and Table 5.2.

For Numeracy and Reading, the within year-level differences between student and school level were -0.06 and -0.05 (from Table 5.1 and Table 5.2), compared with differences between Year 5 and Year 7, which were -0.15 and -0.14 respectively (from Table 5.3). Together, school level means of Year 5 values in Year 7 suggested that there was greater increase in selection across schools by achievement than by SES, although the resulting levels were similar at between -0.22 and -0.23 because the sorting across schools by SES was initially higher in primary school, at -0.11.

5.5.3. Local Ranks

Table 5.4 provides an indication of variation in local ranks as students progress through school, showing that the average difference between local and global ranks is 0.0 and has a standard deviation that ranges between 0.114 and 0.150 standard deviations depending on year level and subject domain. The level of variation was smaller for Reading than for Numeracy, while the standard deviations decreased between Year 5 and Year 7, but increased again in Year 9 for both subjects.

Table 5.4: Difference between Global and Local Ranks by Year Level & by Subject^

	Numeracy	Reading
Year 3	0.0 (0.144)	0.0 (0.136)
Year 5	0.0 (0.147)	0.0 (0.134)
Year 7	0.0 (0.129)	0.0 (0.114)
Year 9	0.0 (0.150)	0.0 (0.128)

^ mean (sd); ${r^{l}}_{t}$ - ${r^{g}}_{t}$

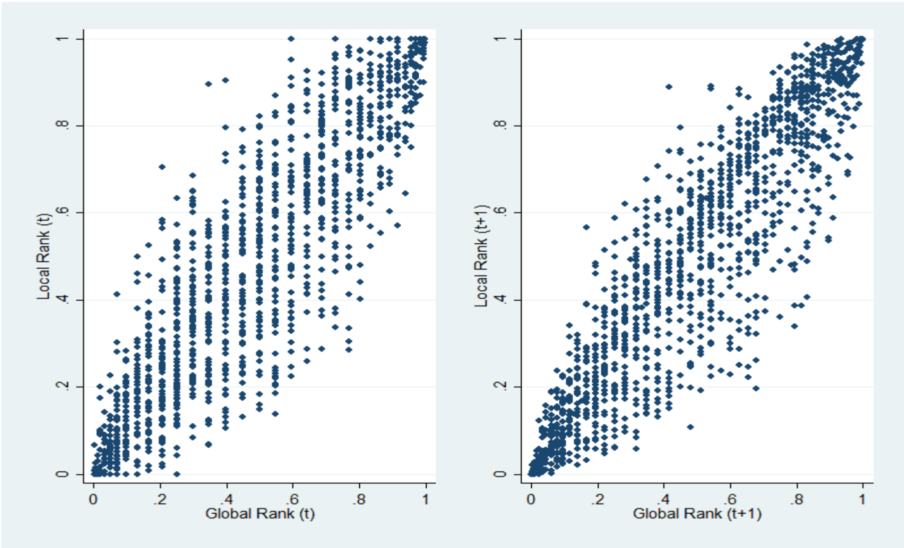
Table 5.5 similarly shows that correlations between local ranks and global ranks, as well as correlations between local ranks and test scores were very high, at over 0.850 and as high as 0.922 for the correlation between local and global ranks for Year 7 Reading. The samples are all students without missing values in each year level from 2008 to 2013.

Table 5.5: Correlation between Local Ranks and each of Global Rank and Achievement

		Local Rank (${r^{l}}_{t}$)			
		YR3	YR5	YR7	YR9
Numeracy	Global Rank (${r^{g}}_{t}$)	0.880	0.874	0.902	0.866
	Achievement (A_t)	0.860	0.850	0.872	0.817
	N	243,610	241,414	205,431	206,106
Reading	Global Rank (${r^{g}}_{t}$)	0.893	0.895	0.922	0.903
	Achievement (A_t)	0.877	0.878	0.904	0.884
	N	244,178	242,120	205,780	206,329

Figure 5.2 shows a scatterplot of local and global ranks in Numeracy for a 5% sample of the 2008 to 2010 cohort, for time periods: t (Year 5) and t + 1 (Year 7). The plot shows both the variation in local ranks arising from different peer levels of achievement, as well as the high levels of correlation between the two variables.

Figure 5.2: Local and Global Ranks, Numeracy (t) and (t+1)^



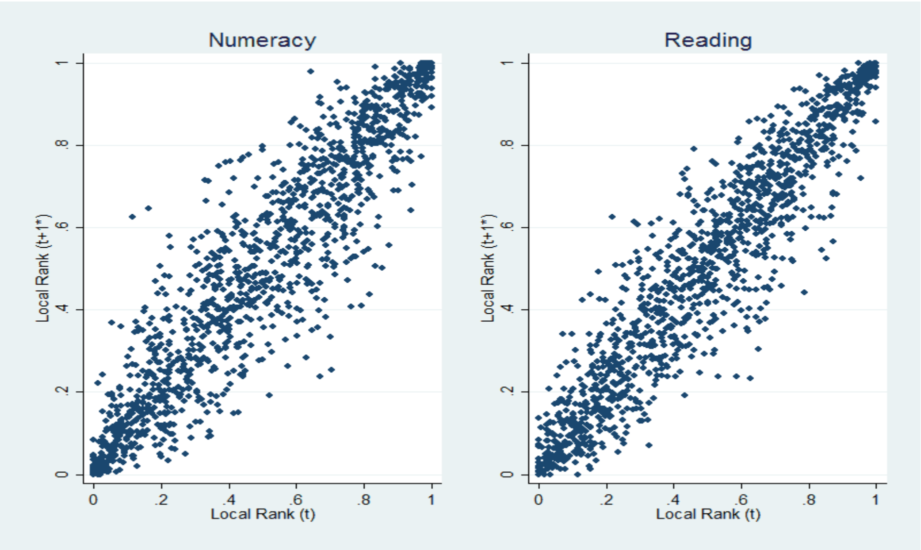
^ 5% sample from cohort 1: 2008 to 2010

5.5.3.1. Illustrative Example

The level of variation between local rank in Year 5 and Year 7 is presented in scatter plots of local ranks in time *t* and local ranks in time t+1*, in Figure 5.3, where the later ranks are based on Year 5 test scores, as per the main analyses. For a clearer picture, a 10% sample from the first cohort, from 2008 to 2010, is shown. Like the earlier correlations in Table 5.5, local ranks in year 5 were also highly correlated with later local ranks in year 7, at 0.91 for Numeracy and 0.92 for Reading. Because the local ranks are based on the same test scores in Year 5, the relationship is actually slightly less noisy than plots of local and global ranks within time periods. i.e. local and global ranks from Year 5 test scores and local and global ranks from Year 7 test scores (Figure 5.2).

Next, the variable of interest, inferred change in local ranks, is presented against global ranks in time *t* (in Year 5) for Numeracy and Reading, in Figure 5.4. There is significant variation in the inferred change in rank – on the vertical axis – across the range of achievement, as measured by global ranks – on the horizontal axis, although the plot is noisier for Numeracy, and there is much less variance at the tail ends, for low and high values of global ranks. This is as expected in that there is less potential for change in rank due to being near the maxima and minima of test scores. Unlike the earlier plots of local ranks and inferred ranks, the correlation between global ranks in Year 5 and inferred changes in ranks are close to zero, and slightly negative, at -0.04 for Numeracy and at -0.01 for Reading.

Figure 5.3: Local Ranks (t) and Inferred Rank (t+1*), Numeracy & Reading



^ 5% sample from cohort 1: 2008 to 2010

Figure 5.4: Inferred Change in Local Ranks (t, t+1*) and Global Ranks (t), Numeracy & Reading^

^ 5% sample from cohort 1: 2008 to 2010, corr = -0.04 (NUM), -0.01 (RDG)

5.5.4. Data Sample

The initial student population is filtered down from 177,204 in Year 5 and 149,476 in Year 7 down to 127,927 after the requirement that students appear in both Year 5 between 2008 and 2011 and in Year 7 between 2010 and 2013. To estimate the value-added regressions of achievement in time *t+1* against prior achievement in *t*, and to calculate the changes in rank, the analyses require that students appear in both time periods: in Year 5 and in Year 7.

A second qualifier that students attend schools with at least 20 students who have test scores from Year 5 in both primary and secondary school (because local ranks are sensitive to school size) reduces the sample further. Table 5.6 shows that these restrictions do not substantially change the demographic characteristics of the sample. The 'final sample' column combines restrictions from both Numeracy and Reading test scores, where in the regressions the restriction of at least 20 students in each of primary and secondary school is applied separately by subject. Note that the local ranks are calculated over all students who appear in each time period, which minimizes missing data problems (discussed further in the limitations).

Table 5.6: Demographic Characteristics for Data Sample^

	Year 5		Year 7	
	Sample	Final Sample	Sample	Final Sample
Male (%)	52.5	51.1	52.5	51.1
Foreign Language (%)	25.3	27.4	22.8	24.8
ATSI (%)	1.7	1.3	2.0	1.5
Std. SES	-0.15 (0.95)	-0.09 (0.96)	-0.01 (1.0)	0.07 (1.0)
Age	10.5 (0.42)	10.5 (0.40)	12.5 (0.42)	12.5 (0.4)
NAPLAN Scores				
Std. Numeracy (SD)	-0.08 (0.98)	-0.04 (0.98)	0.01 (1.0)	0.08 (0.99)
Std. Reading (SD)	-0.08 (0.99)	-0.04 (0.98)	0.02 (1.0)	0.08 (0.97)
Missing (%) – Numeracy	8.7	0.0	9.0	0.0
Missing (%) – Reading	8.5	0.0	8.8	0.0
Missing (%) – Both	7.3	0.0	6.7	0.0
Missing (%) – Either	9.9	0.0	11.1	0.0
Number of students	127,927	93,486	127,927	93,486

^Initial population of 177,204 and 149,476 for Year 5 and for Year 7 respectively.
Sample refers to students who appeared in both years. Final sample includes the additional restriction that they attended schools with at least 20 students with Numeracy and Reading test scores.

Students in the final data sample differed slightly to those from the state population with the main difference being that students with missing values in test scores were excluded. The removal of students attending schools with only few students with test score values resulted in slightly fewer boys in the sample, slightly more students from a foreign language background, fewer indigenous students, and students with slightly higher averages in Numeracy and Reading test scores and slightly higher SES.

The local ranks for both time periods *t* and *t+1* ($r^{\mathrm{l}}{}_{t}$ and $r^{\mathrm{l}}{}_{t+1}$) are shown in Table 5.7 for Numeracy, along with the change in rank ($\Delta r^{\mathrm{l}}{}_{t,t+1}$), inferred rank in t+1, $r^{\mathrm{l}}{}_{t+1^{*}}$, and the inferred change in rank ($\Delta r^{\mathrm{l}}{}_{t,t+1^{*}}$). The local ranks were similar across both samples, from the initial restriction that students appear in both time periods, and the second restriction that they attended schools with at least 20 students who received test scores (the final sample). For differences in local ranks based on test scores within each time period, the difference (or change) was 0.02 ranks with a standard deviation of 0.215, within the final sample.

Table 5.7: Local and Inferred Ranks, Numeracy^

		Sample	Final Sample
Local Rank(t)	$r^{\mathrm{l}}{}_{t}$	0.487 (0.296)	0.493 (0.292)
Local Rank(t+1)	$r^{\mathrm{l}}{}_{t+1}$	0.505 (0.290)	0.515 (0.287)
Change in Rank (t, t+1)	$\Delta r^{\mathrm{l}}{}_{t,t+1}$	0.021 (0.221)	0.021 (0.215)
Inferred Local Rank(t+1*)	$r^{\mathrm{l}}{}_{t+1^{*}}$	0.500 (0.291)	0.505 (0.290)
Inferred Change in Rank (t, t+1*)	$\Delta r^{\mathrm{l}}{}_{t,t+1^{*}}$	0.013 (0.128)	0.012 (0.116)

^ Sample refers to students appearing in both Year 5 and Year 7. Final sample includes the additional restriction that they attended schools with at least 20 students with test scores.

The inferred change in rank, based on only Year 5 test scores, was smaller, at 0.012 with a standard deviation of only 0.116. In other words, assuming a normal distribution in inferred change in rank, the average student experienced only a 1.2 percentile change in rank when moving from primary to secondary school, while 66% (within 1 SD) experienced a change of up to 11.6 percentile points, and 95% experienced less than 23.2 percentile point changes.

In the main analyses, I estimate the inferred changes in ranks in both quartile and decile forms as it is not immediately apparent what its relationship is with later achievement (discussed in section 5.3.2). By estimating quartiles and deciles indicators, the effects are interpreted against the omitted category, which is the first category relating to large inferred decreases in rank. Table 5.8 and Table 5.9 present the mean inferred changes in ranks by quartile and by decile, respectively. The percentile ranks have been scaled down by 100.

Table 5.8: Inferred Changes in Ranks by Quartile of Inferred Change in Rank, by Subject

	Quartiles of Inferred Changes in Rank			
	Q1	Q2	Q3	Q4
Numeracy	-0.13	-0.02	0.04	0.16
	(0.07)	(0.02)	(0.02)	(0.08)
Reading	-0.11	-0.01	0.04	0.14
	(0.07)	(0.02)	(0.02)	(0.07)

From Table 5.9, for Numeracy the inferred change in rank was -0.21 for the bottom quartile (D1), reflecting a large decrease in rank (or increase in peer achievement), while students in the top decile of inferred change in rank experienced increases of 24% percentile ranks. For Reading, the magnitudes of the inferred change in ranks were smaller in the bottom and top deciles, at -0.18 and 0.21, respectively.

Table 5.9: Inferred Changes in Ranks by Deciles of Inferred Change in Rank by Subject

Decile of Inferred Change in Rank	Numeracy	Reading
1	-0.21 (0.07)	-0.18 (0.06)
2	-0.10 (0.02)	-0.08 (0.02)
3	-0.05 (0.01)	-0.04 (0.01)
4	-0.02 (0.01)	-0.02 (0.01)
5	0.00 (0.01)	0.00 (0.00)
6	0.02 (0.01)	0.02 (0.01)
7	0.04 (0.01)	0.04 (0.01)
8	0.08 (0.01)	0.07 (0.01)
9	0.13 (0.02)	0.11 (0.02)
10	0.24 (0.08)	0.21 (0.07)

5.5.5. Exploratory Analyses; Visualization

This section presents a visualization of residualized achievement against the key variables of interest: achievement, local ranks and inferred change in ranks. I apply lowess smoothing with multiple predictors[132] to residualised later achievement with each of the key variables. The scatterplots foreshadow the regression results as the smoothing provides an indication of the direction of the relationship between later achievement and each variable.

From the mlowess command in STATA[133], lowess smoothing with multiple predictors adjusts for the predictors simultaneously and is based on generalized additive models (Fox and Weisberg, 2010, citing Hastie and Tibshirani, 1990). Specifically, locally weighted regression (lowess) is applied using a backfitting algorithm, cycling over each of the predictors, to estimate the fit for partial residuals of later achievement. These partial residuals, which I also refer to as residualised later achievement, are the values of later achievement less the smoothed values of the set of other predictors.

To show the varying influence of each variable on later achievement, I incrementally add the key variables (achievement, local rank and inferred change in rank) to the model of later achievement against SES to produce the following progression:

(1) later achievement (t+1) against SES (t) and achievement (t);
(2) later achievement (t+1) against SES (t), achievement (t), and local rank (t); and
(3) later achievement (t+1) against SES (t), achievement (t), local rank (t), and inferred change in rank (t, t+1*).

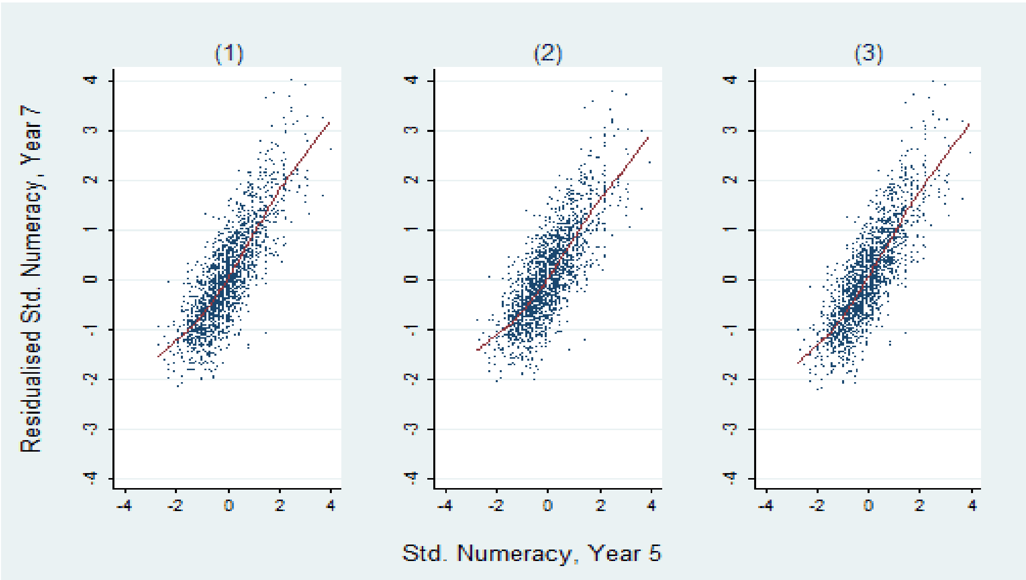
For Numeracy and Reading, Figure 5.5 to Figure 5.9 present scatterplots of residualised later achievement, with lines representing lowess smoothing with multiple predictors, for each of: achievement (t), local ranks (t), and inferred changes in rank (t, t+1*). For each variable, the number of scatterplots presented in each figure reflects the number of models (defined above) that the variable appears in.

For achievement (t), there are three scatterplots of residualised later achievement (t+1) - on the y-axis - and achievement (t) - on the x-axis - i.e. with residualised later achievement subtracting: first, SES (t), second, SES (t) and local rank (t), and third, SES (t), local rank (t), and inferred change in rank (t, t+1*). The pattern in lowess lines is similar, with little differences in

[132] Lowess smoothing with multiple predictors extends lowess in the case with one predictor (described in Chapter 4, section 4.4.5 accompanying Figure 4.5).
[133] Author Cox, N., Durham University. mlowess is closely related to the multivariate scatterplot smoother mrunning (Royston and Cox, 2005).

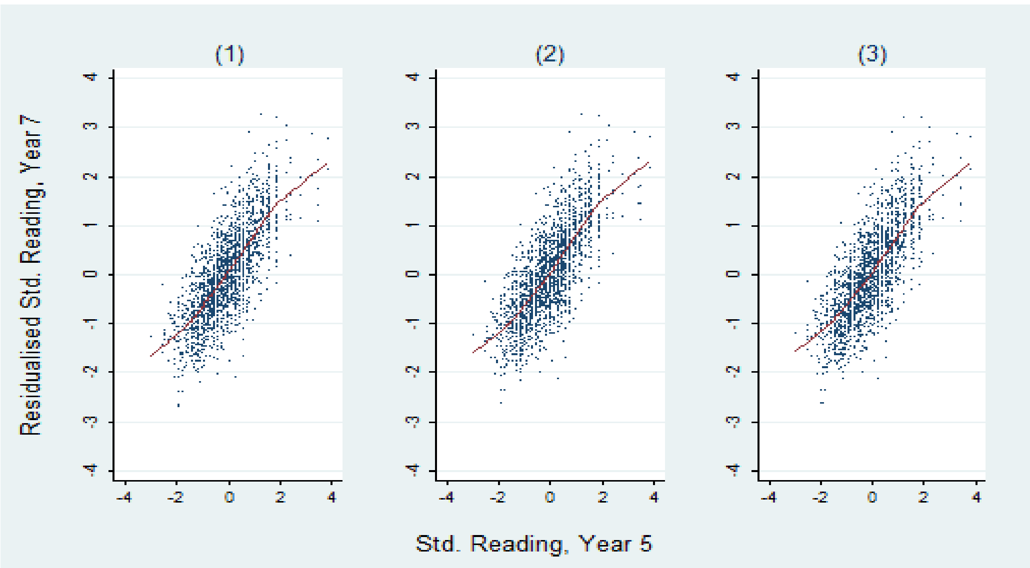
gradient in the three panels, for both Numeracy in Figure 5.5 and Reading in Figure 5.6. The gradient of the lowess lines, first reflecting a simple positive relationship of achievement between time periods in panel (1), becomes marginally steeper from the inclusion of local rank (t) in panel (2), and is similar in gradient to panel (1) in panel (3), with the inclusion of inferred change in rank (t, t+1*), in addition to local rank (t) and SES (t) (which is in the model in all three panels).

Figure 5.5: Lowess with Multiple Predictors of Residualised Numeracy (t+1) and Numeracy (t)^



^ Note that for clarity a randomly chosen 2% sample is shown.
Refer to definitions of (1), (2) and (3) at the start of section 5.5.5.

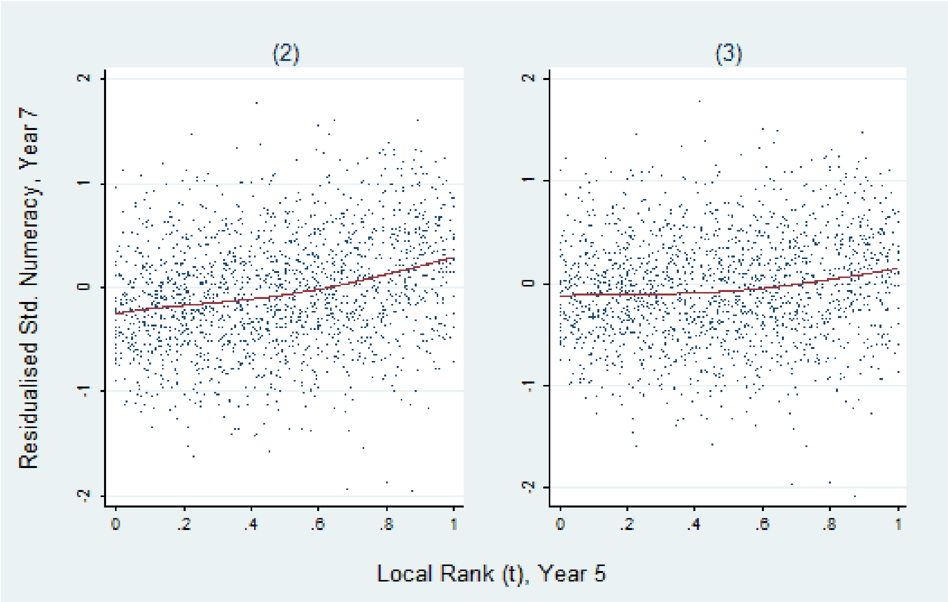
Figure 5.6: Lowess with Multiple Predictors of Residualised Reading (t+1) and Reading (t)^



^ note that for clarity a randomly chosen 2% sample is shown.
Refer to definitions of (1), (2) and (3) at the start of section 5.5.5.

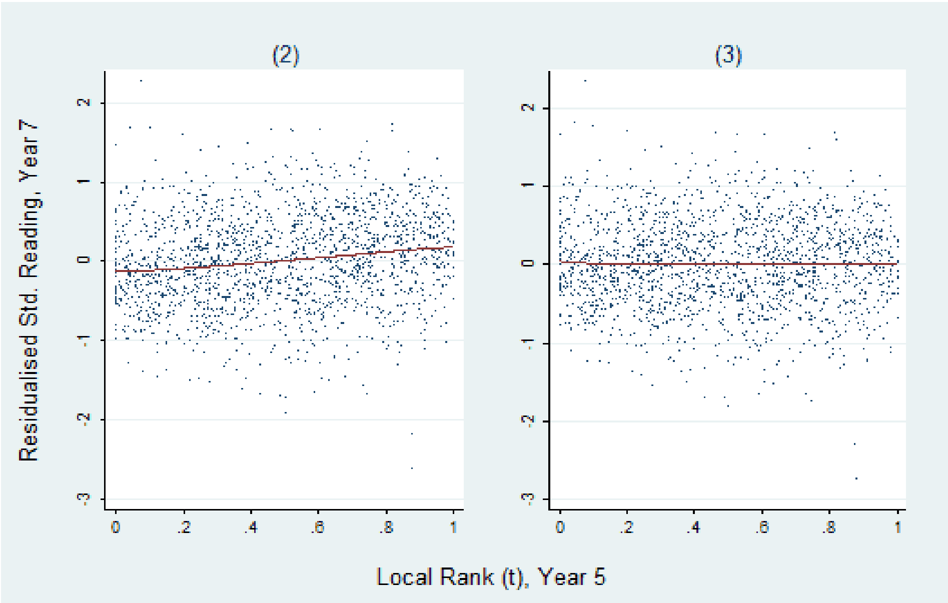
Figure 5.7 and Figure 5.8 present for Numeracy and for Reading respectively, scatterplots of the relationship between residualised later achievement (t+1) and local ranks (t) for model 2 (left panel) and model 3 (right panel). In both figures, the red lines show positive sloping relationships between residualised later achievement (t+1) and local rank (t), consistent with the results from Murphy and Weinhardt (2014).

Figure 5.7: Lowess with Multiple Predictors of Residualised Numeracy (t+1) and Local Rank (t)^



^ note that for clarity a randomly chosen 2% sample is shown.
Refer to definitions of (2) and (3) at the start of section 5.5.5.

Figure 5.8: Lowess with Multiple Predictors of Residualised Reading (t+1) and Local Rank (t)^



^ note that for clarity a randomly chosen 2% sample is shown.

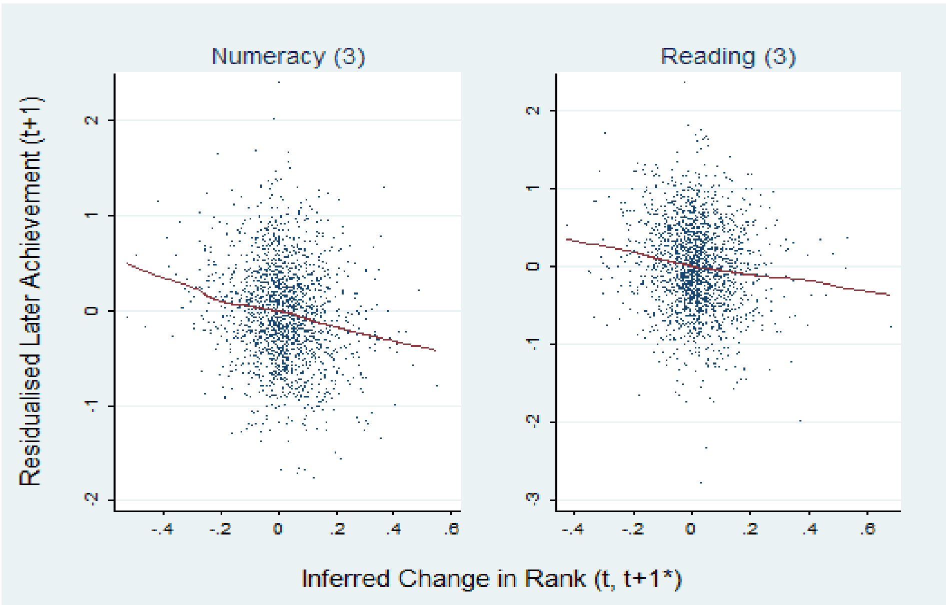
The gradient for Numeracy, in the right panel in Figure 5.7 representing model 3, is slightly smaller from the inclusion of inferred change in rank (t, t+1*), but is essentially flat for

Reading, in the right panel in Figure 5.8. Note that the inclusion of standardized SES was necessary to produce the positive gradients between residualised achievement (t+1) and local ranks (t) for Numeracy and Reading, excluding model 3 for Reading, meaning that higher local ranks are associated with lower SES values.

Figure 5.9 presents the scatterplots and lowess smoothing with multiple predictors for residualised later achievement (t+1) and inferred change in ranks (t, t+1*), with Numeracy in the left panel and Reading in the right panel. This reflects model 3 defined earlier, with all the variables included: SES (t), achievement (t), local rank (t), as well as inferred change in rank (t, t+1*). The smoothing lines are downward sloping for inferred changes in rank, with a steeper gradient for Numeracy than for Reading, which is consistent with the estimates from the main results (in section 5.6, later).

For both the local ranks (t), in Figure 5.7 and Figure 5.8, and inferred change in ranks (t, t+1*), in Figure 5.9, their relationship with (residualised) later achievement was relatively weak. The range of residualised later achievement in the plots, between -2 and 2, reflect this, as does the noisiness of the scatterplots, which includes only 2% of the sample.

Figure 5.9: Lowess w/ Multiple Pred. of Residualised Achievement & Inferred Change in Ranks^



^ note that for clarity a randomly chosen 2% sample is shown.

5.6. Results

5.6.1. Simplified Main Specification

To introduce the key variables and the value-added specifications, Table 5.10 presents estimates from a simplified version of the main specification. The introductory specification has both prior achievement and local ranks in linear forms, while estimates for the inferred change in rank takes the form of quartile indicators, which is used consistently throughout the results. Results from both primary school random effects (RE) and fixed effects (FE) regressions are presented in separate columns for each of Numeracy and Reading. The regressions also include secondary school and cohort indicators, but they are not shown.

Table 5.10: Simplified Version of the Main Specification by Subject and School Effects^

	Numeracy t + 1		Reading t + 1	
	RE	FE	RE	FE
Std. Test Score	0.631***	0.633***	0.481***	0.463***
Male	0.010*	0.009*	-0.035***	-0.035***
Foreign Language Background	0.118***	0.119***	0.028***	0.033***
Std. SES	0.077***	0.073***	0.097***	0.090***
Local Rank (t)	0.442***	0.440***	0.734***	0.796***
Inferred Change in Rank Q1 (t, t+1*)	omitted	omitted	omitted	omitted
Inferred Change in Rank Q2 (t, t+1*)	0.068***	0.063***	0.058***	0.046***
Inferred Change in Rank Q3 (t, t+1*)	0.027***	0.023**	0.063***	0.046***
Inferred Change in Rank Q4 (t, t+1*)	-0.035***	-0.035***	0.071***	0.050***
Constant	-0.338***	-0.310***	-0.524***	-0.495***
N	93,425	93,425	93,997	93,997

^ Secondary school and cohort indicators are included but not shown.
With primary school random effects (RE) and secondary school fixed effects (FE).
Std. Test Score refers subject domain indicated by column name at time t.

For the key variable of interest, the inferred change in rank, the estimates of the quartile indicators were also consistent across school effect variations (fixed or random effects). As discussed earlier, I adopted quartile indicators since the effects from inferred changes in rank are likely to be non-linear as these changes can be positive or negative.

For Numeracy and primary school random effects (in the first column), experiencing a change in the second quartile, which implies a decrease in perceived local rank or increase in peer achievement, had a positive effect of 0.068; the coefficient of the quartile indicator is relative to the omitted category of experiencing the largest decreases in rank. Relative to the omitted first quartile, the effect was smaller for the third quartile, which relates to students who perceived on average an increase in local rank, at 0.027, and negative for the fourth quartile, at -0.035.

For Reading and Numeracy, the estimates were very similar regardless of adopting random effects or fixed effects to reflect primary school attended. Given that most students attend their local primary school, the assumption of random effects that which corresponds to observations drawn from a normal distribution is appropriate, whereas fixed effects would be suitable if there is uneven sorting or unusually large or small influences on achievement at certain schools. I adopt fixed effects for the secondary schools that students attend, where there appears to be greater intentional sorting and limits on student places.

Concentrating on the random effects estimates for Numeracy from Table 5.10, being male and having a foreign language background was associated with increases in Year 7 test scores, by 0.01 and 0.12 SD respectively, while a 1 standard deviation increase in socioeconomic index was positively associated with increases in scores by 0.077. The regression produces a large coefficient for prior achievement at 0.631, while the effect from having higher local ranks in primary school is 0.128 for an increase with one standard deviation of within-school rank (0.29 x 0.442); note that 0.29 is the standard deviation of local rank which has a standard uniform distribution.

For Reading and primary school random effects, being male was negative for later achievement, at -0.035, and the effect of having a foreign language background was smaller than for Numeracy, at 0.028, although the effect of socioeconomic background was larger, at 0.097. Compared to Numeracy, the effect of prior achievement for Reading was smaller (0.481), while the effect from local ranks was higher (0.734); the 1-SD equivalent effect is 0.21 (calculated by: 0.734 x 0.29).

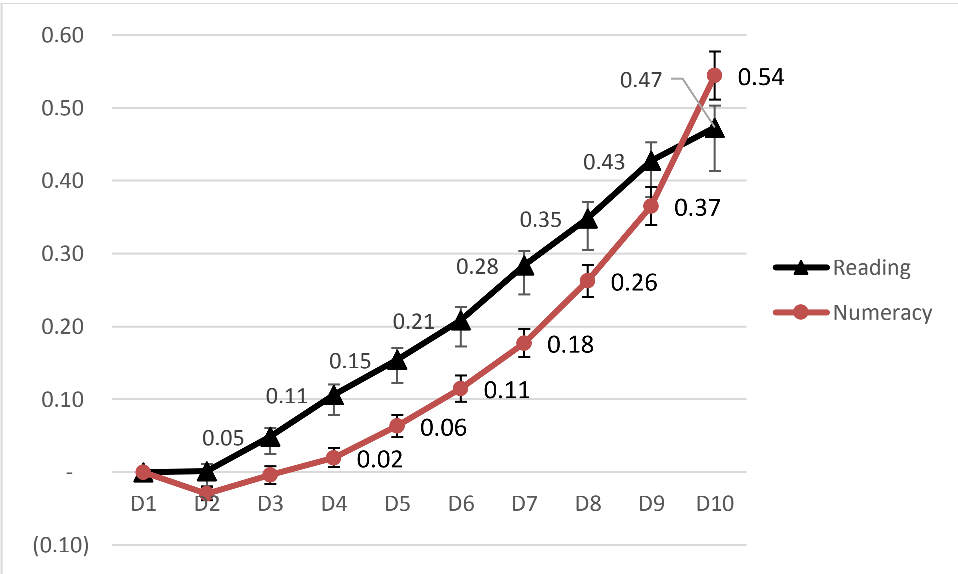
The preliminary results from the simplified main specification suggest that it is best for students in terms of later achievement in Numeracy to experience small decreases in local rank (falling into the second quartile of inferred change in rank). The pattern for Reading was different to that of Numeracy, with increasingly positive effects from larger increases in inferred local rank, relative to the omitted first quartile which relates to decreases in rank. The largest effect for Reading was 0.071 from the fourth quartile of inferred change in rank.

5.6.2. Rank Effects

Next, we turn our attention to the functional form of the local rank effect from the baseline specification, from Murphy and Weinhardt (2014). Although Murphy and Weinhardt found support for a linear functional form, I investigate several variations, given the different institutional context, and the additional complexity introduced by focusing on inferred changes in rank. For ease of interpretation, a linear form for prior achievement is assumed for this set of regressions. However, this issue is revisited later in the comparisons of estimates of inferred change in rank in section 5.6.3.2.

Figure 5.10 plots the estimates (with error bars) from decile indicators for Numeracy and Reading (t+1), with the first decile being the omitted category (from Tables 5.11 and 5.12). A within school standard deviation is equivalent to almost 3 deciles at 0.29, for a standard uniform distribution, which the local ranks are. For Numeracy, the difference in the size of estimates between deciles 7 and 10 is 0.43, but only 0.02 between the decile 1 and 4. The pattern in the estimates of local rank for Numeracy is clearly not linear, while the pattern for Reading appears to be almost linear from the second decile onwards.

Figure 5.10: Local Rank Effects, Numeracy (t+1) & Reading (t+1)^



^ estimates of decile indicators of local ranks from Tables 5.11 and 5.12

The regression results from the baseline specifications with local ranks (from Murphy and Weinhardt) are presented in Table 5.11 for Numeracy and Table 5.12 for Reading. The regressions do not include the inferred change in local ranks, while each column presents a different functional form for the local ranks. These are: linear, quadratic, quartile and decile. The estimates for decile indicators of local rank were plotted in Figure 5.10. The estimates from column 1 for both Numeracy and Reading are very similar to those from the results of the simplified main specification, Table 5.10.

Table 5.11: Local Rank Effects, Numeracy (t+1)^

	(1)	(2)	(3)	(4)
Std. Numeracy	0.622***	0.584***	0.660***	0.593***
Male	0.011**	0.006	0.009*	0.006
Foreign Language Background	0.119***	0.114***	0.116***	0.115***
Std. SES	0.077***	0.077***	0.076***	0.077***
Local Rank(t)	0.471***	-0.290***		
Local Rank(t)^2		0.878***		
Local Rank D1 (t)			omitted	omitted
Local Rank D2 (t)			0.000	-0.030***
Local Rank D3 (t)			0.095***	-0.004
Local Rank D4 (t)			0.283***	0.020*
Local Rank D5 (t)				0.063***
Local Rank D6 (t)				0.115***
Local Rank D7 (t)				0.177***
Local Rank D8 (t)				0.263***
Local Rank D9 (t)				0.365***
Local Rank D10 (t)				0.544***
Constant	-0.340***	-0.254***	-0.191***	-0.250***
N	93,685	93,685	93,685	93,685

^ cohort and secondary school indicators not shown, primary school random effects.
Local rank: linear, quadratic, quartiles and decile indicators. Errors clustered by primary school.

Table 5.12: Local Rank Effects, Reading (t+1)^

	(1)	(2)	(3)	(4)
Std. Reading	0.518***	0.513***	0.580***	0.540***
Male	-0.034***	-0.035***	-0.035***	-0.035***
Foreign Language Background	0.027***	0.027***	0.028***	0.027***
Std. SES	0.097***	0.097***	0.096***	0.096***
Local Rank(t)	0.615***	0.364***		
Local Rank(t)^2		0.266***		
Local Rank D1 (t)			omitted	omitted
Local Rank D2 (t)			0.072***	0.001
Local Rank D3 (t)			0.195***	0.049***
Local Rank D4 (t)			0.331***	0.106***
Local Rank D5 (t)				0.154***
Local Rank D6 (t)				0.208***
Local Rank D7 (t)				0.284***
Local Rank D8 (t)				0.349***
Local Rank D9 (t)				0.428***
Local Rank D10 (t)				0.473***
Constant	-0.408***	-0.372***	-0.243***	-0.302***
N	94,259	94,259	94,259	94,259

^ cohort and secondary school indicators not shown, primary school random effects.
Local rank: linear, quadratic, quartiles and decile indicators. Errors clustered by primary school.

5.6.3. Inferred Change in Ranks

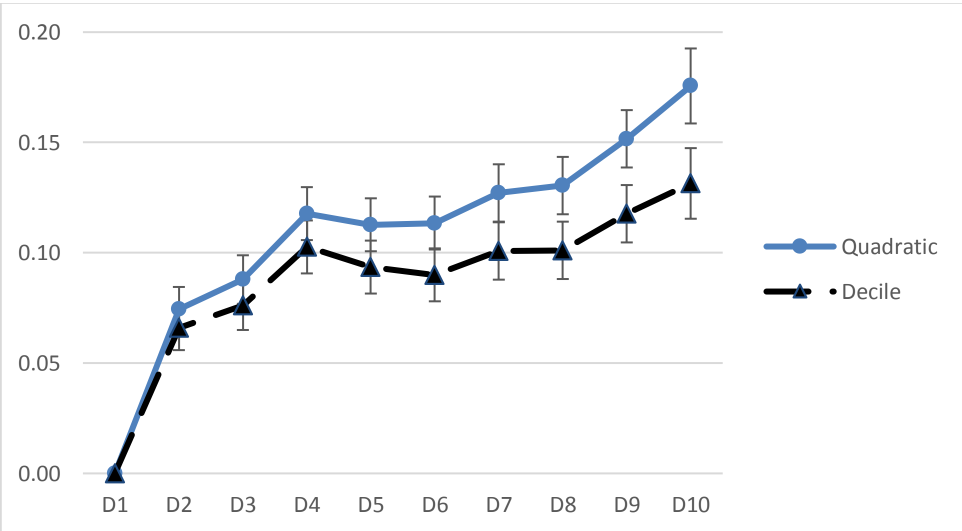
This section presents the estimates for inferred change in ranks in two sets of results. The first set of results relates to the varying of functional forms in local rank, while in the second set of results I introduce a cubic specification of prior achievement to account for the possibility that the estimates for inferred change in rank are reflecting unexplained variation in achievement[134]. As noted earlier, there is an interdependency between achievement and rank, with higher achieving students likely having higher ranks. In a third subsection (section 5.6.3.3) I investigate the interaction between local ranks and inferred change in ranks.

As it is not clear how the perceived changes in rank should influence later achievement and because changes can be either positive or negative, I plot the estimates of decile indicators for the inferred change in rank to ascertain the overall pattern and its relation to later achievement. For ease of comparison, I adopt quartile indicators for the regression estimates.

5.6.3.1. Inferred Change in Rank with Varying Functional Form of Local Rank

Deciles of estimates of inferred change in rank for Reading are presented in Figure 5.11, which are similar to the earlier plot of local ranks in Figure 5.10, in section 5.6.2. The plot shows a non-linear pattern in the effect of inferred changes in rank, with a sharp increase in estimate between the first and fourth deciles, followed by a flat section between the fourth decile (D4) and the eighth decile (D8), before increases in estimate to the tenth decile (D10).

Figure 5.11: Deciles of Inferred Change in Rank, different functional forms of Local Rank, Reading^



^ estimates of decile indicators of inferred change in rank from Table 5.39 (A), Appendix A.5.

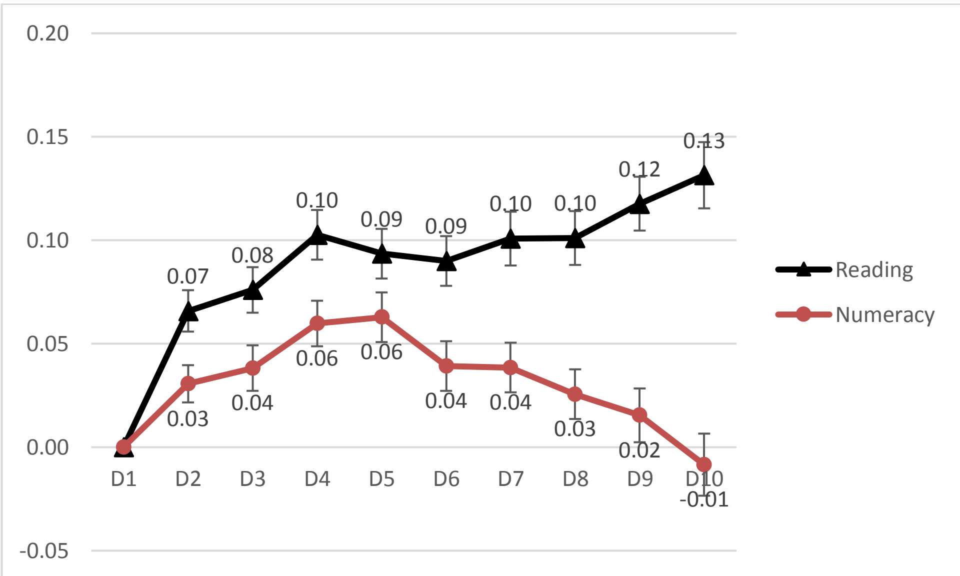
[134] For reference, estimates of different functional forms of local ranks with cubic specifications of prior achievement are included in Appendix A.3.

The increases in estimate for inferred change in rank are more pronounced for the quadratic functional form of local rank, which are represented by the circles. The decile estimates are represented by triangles, along with a dashed line.

For the preferred results, I adopt the decile indicators of local rank for their flexibility in anticipation of adding a cubic specification for the prior achievement in the next subsection (section 5.6.3.2). In separate analyses (see Semi-nonparametric Estimation of Local Rank, Appendix A.4) I found that there was no statistical difference between a quadratic form and a non-parametric form for local ranks, when prior achievement was included as a linear variable. For Numeracy, a parametric form (linear, quadratic or cubic) of local ranks that was statistically equivalent to one identified non-parametrically could not be identified.

Figure 5.12 plots the estimates of decile indicators for inferred change in ranks, when local rank is also specified as decile indicators, in the value-added regressions for Numeracy and Reading. The pattern for Reading, represented by triangles, is generally increasing in effect size. For Numeracy the effects, represented by dots, are more muted and are largest at deciles 4 and 5. In contrast to Reading, the estimates for Numeracy at the top deciles (D9 and D10) are near zero, meaning that large inferred increases in rank have a similarly low impact on later achievement, as for large inferred decreases in rank.

Figure 5.12: Deciles of Inferred Change in Rank, Numeracy and Reading^



^ the Reading decile estimates of inferred changes in ranks are the same as from Figure 5.11, which showed inferred changes in rank for different functional forms of local rank. Refer to Table 5.39 (A), Appendix A.5.

Table 5.13 and Table 5.14 show the regression results with quartile indicators for inferred change in ranks. As per the local ranks section (section 5.6.2), each column refers to a different functional form for local rank. Consistent with Figure 5.12, the results support an inverse U-shape for Numeracy in the effect of inferred change in rank, and positive effects with increasing inferred change in rank for Reading, regardless of the functional form for local rank.

Table 5.13: Quartiles Indicators of Inferred Change in Rank, Numeracy^

	(1)	(2)	(3)	(4)
Std. Numeracy	0.631***	0.582***	0.673***	0.596***
Male	0.010*	0.006	0.008*	0.006
Foreign Language Background	0.118***	0.114***	0.115***	0.114***
Std. SES	0.077***	0.077***	0.077***	0.077***
Local Rank (t)	0.442***	-0.229***	quartile^	decile^
Local Rank (t)^2	N/A	0.822***	N/A	N/A
Inferred Change in Rank Q1 (t, t+1*)	omitted	omitted	omitted	omitted
Inferred Change in Rank Q2 (t, t+1*)	0.068***	0.039***	0.032***	0.038***
Inferred Change in Rank Q3 (t, t+1*)	0.027***	0.018*	-0.012	0.014
Inferred Change in Rank Q4 (t, t+1*)	-0.035***	0.005	-0.060***	-0.006
Constant	-0.338***	-0.285***	-0.166***	-0.268***
N	93,425	93,425	93,425	93,425

^Local rank variables: linear, quadratic, quartiles indicators, deciles indicators
With primary school random effects (RE) and secondary school fixed effects (FE). Errors clustered by primary school.

Table 5.14: Quartiles Indicators of Inferred Change in Rank, Reading^

	(1)	(2)	(3)	(4)
Std. Reading	0.481***	0.473***	0.578***	0.512***
Male	-0.035***	-0.035***	-0.036***	-0.036***
Foreign Language Background	0.028***	0.028***	0.029***	0.028***
Std. SES	0.097***	0.096***	0.096***	0.096***
Local Rank (t)	0.734***	0.489***	quartile^	decile^
Local Rank (t)^2	N/A	0.274***	N/A	N/A
Inferred Change in Rank Q1 (t, t+1*)	omitted	omitted	omitted	omitted
Inferred Change in Rank Q2 (t, t+1*)	0.058***	0.045***	0.023***	0.037***
Inferred Change in Rank Q3 (t, t+1*)	0.063***	0.056***	0.013	0.040***
Inferred Change in Rank Q4 (t, t+1*)	0.071***	0.080***	0.009	0.054***
Constant	-0.524***	-0.491***	-0.258***	-0.388***
N	93,997	93,997	93,997	93,997

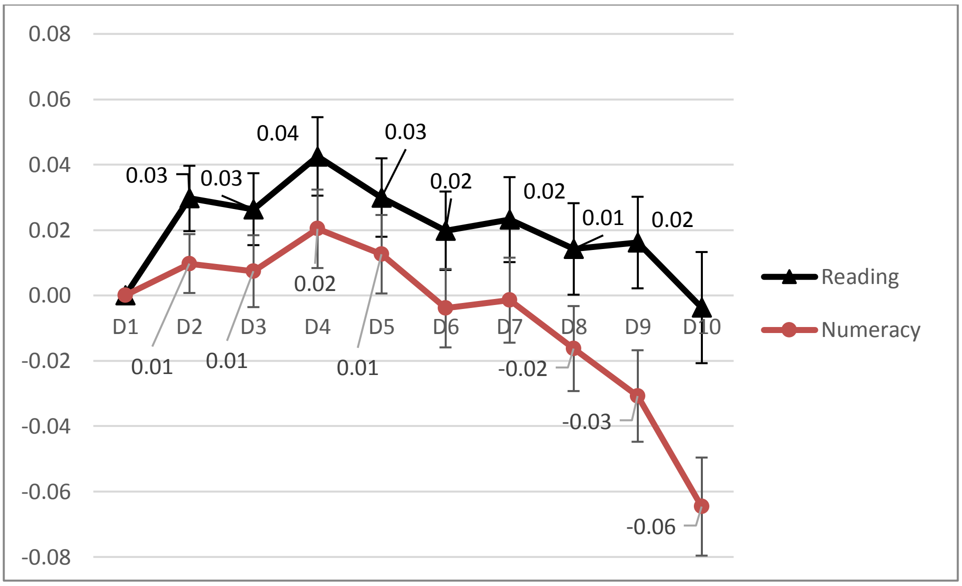
^Local rank variables: linear, quadratic, quartiles indicators, deciles indicators
With primary school random effects (RE) and secondary school fixed effects (FE). Errors clustered by primary school.

5.6.3.2. Cubic Functional Form for Prior Achievement

For the second main set of results on inferred change in rank, I adopt decile indicators for local rank but instead specify the more flexible cubic functional form for prior achievement. This is done to address the earlier observation that local ranks appear to reflect to a large extent the same underlying natural academic ability as the test scores. As a result, the patterns in inferred change in rank from the value-added regressions may be identifying non-linearities in prior achievement, or in local ranks.

The addition of quadratic and cubic terms for prior achievement results in a change to the patterns in inferred change in rank, as shown in Figure 5.13 for Numeracy and Reading, respectively. These decile estimates are included in Appendix A.5, in Table 5.40 (A) for Numeracy and Table 5.41 (A) for Reading.

Figure 5.13: Inferred Change in Rank, Numeracy & Reading



In the new regression estimates, the results for Reading now resemble those of Numeracy with a flatter pattern in decile estimates (than in Figure 5.11 and Figure 5.12), but with generally more positive estimates (reaching 0.043). In contrast to Numeracy, the top decile (D10) for Reading had an effect which was statistically indistinguishable from the bottom decile (D1). For Numeracy, being in the top decile of inferred change in rank had a negative impact for later achievement, at -0.065. Overall, the estimates were somewhat muted for Reading, ranging from 0.00 to 0.04, with the maximum estimate of 0.04 in decile 4. The estimates for Numeracy provided a larger range, between 0.02 in decile 4 and -0.06 in the top decile.

Table 5.15 and Table 5.16 show the regression results for quartile indicators of inferred change in rank. Like the first set of results, the estimates for the quartile indicators match the observed patterns plotted for the decile indicators in Figure 5.13. The negative effect on achievement from large inferred changes in rank (from the fourth quartile, relative to the omitted quartile) appeared stronger, however, for both Numeracy and Reading, at -0.083 and -0.046 respectively (in the third columns). Note that the estimated effect of inferred changes in rank were smaller than that of local ranks; a within school SD of 0.29 translated to positive effects of 0.126 for Numeracy and 0.097 for Reading.

Table 5.15: Quartiles Indicators of Inferred Change in Rank w/ Cubic Std. Test Scores, Numeracy^

	(1)	(2)	(3)	(4)
Std. Numeracy	0.671***	0.655***	0.752***	0.668***
Std. Numeracy ^2	0.055***	0.038***	0.045***	0.040***
Std. Numeracy ^3	-0.018***	-0.014***	-0.019***	-0.016***
Male	0.006	0.005	0.006	0.005
Foreign Language Background	0.111***	0.112***	0.112***	0.112***
Std. SES	0.076***	0.076***	0.076***	0.076***
Local Rank (t)	0.436***	0.036	quartile^	decile^
Local Rank (t)^2		0.434***	N/A	N/A
Inferred Change in Rank Q1 (t, t+1*)	omitted	omitted	omitted	omitted
Inferred Change in Rank Q2 (t, t+1*)	0.021**	0.017*	-0.001	0.015*
Inferred Change in Rank Q3 (t, t+1*)	-0.002	-0.003	-0.038***	-0.006
Inferred Change in Rank Q4 (t, t+1*)	-0.032***	-0.023*	-0.083***	-0.030**
Constant	-0.354***	-0.288***	-0.149***	-0.287***
N	93,425	93,425	93,425	93,425

^Local rank variables: linear, quadratic, quartiles indicators, deciles indicators
With primary school random effects (RE) and secondary school fixed effects (FE).

Table 5.16: Quartiles Indicators of Inferred Change in Rank w/ Cubic Std. Test Scores, Reading^

	(1)	(2)	(3)	(4)
Std. Reading	0.664***	0.659***	0.727***	0.682***
Std. Reading ^2	0.011***	0.004	0.007**	0.005
Std. Reading ^3	-0.021***	-0.020***	-0.023***	-0.021***
Male	-0.036***	-0.036***	-0.036***	-0.036***
Foreign Language Background	0.030***	0.030***	0.030***	0.030***
Std. SES	0.094***	0.094***	0.094***	0.094***
Local Rank (t)	0.336***	0.137**	quartile^	decile^
Local Rank (t)^2		0.211***	N/A	N/A
Inferred Change in Rank Q1 (t, t+1*)	omitted	omitted	omitted	omitted
Inferred Change in Rank Q2 (t, t+1*)	0.015*	0.012	-0.001	0.008
Inferred Change in Rank Q3 (t, t+1*)	0.007	0.005	-0.019**	-0.003
Inferred Change in Rank Q4 (t, t+1*)	-0.009	-0.006	-0.046***	-0.020*
Constant	-0.271***	-0.234***	-0.120***	-0.191***
N	93,997	93,997	93,997	93,997

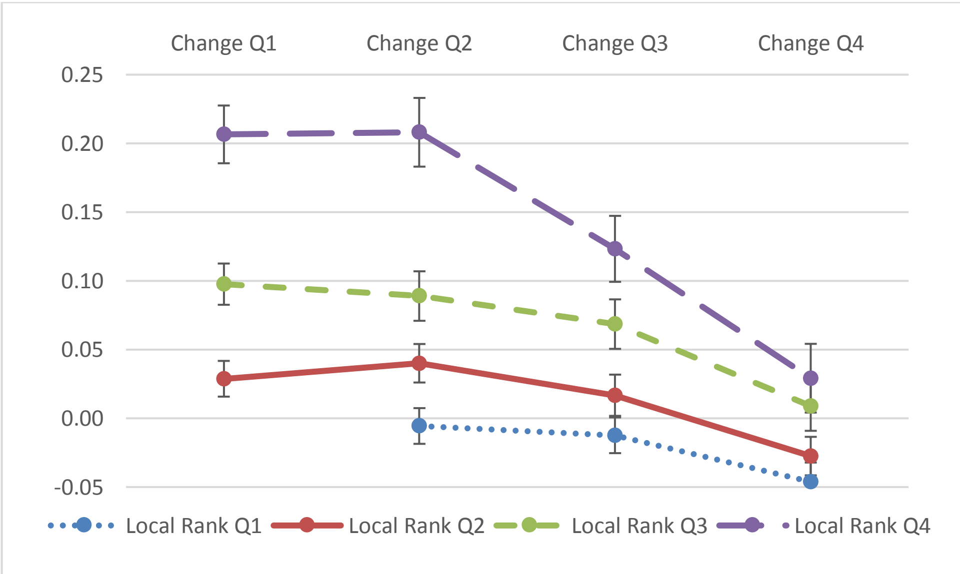
^Local rank variables: linear, quadratic, quartiles indicators, deciles indicators
With primary school random effects (RE) and secondary school fixed effects (FE).

5.6.3.3. Combinations of Local Rank and Inferred Changes

Recognizing the close relationship between local ranks and its changes, I create variables for interactions with 4x4 indicators between quartiles of local rank and quartiles of inferred change in rank. This last set of results shows clearly that the magnitude of the effect on later achievement is much larger for previous local rank than for the changes in rank.

Relative to the significance of local ranks in the earlier time-period (Year 5), the inferred change in rank appears to have a much smaller effect on later achievement. Figure 5.14 and Figure 5.15 show for Numeracy and for Reading the estimates for combinations between quartiles of local rank and quartiles of inferred changes (see Appendix A.6). There are 15 estimates excluding the omitted category, lowest quartile of local rank and lowest quartile in inferred change. There are four lines connecting the points, which represent each local rank quartile, while the horizontal axis measures the quartile of inferred change in rank. The estimates are from the main specification that includes decile indicators for local ranks and a cubic functional form for prior achievement.

Figure 5.14: Estimates of Combinations between Local Rank and Inferred Changes, Numeracy



For Numeracy, Figure 5.14 indicates that there are quite large differences in the effect of local rank between the 1st and 4th quartiles (Q1 and Q4) when looking at quartile 1 of inferred changes in rank. The four lines are somewhat parallel with a convergence towards zero, as we move towards the top quartile of inferred changes in rank. This means that the largest effect is for being in the top quartile of local rank and remaining in the first two quartiles of inferred changes, which indicate that the change in rank is negative.

There is a similar pattern for Figure 5.15, which presents the same information for Reading; however, the effect sizes are smaller to begin with, meaning the convergence appears more gradual. Like with Numeracy, having a large increase in inferred rank (fourth quartile) and being in the bottom quartile of earlier local rank (1st quartile) is worse, at -0.040 for Reading (-0.046 for Numeracy), than experiencing a decrease in inferred rank (the 1st quartile is the omitted category).

Figure 5.15: Estimates of Combinations between Local Rank and Inferred Changes, Reading

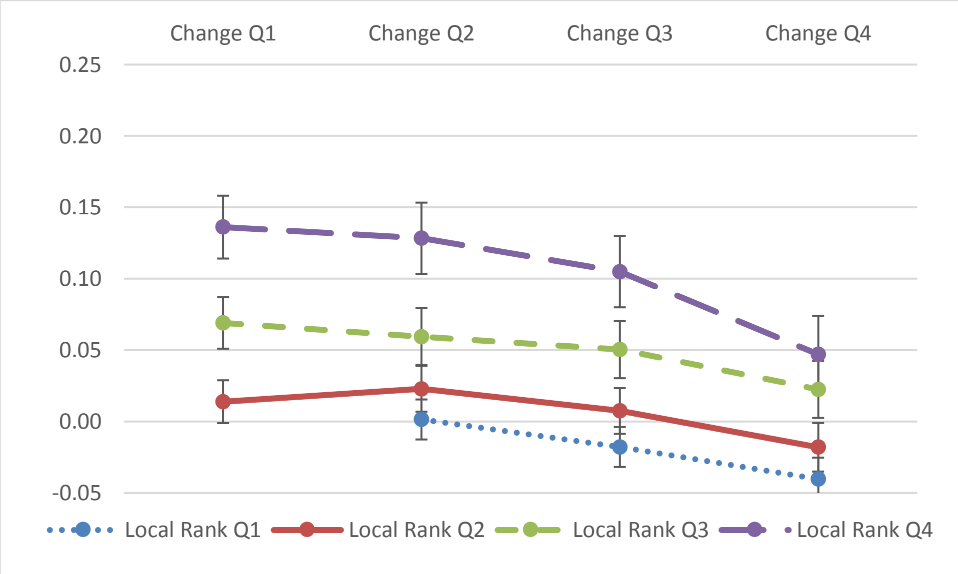


Figure 5.14 also shows that the effect of inferred changes in rank appears to be dependent on student's initial ranks. There were large negative effects in Numeracy between the top and bottom quartiles of inferred changes in rank for students who were in the top two quartiles of local rank, the third and fourth quartiles, at -0.089 and -0.178, respectively. The analogous effects for Reading (see Figure 5.15) were smaller, at a difference of -0.089 between top quartile and bottom quartile of inferred change in rank for the top quartile of local rank, and a difference of -0.046 for the third quartile of local rank. Note that top and bottom quartiles of inferred changes in rank can be interpreted as decreases and increases in peer achievement, respectively.

5.6.4. Goodness-of-fit

For reference, the goodness-of-fit across the different regression specifications are discussed in this section.

Controlling for Primary and Secondary Schools

For a sense of the increase in predictive value from including primary and secondary school (fixed effect) indicators, Table 5.17 and Table 5.18 present the estimates for the plain value-added regressions for Numeracy and Reading, respectively, without local ranks or inferred change in ranks, and their associated adjusted R-squares.

For Numeracy, the adjusted R-squared was 0.658 with the same explanatory variables as before: standardized Numeracy, sex, language background, index of socioeconomic background, and cohort indicators. The inclusion of primary school indicators only increased the goodness of fit to 0.679, which is a difference of 0.021, while the improvement of 0.012 was lower for the inclusion of secondary school indicators, leading to an adjusted R-squared of 0.670. With both primary and secondary school indicators combined, the adjusted R-squared was 0.683.

Table 5.17: Value-Added Regressions with Primary & Secondary School Fixed Effects, Numeracy

	(1)	(2)	(3)	(4)
Std. Numeracy	0.771***	0.775***	0.751***	0.766***
Male	0.005	0.004	0.013**	0.010*
Foreign Language Background	0.169***	0.129***	0.119***	0.118***
Std. SES	0.116***	0.081***	0.081***	0.074***
Cohort Indicators	Y	Y	Y	Y
Primary Indicators	N	Y	N	Y
Secondary Indicators	N	N	Y	Y
Constant	0.055***	-0.025***	-0.121	-0.17
N	93,437	93,437	93,437	93,437
Adj. R-Square	0.658	0.679	0.670	0.683

Increases in the adjusted R-squared for the Reading regressions from the inclusion of primary and secondary school fixed effects were even smaller than for Numeracy. The inclusion of primary school indicators improved the adjusted R-squared from 0.583 to 0.595, which is a difference of 0.012, whereas the difference from the inclusion of secondary school indicators was only 0.09.

Table 5.18: Value-Added Regressions with Primary & Secondary School Fixed Effects, Reading

	(1)	(2)	(3)	(4)
Std. Reading	0.710***	0.707***	0.695***	0.701***
Male	-0.040***	-0.041***	-0.038***	-0.036***
Foreign Language Background	0.063***	0.042***	0.030***	0.033***
Std. SES	0.134***	0.098***	0.096***	0.091***
Cohort Indicators	Y	Y	Y	Y
Primary Indicators	N	Y	N	Y
Secondary Indicators	N	N	Y	Y
Constant	0.102***	0.106***	-0.081***	0.056
N	94,006	94,006	94,006	94,006
Adj. R-Square	0.583	0.595	0.592	0.597

Varying Functional Forms of Local Rank

Table 5.19 shows the adjusted R-squares for the primary school fixed effects equivalents for the different regressions presented for the baseline specifications with local ranks in Table 5.11 and Table 5.12. For Numeracy, the adjusted r-squared is 0.683 without local rank as an explanatory variable, and increases slightly to 0.685, with the inclusion of linear rank, and to 0.687, for quartile indicators of local rank. The decile indicators of local rank provide the highest levels in goodness of fit but are only larger by 0.006 than the plain value-added model. The difference between the plain value-add model and the regression with decile indicators for Reading is also low, at 0.005.

Table 5.19: Adjusted R-Squared for Numeracy and Reading with varying forms for local rank^

	Std. Numeracy	Adj. R-Square	Std. Reading	Adj. R-Square
plain value-add	0.766***	0.683	0.701***	0.597
linear	0.627***	0.685	0.484***	0.602
quadratic	0.583***	0.689	0.477***	0.603
quartile	0.666***	0.687	0.574***	0.601
decile	0.593***	0.689	0.511***	0.602

^ Std. Test Score, Male, Foreign Language Background, Std. SES, with Primary & Secondary Fixed Effects

Decreases in coefficient for prior achievement in Numeracy and in Reading are also apparent with decreases of at least 0.10 for Numeracy, falling from 0.766 (plain value-add) to 0.666 (quartile), and decreases of at least 0.127 for Reading, falling from 0.701 (plain value-add) to 0.574 (quartile).

Quartiles and Deciles of Inferred Change in Rank

Table 5.20 presents the adjusted R-squares for the main specifications with quartile and decile indicators of inferred change in ranks. These correspond to the regression results from Tables 5.13 and 5.14, and those from Figure 5.12, respectively, with the same set of explanatory variables from the baseline specification, along with decile indicators of local ranks and fixed effects for primary school rather than random effects.

The addition of inferred change in rank to the specifications with local ranks had little effect on the goodness of fit of the regressions. The R-squares were unchanged for Numeracy for both quartile and decile indicators of inferred change in rank (from Table 5.19), and were essentially the same for Reading, at 0.602 and 0.603 for quartile and decile indicators of inferred change in rank, respectively.

Table 5.20: Adjusted R-Squared for Quartiles & Deciles of Inferred Change in Rank by Subject^

	Numeracy		Reading	
Inferred change in rank	Quartile	Decile	Quartile	Decile
Adj. R-Square	0.689	0.689	0.602	0.603

^ Std. Test Score, Male, Foreign Language Background, Std. SES, deciles indicators of local rank, with Primary & Secondary Fixed Effects

5.7. Discussion

The main analyses consisted of value-added regressions of inferred changes in ranks, with flexible forms for local ranks and for prior achievement, and regressions with combinations of local ranks and inferred ranks. For the regressions with decile indicators for inferred changes in rank, the effect sizes ranged between 0.02 and -0.065 for Numeracy and between 0.04 and 0.0 for Reading.

The analyses showed for Numeracy that students who experienced large increases in inferred rank had worse achievement in Year 7, with negative estimates of -0.065 and -0.03 for students being in the top two deciles (10^{th} and 9^{th}) respectively. For Reading, the relationship between changes in inferred rank and later achievement was much flatter, with a range of between 0.02 and 0.04 for the 2^{nd} to 9^{th} deciles, suggesting that large positive or negative changes resulted in small negative effects on achievement, reflected by the 1^{st} and 10^{th} deciles.

When the inferred changes in ranks were measured as quartile indicators, there was a benefit of between 0.015 and 0.021 in standardized Numeracy from experiencing changes in the second quartile, depending on functional form for local rank, and a negative effect of around -0.03 from being in the top quartile. For Reading, experiencing large changes in inferred rank, in the top quartile, translated to a small negative effect of -0.02 (relative to the omitted category of first quartile, which reflected large decreases)[135].

When considering combinations of local ranks and inferred changes in ranks together, there were large negative effects for students from experiencing large increases in rank in Numeracy (top quartile of inferred change in rank), at -0.089 and -0.178, for the top two (third and fourth) quartiles of initial local rank, respectively. There were similar but smaller negative effects from experiencing large inferred increases in rank for Reading, for those in the top two quartiles of initial local ranks.

Overall, the results suggest that it is preferable not to experience large inferred increases or decreases in rank for Reading, and that it is preferable not to experience large inferred increases in rank for Numeracy, but that large decreases do not have a large effect on achievement. The results are also supportive of the earlier research that emphasizes the importance of local ranks through its influence on later student achievement.

[135] Note that there was an overlap in Numeracy and Reading samples of over 95% of students and that 46% of students appeared in the same quartile of inferred change in rank in both Numeracy and Reading.

While large decreases in local within-school rank can be readily interpreted as suggesting decreases in self-concept that arises from a diminished group frame-of-reference, the negative impact from large increases in rank are perhaps more difficult to interpret. The drop in performance in academic achievement from an elevation in relative status could possibly be explained by individuals' decision-making in the area of effort allocation. Within a strand of research that concentrates on feedback and task performance, Atkinson (1966) suggests that achievement-motivated individuals are interested in efficiency to accomplish more with less effort.

Taking this line of reasoning further, it is possible that the perceived increases in rank that students experience may lead them to thinking that less effort is required in terms of performance within that subject domain. The increases in rank are artificial in the sense that their underlying performance (or ability) has not changed. This interpretation is also consistent with Atkinson's proposed theoretical framework, in which he defines the optimal level of difficulty as one where success is possible, but the task is still challenging. In support, there is evidence that achievement-oriented individuals respond to feedback of declines in performance but not to performance increases (Schultheiss and Brunstein 2007, describing the findings of the Brunstein and Hoyer's 2002 German paper).

5.8. Robustness; Selection

In this section, I explore the idea that the estimates from the main analyses were driven by sorting of students across schools rather than random variation in local ranks. I first review the background characteristics of students as well as school level characteristics by quartile of inferred change in rank, and secondly estimate further regressions which account for differences in schools as measured by student characteristics.

The main limitation of the estimation approach is that it relies on the assumption that there was random variation in the local ranks that were assigned to individuals as they moved from primary to secondary school. It is probable that there are some specific patterns when students change schools. For example, the earlier descriptive statistics indicated that students generally moved from smaller to larger schools, while there is also likely to be potential sorting of students into schools by SES.

5.8.1. Descriptive Statistics

Analysis of the variance for inferred changes in rank shows that 31.3% of the variance in Year 7 Numeracy can be explained by differences across primary schools, while even less can be attributed to the secondary schools that they attend, at 12.4%. Between percentages of variance from ANOVA for Reading were similar at 29.3% and 12.3% for the primary and secondary school levels, respectively. This is an intuitive observation as one would expect greater variance in local ranks in smaller schools. We saw earlier that on average students moved from primary schools with 34 students to secondary schools with 94 students (see Table 5.2, section 5.5.2), while the standard deviation in the differences between local and global ranks decreased between Year 5 and Year 7, from 0.147 to 0.129 for Numeracy, and from 0.134 to 0.114 for Reading (see Table 5.4, section 5.5.3).

5.8.1.1. Changes in School Level Peer Characteristics

Table 5.21 presents correlations between inferred changes in rank and changes in school level peer characteristics, for each of Numeracy and Reading. There is a slight negative relationship between inferred increases in rank and increases in the size of the school, near -0.2 for both subjects, while the correlation was even larger for increases in mean SES, and larger again for increases in peer level achievement; the correlation was -0.41 for Numeracy and -0.48 for Reading between inferred change in rank and changes in mean SES, and around -0.79 for peer level achievement.

Table 5.21: Correlation between Inferred Changes in Rank & Changes in School Level Variables

	Inferred Change	
	Numeracy	Reading
Change in Number of Students in School	-0.195	-0.190
Change in Std. SES Index^	-0.412	-0.479
Change in Std. Test Score^	-0.790	-0.783
Number of Observations	97,604	97,604

*in both samples. ^ changes based on actual test scores and SES in Year 7.

The high correlation between changes in peer level achievement and inferred changes in rank is unsurprising in that the ranks are calculated from the level of achievement at the schools. The moderate level of correlation between changes in SES and changes in (inferred) rank suggest that sorting on the basis of socioeconomic background may explain some of the magnitude of the estimates from the main results.

In Table 5.22, I take a closer look at schools which were attended by students based on their quartiles of inferred changes in rank, reviewing the mean levels of SES and achievement in

Year 5, in Year 7 (using values from Year 5), and the difference between the two. The standard deviations are shown in parentheses. Table 5.23 repeats this comparison but for local ranks instead of SES and achievement. For reference, other peer characteristics and their changes by quartile of inferred change in rank for Numeracy are included in the Appendix A.7, in Table 5.44 (A) and Table 5.45 (A).

Table 5.22 shows that students in the top quartile experience larger decreases in mean achievement and mean SES, and the opposite pattern for the bottom quartile of inferred change in rank. This is as expected, as larger increases in rank (in the top quartile) suggest that the students attended schools with lower achievement, which is generally correlated with SES. The magnitude of the changes in standardized Numeracy and standardized Reading are similar, as are the values for standardized socioeconomic background.

Table 5.22: Peer Characteristics by Quartile of Inferred Change in Rank, Numeracy & Reading

	Quartiles of Inferred Change in Rank			
	Q1	Q2	Q3	Q4
Std. Numeracy (t)	-0.26 (0.38)	-0.09 (0.37)	0.03 (0.37)	0.24 (0.38)
Std. Numeracy (t+1*)	0.05 (0.39)	-0.04 (0.36)	-0.10 (0.34)	-0.15 (0.33)
Change in Std. Numeracy	0.31 (0.22)	0.05 (0.18)	-0.12 (0.18)	-0.39 (0.24)
Std. SES (t)	-0.13 (0.55)	-0.03 (0.52)	0.00 (0.51)	0.05 (0.54)
Std. SES (t+1*)	-0.04 (0.51)	-0.10 (0.46)	-0.15 (0.46)	-0.20 (0.45)
Change in Std. SES	0.08 (0.33)	-0.07 (0.26)	-0.15 (0.26)	-0.25 (0.33)
Std. Reading (t)	-0.24 (0.36)	-0.08 (0.33)	0.03 (0.32)	0.19 (0.34)
Std. Reading (t+1*)	0.03 (0.34)	-0.05 (0.31)	-0.09 (0.30)	-0.16 (0.30)
Change in Std. Reading	0.27 (0.21)	0.03 (0.17)	-0.12 (0.16)	-0.35 (0.21)
Std. SES (t)	-0.14 (0.55)	-0.04 (0.52)	0.01 (0.52)	0.06 (0.53)
Std. SES (t+1*)	-0.03 (0.52)	-0.11 (0.47)	-0.14 (0.46)	-0.21 (0.45)
Change in Std. SES	0.11 (0.32)	-0.07 (0.26)	-0.15 (0.25)	-0.27 (0.32)

* indicates that the variable is calculated from Year 5 values.

For peer averages of rank of Numeracy, as shown in Table 5.23, the expected pattern was also found, with lower means for the top quartile in Year 5 (time t), and conversely higher means in Year 7 (time t+1* or t+1), given the way that the change in rank is defined, and the quartiles are assigned. The mean values in local rank based on actual test scores (local ranks (t+1)) are consistent with the negative estimates that were produced in the regressions for the top quartile of inferred change in rank. i.e. for the top quartile, the mean local rank of 0.54 is lower than 0.58 which is the local rank based on Year 5 test scores.

Table 5.23: Local Ranks of Peers by Quartile of Inferred Change in Rank, Numeracy

	Quartiles of Inferred Change in Rank			
	Q1	Q2	Q3	Q4
Local Ranks (t)	0.57 (0.25)	0.50 (0.34)	0.47 (0.31)	0.42 (0.23)
Local Ranks (t+1*)	0.44 (0.24)	0.48 (0.34)	0.50 (0.32)	0.58 (0.23)
Local Ranks (t+1)	0.49 (0.26)	0.52 (0.32)	0.52 (0.30)	0.54 (0.26)
Inferred Change in Rank	-0.13 (0.08)	-0.02 (0.02)	0.04 (0.02)	0.16 (0.08)

5.8.1.2. Student Characteristics by Quartile of Inferred Change in Rank

Table 5.24 presents the achievement levels and demographic characteristics of the actual students by quartile of inferred rank of Numeracy, except for the last two rows, which relate to quartiles of inferred rank of Reading. For clarity, only means and standard deviations are shown for Year 5, along with differences between Year 7 and Year 5 for Numeracy and Reading. The percentage of males, percentage foreign language background, percentage indigenous background, the mean levels of SES, were all similar across the four quartiles of inferred change in rank.

Table 5.24: Student Characteristics (YR5) by Quartile of Inferred Change in Rank

	Q1	Q2	Q3	Q4
Male	49.5 (50.0)	52.9 (49.9)	51.9 (50.0)	50.7 (50.0)
Foreign Language Background	29.8 (45.7)	27.2 (44.5)	25.2 (43.4)	26.4 (44.1)
ATSI	1.6 (12.4)	1.5 (12.2)	1.3 (11.2)	1.2 (11.0)
Std. SES	-0.09 (0.97)	-0.08 (0.97)	-0.10 (0.94)	-0.12 (0.93)
Std. Numeracy	-0.12 (0.78)	-0.04 (1.25)	-0.05 (1.05)	0.02 (0.70)
Change in Std. Numeracy^	0.20 (0.01)	0.18 (0.01)	0.11 (0.01)	-0.04 (0.01)
Std. Reading	-0.11 (0.79)	-0.02 (1.27)	-0.05 (1.03)	-0.02 (0.69)
Change in Std. Reading^	0.19 (0.01)	0.13 (0.01)	0.11 (0.01)	0.04 (0.01)

^ changes between Year 5 and Year 7 are calculated using test scores and SES in Year 7

What was notable was decreasing values in the change in Numeracy and change in Reading, with increasing quartile of inferred change in rank. In combination with increasing values in Year 5 achievement, for both Numeracy and Reading, this suggests that on average students in the top quartiles of change had higher initial levels of achievement, and then attended schools with lower peer achievement and lower peer SES (from Table 5.22), and experienced decreases in achievement.

5.8.1.3. Differences between Numeracy and Reading Samples

When comparing quartiles of inferred rank between Reading and Numeracy, 46.1% of students place along the diagonal cells (see Table 5.25), showing a general tendency for students who experienced inferred increases in rank in Numeracy to experience similar changes in Reading. This correlation is consistent with a view that peer achievement across schools is similar for both Reading and Numeracy, with some variation in the strengths and weaknesses in students and in schools.

Table 5.25: Cross-Tabulation of Quartiles in Inferred Change in Rank (%), Numeracy & Reading

Quartiles of Inferred Change in Numeracy	Quartiles of Inferred Change in Reading			
	Q1	Q2	Q3	Q4
Q1	14.2	6.2	3.2	1.4
Q2	6.2	9.1	6.8	3.0
Q3	3.0	6.7	8.7	6.6
Q4	1.6	3.0	6.4	14.1

Observations: 115,159

The difference in effect on achievement from inferred change in ranks between subjects could however be due to underlying differences in the subjects. For example, Numeracy might require greater ongoing effort than Reading in order for students to maintain test score performances, meaning that students might be more sensitive to feedback from local ranks in Numeracy. Recall, that there was a negative effect on achievement from increases in rank for Numeracy and small negative effects for large positive and negative changes in rank for Reading. The regression analyses for Reading also appeared different from those for Numeracy in other ways. Estimates of school level characteristics were smaller in effect size for Reading, as was the predictive power indicated by the R-square values (see section 5.6.4, Goodness-of-fit).

5.8.2. Robustness Regressions

5.8.2.1. Overview

To test the robustness of the results, I undertake the following analyses:

1) First, I re-estimate the main regression model excluding secondary school fixed effects. This analysis provides an indication of the extent that inferred change in ranks indirectly measures school effects.
2) Second, I introduce indicators for transitions between primary and secondary schools, which show whether the estimates of inferred change in rank remain statistically significant. These are essentially the inclusion of interactions of primary and secondary school fixed effect to the value-added regressions.
3) Next, I explicitly include the (school level) peer characteristics in two parts, which also provide an indication of whether the effect of inferred change in ranks can be explained by peer characteristics.
    a. In the first part, I control for peer characteristics, which consists of: language and socioeconomic background, percentage male, school size, and achievement.
    b. In the second part I add the change in peer characteristics to the levels from the first part.

To reflect mean achievement in the later time period, I take the same approach of defining later peer achievement using results from the earlier time-period. Note that I present the proportion male, the proportion foreign language background and divide the number of students in each school (in each year level) by 100 for easier reporting of the coefficients.

5.8.2.2. Specifications

The specifications are presented with notation below to formalize the proposed robustness regression analyses.

1. Without Secondary Fixed Effects

(1) $A_{ik}(\boldsymbol{t}+\mathbf{1}) = a + \beta_1 \boldsymbol{X}_i(\boldsymbol{t}) + \beta_2 A_{ij}(\boldsymbol{t}) + \beta_3 r^l_{ij}(\boldsymbol{t}) + \beta_4 \Delta r^l_{ijk}(\boldsymbol{t}+\mathbf{1}^*) + p_j + c_v + \varepsilon_i$

2. Primary and Secondary Fixed Effects with interactions

First, an interaction term between primary and secondary school indicators of $(\boldsymbol{p_j}\,.\,\boldsymbol{s_k})$ is inserted into the specification from the main analyses earlier, producing (2):

(2) $A_{ik}(\boldsymbol{t}+\mathbf{1}) = a + \beta_1 \boldsymbol{X}_i(\boldsymbol{t}) + \beta_2 A_{ij}(\boldsymbol{t}) + \beta_3 r^l_{ij}(\boldsymbol{t}) + \beta_3 \Delta r^l_{ijk}(\boldsymbol{t}+\mathbf{1}^*) + p_j + s_k + (p_j\,.\,s_k) + c_v + \varepsilon_i$

3. Peer Characteristics

School level peer characteristics and achievement: $\overline{\boldsymbol{X}}_j(\boldsymbol{t})$ and $\overline{\mathbf{A}}_j(\boldsymbol{t})$, are inserted into (2) leading to (3). Specification (3) is analogous to (4.8) from Snidjers and Bosker (2012). The interactions $(\boldsymbol{p_j}\,.\,\boldsymbol{s_k})$ are dropped due to concerns of overfitting and for being too computationally intensive.

(3) $A_{ik}(\boldsymbol{t}+\mathbf{1}) = a + \beta_1 \boldsymbol{X}_i(\boldsymbol{t}) + \beta_2 A_{ij}(\boldsymbol{t}) + \beta_3 r^l_{ij}(\boldsymbol{t}) + \beta_3 \Delta r^l_{ijk}(\boldsymbol{t}+\mathbf{1}^*) + \beta_4 \overline{\boldsymbol{X}}_j(\boldsymbol{t}) + \beta_5 \overline{\mathbf{A}}_j(\boldsymbol{t}) + p_j + s_k + c_v + \varepsilon_i$

4. Changes in Peer Characteristics

This is followed by adding changes in school level characteristics and changes in achievement, $\Delta\overline{\boldsymbol{X}}_{jk}(\boldsymbol{t},\boldsymbol{t}+\mathbf{1})$ and $\Delta\overline{\mathbf{A}}_{jk}(\boldsymbol{t},\boldsymbol{t}+\mathbf{1}^*)$, to (2) to produce (3).

(4) $A_{ik}(\boldsymbol{t}+\mathbf{1}) = a + \beta_1 \boldsymbol{X}_i(\boldsymbol{t}) + \beta_2 A_{ij}(\boldsymbol{t}) + \beta_3 r^l_{ij}(\boldsymbol{t}) + \beta_3 \Delta r^l_{ijk}(\boldsymbol{t}+\mathbf{1}^*) + \beta_4 \overline{\boldsymbol{X}}_j(\boldsymbol{t}) + \beta_5 \overline{\mathbf{A}}_j(\boldsymbol{t}) + \beta_6 \Delta\overline{\boldsymbol{X}}_{jk}(\boldsymbol{t},\boldsymbol{t}+\mathbf{1}) + \beta_7 \Delta\overline{\mathbf{A}}_{jk}(\boldsymbol{t},\boldsymbol{t}+\mathbf{1}^*) + p_j + s_k + c_v + \varepsilon_i$

Note: random effect versions of (3) and (4) for both primary and secondary school were also estimated but not presented, with $(1|p_j) + (1|s_k)$ in place of the school fixed effects: $p_j + s_k$. School random effects reflect the scenario where there are few idiosyncrasies in schools, and where school choice does not influence the change in ranks[136]. As such, the estimates for inferred change in ranks can be thought of as upper bounds, which are likely to be unrealistic

[136] Analyses from multi-level modelling were similarly not included because the normality assumption imposed by the random slopes and intercepts regression model is somewhat dubious for the school effects, particularly at the secondary level where there is a tendency for the schools to specialize, and with parents and students appearing to place greater emphasis on the school they attend.

given the earlier descriptive analyses, which showed some sorting of students across the different quartiles (see section 5.8.1).

5.8.2.3. Without Secondary Fixed Effects (1)

While the regressions in the main analyses control for the influence of secondary school environment, both the inferred change in rank and the secondary school attended are likely inter-related. The choice of secondary school attended affects the peers that a student has and subsequently their local ranks. As noted in the methods section (Regression Model, section 5.4.1.2), secondary school indicators were included to address the possibility that the estimates of inferred change in rank are picking up some form of specific individual school effects.

Table 5.26 presents by subject estimates of these quartile and decile indicators of inferred change in rank, for the regression specification without secondary school fixed effects and with decile local ranks. Consistent with the possibility that inferred change in rank also reflects individual school effects, excluding the secondary school fixed effects from the regressions leads to larger estimates of the inferred change in ranks, at up to -0.154 for the top quartile for Numeracy, and at up to -0.119 for the top quartile for Reading. The impact on the decile indicators of inferred change in rank was similar, with larger estimates of up to -0.250 and -0.176 for the top decile of inferred change in ranks, for Numeracy and Reading respectively.

Table 5.26: Quartiles and Deciles of Inferred Change in Rank, without Secondary Fixed Effects ^

Quartile	Numeracy	Reading	Decile	Numeracy	Reading
1	omitted	omitted	1	omitted	omitted
2	-0.033*	-0.038*	2	-0.039***	-0.019*
3	-0.080***	-0.073***	3	-0.061***	-0.042***
4	-0.142***	-0.122***	4	-0.063***	-0.043***
			5	-0.080***	-0.065***
			6	-0.107***	-0.084***
			7	-0.115***	-0.091***
			8	-0.142***	-0.110***
			9	-0.173***	-0.125***
			10	-0.241***	-0.176***
N	93,425	93,997	N	93,425	93,997

^ Explanatory variables of Std. Test Score, Male, Foreign Language Background, Std. SES and deciles indicators of local rank not shown.

5.8.2.4. Primary and Secondary School Interactions (2)

Table 5.27 presents the results of the fixed effects specification with primary and secondary school interactions. I take a 10% sample to reduce the initial 13,759 combinations of interactions between primary and secondary schools, across the four cohorts, down to 3,817. With samples of around 90,000, the number of interactions is too computationally intensive for STATA and is also likely to result in overfitting of the primary to secondary school transitions due to likely possibility that many of the transitions are idiosyncratic, reflecting the experiences (or value-added relationships) for a small number of students. As per the main analyses, I apply clustered standard errors on primary school, noting that there is likely to be unobserved correlation in the outcomes of students at the same primary schools.

Table 5.27: Inferred Change in Rank with Primary to Secondary School Interactions^

	Numeracy (t+1)	Reading (t+1)
Std. Test Score	0.705***	0.637***
Std. Test Score^2	0.036***	-0.019
Std. Test Score^3	-0.014***	-0.012*
Male	0.009	-0.048**
Foreign Language Background	0.102***	0.055*
Std. SES	0.068***	0.072***
Inferred Change in Rank Q1 (t, t+1*)	omitted	omitted
Inferred Change in Rank Q2 (t, t+1*)	-0.001	-0.011
Inferred Change in Rank Q3 (t, t+1*)	-0.017	-0.032
Inferred Change in Rank Q4 (t, t+1*)	-0.038	-0.046
constant	-4.860***	-0.898*
N	9,345	9366
Adjusted R-square	0.652	0.567

^ Cohort and Primary to Secondary School Interaction Fixed Effects.
10% sample. Decile indicators of local rank (t) not shown.
Std. Test Score refers subject domain indicated by column name at time t.

Unlike the earlier results which showed an initial positive effect with increasing quartile of inferred change in rank for Numeracy, there is a pattern of increasingly negative estimates. The slight benefit from experiencing a decline in inferred change in rank (being in the second quartile), between 0.015 and 0.02 depending on functional form of local rank, is not evident. The negative estimate for the fourth quartile is consistent with the earlier results, at -0.065 from the main analyses, but is smaller in size, at -0.038. The estimates, however, are not statistically significant, which can probably be attributed to the limited sample size and also from the interactions soaking up the variation in later achievement.

5.8.2.5. Peer Characteristics (3)

The regression results from the addition of school level peer characteristics of proportion male, proportion foreign language background, and number of students in year level, along with mean values in achievement and SES, are shown in Table 5.28 for Numeracy and Table 5.29 for Reading. Primary and secondary school fixed effects are adopted as before.

For both Numeracy and Reading, in columns 1 and 2 the pattern in results is similar to the previous regression with interactions between primary and secondary school indicators, but the negative estimate for experiencing large inferred changes in rank is now statistically significant, at around -0.030 for Numeracy, and around -0.028 for Reading.

For the peer characteristics, the regression produced a positive statistically significant estimate of 0.059 for a 1-SD increase in the SES index value, while the proportion male, proportion foreign language background and number of students were all indistinguishable from zero. The positive effect from peer SES is consistent with the conventional interpretation of peer effects, which has a positive relationship with achievement.

Table 5.28: Inferred Change in Rank with Peer Characteristics (Fixed Effects), Numeracy^

	(1)	(2)	(3)	(4)
Inferred Change in Rank Q1 (t, t+1*)	omitted	omitted	omitted	omitted
Inferred Change in Rank Q2 (t, t+1*)	0.008	0.008	0.021**	0.020**
Inferred Change in Rank Q3 (t, t+1*)	-0.012	-0.012	0.014	0.013
Inferred Change in Rank Q4 (t, t+1*)	-0.029**	-0.030**	0.015	0.015
School Level Chars				
Number of Students	0.042			0.026
Male (proportion)		-0.029		0.030
Foreign Language Background (prop.)		0.041		0.060
Peer SES		0.059*		0.120***
Peer Achievement			-0.356***	-0.368***
N	93,425	93,425	93,425	93,425

^Local Rank (t) deciles, cohort indicators, primary school and secondary school fixed effects.

For column 3 in Table 5.28 for Numeracy, when the peer level of achievement from Year 5 is included, the negative estimates for the top two quartiles of inferred change in rank become positive but are not statistically significant. With peer level of achievement included, being in the second quartile improves Year 7 test scores by 0.021, relative to the omitted first quartile. The estimate for peer achievement is negative, consistent with the BFLPE and with rank effects, where higher (percentile) ranks are associated with improved later achievement. i.e. higher peer achievement is associated with lower individual ranks.

Given that a high peer level of achievement necessarily decreases the likelihood of decreases in rank, and increases the likelihood of increases in inferred rank, then the negative effect in peer level can be viewed as offsetting the negative estimates of inferred change in rank from columns 1 and 2, which do not account for peer achievement. In column 4 when peer SES and peer achievement are both included, the estimates of inferred change in rank largely remain the same, and the direction of the coefficients for peer SES (positive) and peer achievement (negative) remain consistent with expectation.

The results for peer achievement and peer SES in columns 3 and 4 for Reading in Table 5.29 are similar to those from Numeracy, though none of the quartile estimates are statistically significant and the magnitudes in coefficients for peer SES and peer achievement are smaller than for Numeracy.

Table 5.29: Inferred Change in Rank with Peer Characteristics (Fixed Effects), Reading^

	(1)	(2)	(3)	(4)
Inferred Change in Rank Q1 (t, t+1*)	omitted	omitted	omitted	omitted
Inferred Change in Rank Q2 (t, t+1*)	-0.001	-0.001	0.006	0.006
Inferred Change in Rank Q3 (t, t+1*)	-0.012	-0.013	0.002	0.002
Inferred Change in Rank Q4 (t, t+1*)	-0.027**	-0.028**	-0.004	-0.003
School Level Chars				
Number of Students	0.018			0.017
Male (proportion)		0.005		-0.014
Foreign Language Background (prop.)		0.048		0.027
Peer SES		0.03		0.069**
Peer Achievement			-0.214***	-0.222***
N	93,997	93,997	93,997	93,997

^Local Rank (t) deciles, cohort indicators, primary school and secondary school fixed effects.

5.8.2.6. Changes in Peer Characteristics (4)

In Table 5.30 and Table 5.31, there is little change in the estimates of inferred change in ranks for Numeracy or Reading from the addition of changes in school level characteristics to the school fixed effects regressions. The negative effect on Year 7 achievement from experiencing an inferred change in rank in the top quartile was still around -0.03 for Numeracy and -0.028 for Reading. Once again, the negative estimate for the top quartile of inferred change in rank became positive and statistically indistinguishable from zero from the addition of peer level achievement in Year 5 (discussed in the previous section, section 5.8.2.5).

Only the estimate for the change in the number of students for Reading was statistically significant among the change in peer variables, with increases in student numbers having a negative effect on later achievement, by 0.026 for an increase of 100 students. For the direction of the estimates of changes in peer characteristics, there was a similar negative estimate of -0.021 from an increase in student numbers for Numeracy, while increases in the proportion male was positive, at 0.031, for Numeracy, and negative, at -0.033, for Reading. Increases in the proportion of students from a foreign language background produced positive estimates of 0.081 and 0.067 for Numeracy and Reading, respectively. Increases in peer SES were similarly positive but with smaller magnitudes for both subjects.

Note that the inclusion of all five school-level variables and their changes did not provide any additional insights in either school fixed effects or random effects specifications (discussed next) and were not presented for brevity.

Table 5.30: Inferred Change in Rank with Peer Characteristics & Changes (FE), Numeracy^

	(1)	(2)	(3)	(4)
Inferred Change in Rank Q1 (t, t+1*)	omitted	omitted	omitted	omitted
Inferred Change in Rank Q2 (t, t+1*)	0.008	0.009	0.009	0.021**
Inferred Change in Rank Q3 (t, t+1*)	-0.012	-0.012	-0.011	0.015
Inferred Change in Rank Q4 (t, t+1*)	-0.029**	-0.030**	-0.029**	0.017
School Level Characteristics				
Number of Students	0.027			
Male (prop.)		0.026		
Foreign Language Background (prop.)		0.111		
Peer SES			0.111*	
Peer Achievement				-0.349***
School Level Change				
ΔNumber of Students	-0.021			
ΔMale (prop.)		0.031		
ΔForeign Language Background (prop.)		0.081		
ΔPeer SES			0.060	
ΔPeer Achievement				0.009
N	93,425	93,425	93,425	93,425

^Local Rank (t) deciles, cohort indicators, primary sch. random effects & secondary sch. fixed effects (FE)

Table 5.31: Inferred Change in Rank with Peer Characteristics & Changes (FE), Reading^

	(1)	(2)	(3)	(4)
Inferred Change in Rank Q1 (t, t+1*)	omitted	omitted	omitted	omitted
Inferred Change in Rank Q2 (t, t+1*)	-0.001	-0.0005	-0.0002	0.004
Inferred Change in Rank Q3 (t, t+1*)	-0.013	-0.012	-0.012	-0.001
Inferred Change in Rank Q4 (t, t+1*)	-0.028**	-0.027**	-0.027**	-0.008
School Level Characteristics				
Number of Students	-0.001			
Male (prop.)		-0.023		
Foreign Language Background (prop.)		0.107		
Peer SES			0.056	
Peer Achievement				-0.238***
School Level Change				
ΔNumber of Students	-0.026*			
ΔMale (prop.)		-0.033		
ΔForeign Language Background (prop.)		0.067		
ΔPeer SES			0.031	
ΔPeer Achievement				-0.031
N	93,997	93,997	93,997	93,997

^Local Rank (t) deciles, cohort indicators, primary sch. random effects & secondary sch. fixed effects (FE)

Note that the regressions results from the specifications with secondary school random effects and with levels and changes in peer characteristics (not shown) produced more statistically significant estimates for school level variables and their changes than the analogous fixed effects specifications, in Table 5.30 and Table 5.31. This is to be expected with the fixed effects absorbing a greater proportion of the variance in achievement. The estimates of inferred change in rank, from the random effects with levels and changes, were similar to the regressions with secondary school fixed effects and peer characteristics, but were smaller in magnitude, ranging between -0.106 and 0.022 for Numeracy, and ranging between -0.104 and -0.009 for Reading.

5.9. Limitations

This section focuses on the potential of peer effects to explain the main result and a second concern relating to non-random variation in school choice and functional form of the regression specifications.

5.9.2. Peer Effects & Local Ranks

The methodological limitations relating to peer effects within the literature on educational achievement have been well documented (e.g. Ryan 2017 and Feld and Zolitz, 2017), the most serious of which relates to the reflection problem such that individual and peer achievement simultaneously influence each other (see Angrist 2014, and earlier Manski, 1993). More broadly, the recent research has collectively shown that the peer effects on achievement are modest (Sacerdote, 2011), as noted in the relevant literature (section 5.2.1).

As this study focuses on the effect of changes in rank on later achievement, there are two relevant time periods. For the later time-period, the reflection problem is avoided by using previous tests scores to calculate later ranks, meaning that this definition of later ranks does not measure the influence that individuals and their peers have on each other in the new school environment. McVicar, Moschion and Ryan (2016) adopt a similar approach, estimating the effect on achievement from peer quality as measured by achievement from a previous time-period that also reflects a different school environment.

The reflection problem may, however, positively influence the estimated local ranks in the initial time-period, with local ranks, peer achievement and test scores measured together. Murphy and Weinhardt noted a similar concern that peer quality jointly determined student rank and prior achievement[137]. This point posits a positive relationship between peer and individual achievement such that lower peer quality in primary school negatively influences the individual's achievement[138]. Lower peer achievement additionally equates to high local ranks, with local ranks calculated within school. This in turn, results in the appearance that students' achievement gain is higher from a change in peer environment for students with higher ranks.

If we assume positive peer effects on achievement and assume that students in the middle of the distribution are most sensitive to changes in local ranks, then it appears that the local rank effect on later achievement could be equivalent to a peer effect narrative. i.e. moving from a low peer to high peer achievement environment could be generating the positive local rank effect, with low peer achievement environments first pulling down test scores and high peer achievement environments later pulling up test scores, and vice versa. Positive peer effects then correspond to initially having high local ranks, while negative peer effects would correspond to having low local ranks, which together produce the positive estimate.

[137] Appendix 2, CEP Discussion Paper 1241. It seems reasonable to suggest that school environment jointly influences student and peer achievement and that selection across schools creates the appearance that peer achievement quality influences achievement.
[138] For example, McVicar, Moschion, and Ryan (2013) find within-year level (endogenous) peer effects of magnitude 0.10 SD for a 1 SD increase peer achievement, using similar data. For reference, McVicar, Moschion and Ryan (2016) estimate a general effect of 0.8 percentile increase in rank for a 10 percentile rank increase in peer quality (near 1 SD**)**.

The critique appears to have merit for the interpretation of positive effects for local ranks, however, this study produced estimates of the inferred change in ranks that are only partially supported by a positive peer effects narrative. When students move from a high to low peer achievement environment, thus experiencing an increase in local ranks, the effect on later achievement is negative, as predicted by a positive peer effect on achievement. However, when students move from a low to high peer achievement environment and consequently experience perceived decreases in local ranks, there is largely an absence of effect and most definitely not a positive effect, as would be predicted from the conventional literature relating to the effect of peer effects on academic achievement. That is, the interpretation that moving to higher peer achievement environments is beneficial for later academic performance is not supported by the analyses in this study, while there does appear to be a negative effect from moving to low peer achievement environments with the caveat that the study has a focus on short-term achievement.

There is support from the robustness analyses that estimates of local ranks (and changes in ranks) and mean peer achievement cannot be cleanly interpreted because they measure the same underlying within-school achievement distributions. Under the fixed effects specifications that accounted for peer achievement (Table 5.28, Table 5.29, Table 5.30 and Table 5.31), there was no negative effect from inferred increases in rank (top quartile estimates), unlike what was suggested by the negative estimates from the main results. Counter to the positive peer effects narrative, these regressions produced negative estimates of peer achievement in primary school, indicating that there was a negative association between attending a high peer achievement primary school and later achievement.

5.9.3. Selection and Functional Form

The second main limitation was that some of the estimated effects from inferred change in rank is likely to reflect systematic changes in peer achievement and peer SES that are intentional on the part of students and parents, rather than random variation in school choice. For example, students in the top quartile of perceived change in ranks had higher levels of achievement and experienced greater decreases in peer achievement and peer SES. In the robustness section, I investigated this issue and sought to address the selection by accounting for school level differences and the changes in school environment experienced by students between primary and secondary school. In addition, the fixed effects regressions with school level achievement are also limited in their ability to reliably disentangle the correlated relationship between school level and individual achievement.

In relation to the possible limitation from relying on analyses of only students from the government sector, I noted in the descriptive statistics (see Table 5.3) that was not problematic because the approach relies on variation generated by differences in peer level achievement, which still exists within the government sector, and also because the local ranks are calculated within school. Table 5.3 also showed that the degree of selection, by either SES or achievement, of students moving from the government sector to the non-government sector was not especially large.

An important limitation is the level of abstraction involved in estimating and interpreting the effect of inferred change in rank. The key variable of interest relies on differences in local ranks between two time periods and different school environments, where the later rank is calculated from test score achievement from the earlier time-period.

While the issue was that inferred change in rank is likely to be correlated with peer level achievement between time periods, another complexity in the regression estimation was determining the functional forms of local rank and prior achievement, given that local ranks are to some degree dependent on prior achievement. I.e. high-achieving students will generally have top ranks unless they are attending schools with unusually high levels of peer achievement.

In the results, I concentrated on varying specifications for the functional forms of each of local ranks and achievement to account for the effect of estimates of the inferred change in rank. The main specification with a linear form of prior achievement produced a positive effect on later achievement from increases in inferred changes in rank for Reading, and an inverse U-shape for Numeracy. After adjusting for non-linear forms of prior achievement, the pattern in inferred change in rank for Reading became much flatter, bearing a closer resemblance to the results for Numeracy.

Further, the specification problem also applied to the inferred change in rank itself, as it was not apparent what the expected effect should be. For example, one possible interpretation was that a positive increase in rank should improve later achievement. In consideration of the possible interdependence between initial ranks and the inferred changes, I also estimated combinations between quartiles of each variable.

Finally, it is likely that the timing of when the tests were taken also has an impact on the interpretation of the magnitude of the estimates. The standardized tests are taken around May which is several months after the start of the school year in February, meaning that students have had a significant period of exposure to their new peers. Given recent research it is likely that the rank effects from both previous and later environments are exerting an influence on students; Becker and Neumann (2016) recently showed that the Big-Fish-Little-Pond-Effect did not appear to persist very long, with the positive BFLPE from primary school fading after one year after attending their new school.

5.10. Conclusion

In this study I propose an approach that extends Murphy and Weinhardt (2014) to estimate the effect on later achievement from a change in local ranks for students when they change school environments. I interpret the local ranks as reflecting academic status and calculate the within-school ranks in both time periods from the earlier test scores. The importance of relative status, which is often reflected by local ranks, has been posited as a fundamental human motive (Anderson et al. 2015), and there is evidence that individuals vigilantly monitor their status in various contexts.

This study found that large perceived increases in rank, which were arguably unexpected and attributable to random variation, tended to negatively affect students' later achievement. In the analyses, I found that large inferred increases in rank (top quartile) negatively affected later achievement by between -0.023 and -0.083 on standardized test scores in Numeracy and found a similar negative effect of up to -0.046 for students experiencing large changes in rank for Reading.

When estimated as decile indicators, the pattern in inferred change in rank for Numeracy was consistent with the quartile estimates, while the pattern was much flatter for Reading, suggesting small negative effects for the top and bottom deciles. Large inferred increases in rank of around 24 percentiles had a significant negative effect for Numeracy of up to -0.065 SD for the top decile of change, while there were only small negative effects for the top and bottom deciles for Reading, at -0.02 SD.

Estimates of combinations in local ranks and inferred change in ranks showed that large increases had a greater impact for students who were in the top quartile of local ranks in Year 5. Overall, the effect from a perceived change in rank was smaller than the influence from the initial local rank, the estimates of which were consistent with the Big-Fish-Little-Pond-Effect (BFLPE) and Murphy and Weinhardt (2014).

The main limitation of the approach was that it relied on random variation in the changes in rank when students move from primary to secondary school, although it is probable that students sort into secondary school. For example, students in the top quartile of perceived change in ranks had higher levels of achievement and experienced greater decreases in peer achievement and peer SES. A related concern is that the main result is consistent with a traditional positive peer effects story, whereby the negative effect on achievement could be explained by students initially having high-achievement peers (low ranks) and low-achievement peers (high ranks) in the later environment.

To address this concern, additional regression analyses accounted for peer level achievement and peer level SES. The fixed effects specifications suggested that the negative effects from large inferred increases in rank in both Numeracy and Reading were closer to -0.03 for being in the top quartile of change. These effects were similar in size to the effects of being male for Reading, at -0.035, or from having foreign language background at 0.03.

The study of academic selection and academic achievement intersects several concepts in educational psychology. Extending the metaphor offered by the BFLPE, students generally move from smaller to bigger ponds by virtue of attending smaller primary schools and larger secondary schools, and there is also a tendency for some students to sort into schools on a socioeconomic or academic basis.

The comparisons that students make with reference to their peers have been shown to influence their beliefs in their own academic ability, and also appear to be unsurprisingly context and time dependent. Becker and Neumann (2016) showed that the effect on self-concept from the BFLPE diminished in new school environments after one year and was replaced by a new one. There are indications that relative performance in academic achievement may also influence individuals' longer term decision-making regarding subject choices and career specialization, particularly when the results re-enforce race or gender stereotypes[139].

This study found evidence of immediate short-term impacts on academic achievement from changes in academic status, and also provided new evidence of short-term effort allocation, particularly for Numeracy. The results provide new evidence of self-referencing behaviour, where individuals make comparisons with their past performances (which could be relative performances, like local ranks), over short time horizons. For instance, moderate decreases in local rank may prompt students to try harder to maintain their previous standing or self-referenced benchmark. The behaviour of reduced effort allocation from students who experienced large perceived increases in rank makes sense in a context where the new information was unexpected and attributable to random variation.

[139] See section 2.3.4 on Self-Identity and Academic Specialization in Chapter 2: Literature Review.

Appendix

A.1. Simplified Description of Methodology

A simplified overview of the steps building up to the main relationship is presented here (details of the regressions are presented in the main text); I specify the relationships as enumerated relationships (in bold), describing the positive and negative associations between the key variables, rather than as regression specifications.

Marsh and Hau (2003) estimate the Big-Fish-Little-Pond effect as a negative relationship between individual and group achievement; $\mathrm{Self\ Concept_{it}}$ is self-concept for individual *i,* which is negatively related to the group average ($\overline{\mathrm{A}}_{\mathrm{jt}}$ for school *j*):

$$\mathrm{Self\ concept_{it}} \propto -\overline{\mathrm{A}}_{\mathrm{jt}} \qquad \textbf{[a]}$$

Murphy and Weinhardt (2014) extend Marsh by introducing within-school local ranks (r^{l}), which I follow. For the achievement outcome, they use global ranks (r^{g}) and estimate the effect of local rank (${r^{\mathrm{l}}}_{\mathrm{t}}$) on later achievement (${r^{\mathrm{g}}}_{\mathrm{t+1}}$). I call ${r^{\mathrm{l}}}_{\mathrm{t}}$ the rank effect, like Murphy and Weindhardt. Note: *i* subscripts are omitted for clarity.

$$A_{\mathrm{t+1}} \propto A_{\mathrm{t}} + {r^{\mathrm{l}}}_{\mathrm{t}} \qquad \textbf{[b]}$$

Next, I introduce local ranks over different groups of students (schools *j* and *k*) for time periods *t* and *t + 1,* calling them *previous* local ranks and *inferred* local ranks. For notation, the superscript asterisk indicates that ranks are calculated from previous achievement.

${r^{\mathrm{l}}}_{\mathrm{jt}}$ for school *j*; and $r^{\mathrm{l}}_{\mathrm{k(t+1)^*}}$ for school *k*.

We can now introduce the variable of interest: the inferred change in local rank which reflects changes in relative status. The difference between inferred rank in *t+1* and previous rank in *t* is then simply:

$$\Delta r^{\mathrm{l}}_{\mathrm{t,t+1^*}} = r^{\mathrm{l}}_{\mathrm{k(t+1)^*}} - {r^{\mathrm{l}}}_{\mathrm{jt}}$$

Adding this change in rank term to the Murphy and Weinhardt leads to the following relationship for student achievement (school subscripts are assumed and omitted):

$$A_{\mathrm{t+1}} \propto A_{\mathrm{t}} + {r^{\mathrm{l}}}_{\mathrm{t}} + \Delta r^{\mathrm{l}}_{\mathrm{t,(t+1)^*}} \quad \textbf{[c]}$$

Equation [c] is the relationship of interest in this chapter.

As an aside, note that substituting for $\Delta r^{\mathrm{l}}_{\mathrm{t,(t+1)^*}}$ into [c] produces: $A_{\mathrm{t+1}} \propto A_{\mathrm{t}} + r^{\mathrm{l}}_{\mathrm{k(t+1)^*}}$

The point of the disaggregation, however, is that individuals experience the change in school environment as two effects: an existing relative status, which informs their academic self-concept, and a subsequent perceived update (or change). This accounts for the time dimension which complicates the analyses in this area. In other words, the effect from local ranks (via self-concept) is a delayed effect. Under the conceptual framework of relative status and self-concept, a regression equation that isolates $r^{\mathrm{l}}_{\mathrm{k(t+1)^*}}$ is more similar to including only $r^{\mathrm{l}}{}_{\mathrm{t}}$, like Murphy and Weinhardt.

A.2. Descriptive Statistics

Table 5.32 (A): School Level Characteristics, Student Weighted^

	Grade 5	Year 7	Change
Number of Students	63.0 (27.5)	183.5 (74.8)	120.5 (73.9)
Male	51.8 (9.1)	52.3 (11.0)	0.7 (12.6)
Foreign Language Background	25.6 (24.5)	25.5 (23.5)	-1.2 (14.2)
Std. SES Index	-0.06 (0.53)	0.04 (0.50)	0.06 (0.34)
ATSI	1.7 (3.8)	1.8 (2.5)	0.3 (2.7)
Age	10.5 (0.10)	12.5 (0.1)	2.01 (0.08)
NAPLAN Scores			
Std. Numeracy	-0.02 (0.42)	0.02 (0.41)	0.04 (0.36)
Std. Reading	-0.02 (0.37)	0.02 (0.35)	0.04 (0.32)
Missing Values – Numeracy	6.7 (5.2)	7.4(4.3)	0.7 (6.2)
Missing Values – Reading	6.5 (5.2)	7.2 (4.4)	0.7 (6.4)
Both Missing	5.5 (4.8)	5.0 (3.4)	-0.4 (5.4)

^ students from both Reading and Numeracy samples.

A.3. Local Rank Estimates with Cubic Standardized Test Score

Table 5.33 (A): Quartiles Indicators of Inferred Change in Rank, Numeracy^

	(1)	(2)	(3)	(4)
Std. Numeracy	0.671***	0.655***	0.752***	0.668***
Std. Numeracy^2	0.055***	0.038***	0.045***	0.040***
Std. Numeracy^3	-0.018***	-0.014***	-0.019***	-0.016***
Male	0.006	0.005	0.006	0.005
Foreign Language Background	0.111***	0.112***	0.112***	0.112***
Std. SES	0.076***	0.076***	0.076***	0.076***
Local Rank (t)	0.436***	0.036	quartile^	decile^
Local Rank (t)^2		0.434***	N/A	N/A
Constant	-0.354***	-0.288***	-0.149***	-0.287***
N	93,425	93,425	93,425	93,425

^Local rank variables: linear, quadratic, quartiles indicators, deciles indicators

Table 5.34 (A): Quartiles Indicators of Inferred Change in Rank, Reading^

	(1)	(2)	(3)	(4)
Std. Reading	0.664***	0.659***	0.727***	0.682***
Std. Reading^2	0.011***	0.004	0.007**	0.005
Std. Reading^3	-0.021***	-0.020***	-0.023***	-0.021***
Male	-0.036***	-0.036***	-0.036***	-0.036***
Foreign Language Background	0.030***	0.030***	0.030***	0.030***
Std. SES	0.094***	0.094***	0.094***	0.094***
Local Rank (t)	0.336***	0.137**	quartile^	decile^
Local Rank (t)^2		0.211***	N/A	N/A
Constant	-0.271***	-0.234***	-0.120***	-0.191***
N	93,997	93,997	93,997	93,997

^Local rank variables: linear, quadratic, quartiles indicators, deciles indicators

A.4. Semi-nonparametric Estimation of Local Rank

In this appendix, I estimate semi-nonparametric regressions to identify if there are functional form equivalents to the non-parametric decile indicators of local ranks used in the main analyses. These results have been omitted from the main text for ease of comprehension, with decile indicators of local rank being adopted for the main analyses.

First, Robinson's (1988) double residual estimator, allows for the estimation of the explanatory variables while removing the influence of the variable of interest, local ranks. Anders Munk-Nielsen provides an intuition for the estimator[140] in that the influence of the variable of interest (in this case the local rank) is subtracted from the dependent and explanatory variables, which residualizes both types of variables (hence the name). The coefficients of the parametric variables can then be estimated from a simple ordinary least squares regression from the residualised variables; the predictions of dependent and explanatory variables from the non-parametric variable can be estimated from local constant or local linear regressions.

Table 5.35 (A) presents the estimates from the value-added regressions from Robinson's (1988) double residual estimator, using Veradi and Debarsy's (2012) semipar STATA command. The difference in estimate for achievement is less than 0.05, compared with the estimates from Tables 6.6 and 6.7, meaning that the decile indicators are not too dissimilar from the semi non-parametric approach.

Table 5.35 (A): Robinson's double residual estimator

	Numeracy	Reading
Std. Test Score	0.551***	0.507***
Male	0.018	-0.013
Foreign Language Background	0.097***	0.041*
Std. SES	0.060***	0.079***
N	9,343	9,375

Cohort indicators, primary and secondary school fixed effects.

10% sample due to computational demands

Table 5.36, Table 5.37, and Table 5.38 (A), reflecting linear, quadratic and cubic forms of standardized test scores, present the results from Hardle and Mammen tests, which test through bootstrapping simulations whether the functional forms are equivalent to the non-parametric forms of local rank. From Table 5.36 (A), the null hypothesis that the two functional forms in the Numeracy regressions were equivalent was rejected for all three forms of local

[140] The Method of Sieves, Lecture Note, Advanced Microeconometrics, Fall 2016, version 1.1. Department of Economics, University of Copenhagen (website): http://www.econ.ku.dk/munk-nielsen/miscellaneous.htm, accessed 5 December 2017.

ranks (linear, quadratic and cubic). The p-value of 0.14 for the quadratic form of local rank in the Reading regression indicated that the quadratic form of local rank is not statistically different from the semi non-parametric form of local rank.

Table 5.37 and Table 5.38 (A), which included quadratic and cubic forms of prior achievement, respectively, showed that the quadratic forms of local rank were equivalent to the semi non-parametric forms, according to the Hardle and Mammen tests, and also showed that cubic forms of local rank were similarly equivalent for the Numeracy regressions.

Table 5.36 (A): Hardle and Mammen; Linear Std. Test Score

		form of local rank		
		linear	quadratic	cubic
Numeracy	T-Statistic	1.96	7.60	2.81
	P Value > 0.05	0.00	0.00	0.00
Reading	T-Statistic	5.23	1.00	1.12
	P Value > 0.05	0.00	0.14	0.24

Table 5.37 (A): Hardle and Mammen; Quadratic Form of Std. Test Score

		form of local rank		
		linear	quadratic	cubic
Numeracy	T-Statistic	20.45	1.57	1.19
	P Value > 0.05	0.00	0.06	0.16
Reading	T-Statistic	13.59	1.15	1.02
	P Value > 0.05	0.00	0.23	0.32

Table 5.38 (A): Hardle and Mammen; Cubic Form of Std. Test Score

		form of local rank		
		linear	quadratic	cubic
Numeracy	T-Statistic	11.96	3.35	1.01
	P Value > 0.05	0.00	0.01	0.23
Reading	T-Statistic	9.84	0.58	0.73
	P Value > 0.05	0.00	0.63	0.57

A.5. Deciles of Inferred Change in Rank

Table 5.39 (A): Decile Indicators of Inferred Change in Rank w/ different functional forms of local rank, Numeracy & Reading

	Reading		Numeracy	
	(1)	(2)	(3)	(4)
Std. Test Score	0.446***	0.492***	0.581***	0.597***
Male	-0.036***	-0.036***	0.006	0.006
Foreign Language Background	0.029***	0.029***	0.114***	0.115***
Std. SES	0.096***	0.096***	0.077***	0.077***
Local Rank (t)	0.578***	decile^	-0.204***	decile^
Local Rank (t)^2	0.271***	N/A	0.801***	N/A
Inferred Change in Rank				
Decile 1 (bottom)	omitted	omitted	omitted	omitted
Decile 2	0.075***	0.066***	0.034***	0.031***
Decile 3	0.088***	0.076***	0.043***	0.038***
Decile 4	0.118***	0.103***	0.065***	0.060***
Decile 5	0.113***	0.094***	0.068***	0.063***
Decile 6	0.113***	0.090***	0.047***	0.039***
Decile 7	0.127***	0.101***	0.047***	0.039**
Decile 8	0.130***	0.101***	0.034**	0.026*
Decile 9	0.152***	0.118***	0.029*	0.015
Decile 10 (Top)	0.176***	0.131***	0.014	-0.008
Constant				
N	93,997	93,997	93,425	93,425

Table 5.40 (A): Decile Indicators of Inferred Change in Rank, Numeracy^

	(1)	(2)	(3)	(4)
Std. Numeracy	0.683***	0.665***	0.765***	0.681***
Std. Numeracy ^2	0.054***	0.037***	0.044***	0.039***
Std. Numeracy ^3	-0.018***	-0.015***	-0.019***	-0.016***
Male	0.006	0.005	0.006	0.005
Foreign Language Background	0.112***	0.112***	0.112***	0.112***
Std. SES	0.076***	0.076***	0.076***	0.076***
Local Rank (t)	0.402***	0.020	quartile^	decile^
Local Rank (t)^2		0.421***	N/A	N/A
Inferred Change in Rank				
Decile 1 (bottom)	omitted	omitted	omitted	omitted
Decile 2	0.015	0.016	-0.007	0.010
Decile 3	0.017	0.016	-0.015	0.007
Decile 4	0.032**	0.028*	-0.007	0.020
Decile 5	0.022	0.020	-0.018	0.013
Decile 6	0.008	0.005	-0.042***	-0.004
Decile 7	0.007	0.007	-0.048***	-0.001
Decile 8	-0.013	-0.008	-0.072***	-0.016
Decile 9	-0.029*	-0.018	-0.094***	-0.031*
Decile 10 (Top)	-0.058***	-0.043**	-0.142***	-0.065***
Constant	-0.336***	-0.278***	-0.119**	-0.266***
N	93,425	93,425	93,425	93,425

^Local rank variables: linear, quadratic, quartiles indicators, deciles indicators

Table 5.41 (A): Decile Indicators of Inferred Change in Rank, Reading^

	(1)	(2)	(3)	(4)
Std. Reading	0.649***	0.644***	0.727***	0.673***
Std. Reading^2	0.011***	0.004	0.006*	0.004
Std. Reading^3	-0.020***	-0.019***	-0.023***	-0.020***
Male	-0.036***	-0.036***	-0.037***	-0.036***
Foreign Language Background	0.030***	0.030***	0.030***	0.030***
Std. SES	0.094***	0.095***	0.094***	0.095***
Local Rank (t)	0.376***	0.185***	quartile^	decile^
Local Rank (t)^2		0.205***	N/A	N/A
Inferred Change in Rank				
Decile 1 (bottom)	omitted	omitted	omitted	omitted
Decile 2	0.036***	0.036***	0.017	0.030**
Decile 3	0.036**	0.034**	0.010	0.026*
Decile 4	0.056***	0.052***	0.023*	0.043***
Decile 5	0.044***	0.042**	0.008	0.030*
Decile 6	0.037**	0.033**	-0.005	0.020
Decile 7	0.040**	0.038**	-0.006	0.023
Decile 8	0.032*	0.031*	-0.019	0.014
Decile 9	0.034*	0.037*	-0.022	0.016
Decile 10 (Top)	0.019	0.024	-0.052***	-0.004
Constant	-0.271***	-0.234***	-0.120***	-0.191***
N	93,997	93,997	93,997	93,997

^Local rank variables: linear, quadratic, quartiles indicators, deciles indicators

A.6. Combinations of Local Rank and Changes in Local Rank

Table 5.42 (A): Combinations of Inferred Change in Rank and Local Rank, Numeracy^

	Local Rank Quartiles (t)			
Numeracy (t+1)	Q1	Q2	Q3	Q4
Inferred Change in Rank Q1 (t, t+1*)	omitted	0.029*	0.098***	0.207***
Inferred Change in Rank Q2 (t, t+1*)	-0.006	0.040**	0.089***	0.208***
Inferred Change in Rank Q3 (t, t+1*)	-0.012	0.017	0.069***	0.123***
Inferred Change in Rank Q4 (t, t+1*)	-0.046***	-0.027	0.009	0.029
N=93,425				

^ cohort and secondary school indicators not shown, primary school random effects

Table 5.43 (A): Combinations of Inferred Change in Rank by Quartile of Local Rank, Reading^

	Local Rank Quartiles (t)			
Reading (t+1)	Q1	Q2	Q3	Q4
Inferred Change in Rank Q1 (t, t+1*)	omitted	0.014	0.069***	0.136***
Inferred Change in Rank Q2 (t, t+1*)	0.001	0.023	0.060**	0.128***
Inferred Change in Rank Q3 (t, t+1*)	-0.018	0.007	0.050*	0.105***
Inferred Change in Rank Q4 (t, t+1*)	-0.040**	-0.018	0.023	0.047
N=93,997				

^ cohort and secondary school indicators not shown, primary school random effects

A.7. Peer Characteristics and Changes between Year 5 and Year 7

Table 5.44 (A): Peer Characteristics by Quartile of Inferred Change in Rank, Numeracy

	Q1	Q2	Q3	Q4
Number of Students	56.7 (25.0)	64.9 (27.5)	66.6 (27.8)	63.5 (28.5)
Male	51.1 (7.9)	51.4 (7.3)	51.6 (7.2)	52.1 (7.4)
Foreign Language	28.6 (24.9)	25.8 (23.7)	25.1 (23.4)	27.4 (24.4)
Std. SES	-0.13 (0.55)	-0.03 (0.52)	0.00 (0.51)	0.05 (0.54)
ATSI	2.0 (3.9)	1.6 (3.2)	1.5 (2.9)	1.2 (2.4)
Std. Numeracy	-0.26 (0.38)	-0.08 (0.37)	0.03 (0.37)	0.24 (0.39)
Std. Reading	-0.19 (0.37)	-0.06 (0.34)	0.02 (0.33)	0.15 (0.35)
Missing Numeracy	7.2 (5.6)	6.6 (5.0)	6.4 (4.9)	6.7 (5.5)
Missing Reading	7.0 (5.5)	6.4 (5.0)	6.1 (4.8)	6.5 (5.5)
Both Missing	5.9 (5.1)	5.3 (4.5)	5.2 (4.4)	5.5 (5.1)

Table 5.45 (A): Changes in Peer Characteristics by Quartile of Inferred Change in Rank, Numeracy

	Q1	Q2	Q3	Q4
Number of Students	140.7 (71.9)	125.1 (74.8)	113.8 (72.2)	102.7 (71.3)
Male	1.3 (12.6)	1.2 (11.5)	0.6 (12.1)	-0.2 (14.1)
Foreign Language	-0.9 (16.4)	-1.1 (13.1)	-1.4 (12.4)	-1.5 (14.7)
Std. SES*	0.25 (0.35)	0.09 (0.29)	0.01 (0.28)	-0.10 (0.34)
ATSI	-0.4 (3.1)	0.2 (2.4)	0.5 (2.4)	0.7 (2.6)
Std. Numeracy*	0.39 (0.28)	0.13 (0.23)	-0.05 (0.23)	-0.33 (0.28)
Std. Reading*	0.30 (0.27)	0.10 (0.24)	-0.03 (0.24)	-0.22 (0.28)
Missing Numeracy*	-0.5 (6.2)	0.6 (5.8)	1.2 (5.9)	1.3 (6.5)
Missing Reading*	-0.4 (6.4)	0.6 (6.0)	1.3 (6.1)	1.5 (6.7)
Both Missing*	-1.3 (5.5)	-0.5 (5.0)	0.0 (5.1)	0.1 (5.8)

* Change in actual values

Chapter 6. Conclusion

6.1. Summary & Contribution

Selective high schools make up a small part of Australian education systems, with their policies varying at the state and territory level. The schools are important for the achievement ideal that is reflected in their students' educational and career successes, while some selective schools are known for their successful alumni, who include among them prominent public figures (e.g. former Prime Ministers and Premiers of State[141]). In contrast, selective schools also attract negative attention for their perceived elitism and for promoting inequality through the grouping of high-achievement students, which has been argued to disadvantage the remaining students who attend their former schools[142].

This thesis asked whether selective schools improved their students' university entrance results beyond what they would have otherwise achieved and sought to answer this question by following two recent cohorts of high-achievement students through high school from an anonymized Australian state. The thesis contributed to the literature by showing that the selective school effect in the Australian context was smaller than might be expected. This is significant because university entrance results are a high stakes education outcome that largely determines the receipt of offers for university courses, and for which newspapers report the annual rankings of school performance.

The main analyses are a comparison of two methods and provided the first estimates from matching and regression discontinuity approaches (roughly contemporaneous with Zen 2016[143]) in the Australian context. These improved statistical methods produced more precise estimates of the selective school effect compared with that of previous research (e.g. value-added regression analyses from Lu and Rickard, 2014).

The results indicated that there were only small positive effects at best from attending the selective schools from the anonymized Australian state. In addition, exploratory analyses and previous research supported the interpretation that selective students had high levels of motivation or educational aspiration, meaning that they are likely to have done equally as well

[141] E.g. Former Prime Ministers Bob Hawke (Perth Modern School) and Edmund Barton (Fort Street High School), Premier of New South Wales Neville Wran (Fort Street High School).

[142] See for example, 'Selective schools are failing our children. We don't need to build more', Opinion, Anne Susskind, abc.net.au, 26 July 2017.

[143] Zen (2016) conducted similar regression discontinuity analyses for New South Wales concurrent to the research in Chapter 4. Zen finds limited and mostly insignificant effects consistent with the research here and previous studies from the UK and the USA. Note that Zen's study was not accessible prior to March 2018.

by attending other schools. Note that the preliminary analyses from the matching approach in Chapter 3 indicated that there were small positive effects on university entrance results for girls from selective school attendance, but the comparisons did not account for the probable higher levels of aspiration and motivation of students attending selective schools.

Expanding on the idea that the group frame-of-reference can influence individuals' academic self-concept, and hence subsequent achievement, the thesis also asked what the effect of changes in local achievement ranks was for students' later academic achievement. The study found evidence of immediate short-term impacts on achievement from increases in academic status, which were arguably unexpected and attributable to random variation. I found that large perceived increases in rank tended to negatively affect students' later achievement; estimates were up to -0.083 SD for Numeracy, and up to -0.046 SD for Reading, in effect size from experiencing top quartile inferred changes in rank, relative to the bottom quartile.

These results provided evidence of self-referencing behaviour, whereby individuals make comparisons with their past performances, over short time horizons. The diminished performance from an increase in local ranks suggested that individuals allocate effort for efficiency reasons, rather than improving their performances from increases in academic self-concept as would be predicted by a positive group frame-of-reference effect.

The literature review chapter placed the Australian selective school policies in their historical and conceptual contexts, documenting their original functions and intentions, and investigated the premise that academically selective schools cater to gifted students. I additionally explored whether the apparent demand for selective schools was related to trends driven by the economic narrative of education.

The polarizing nature of the selective schools appeared to partially stem from the visibly high representation of students of immigrant background, and apparent differences in their cultural attitudes. E.g. selective students and parents were thought to hold an instrumental view of education, including an emphasis on effort and tutoring, compared with that other parents who emphasized a connection to the local community. The geographic placement of the selective schools in certain local areas also potentially accentuated the differences in appearance and culture, creating a sharp contrast between their students and residents (see section 2.5.3).

As well as contributing to the relevant literature, the thesis provides several indirect contributions. From a policy perspective, the new research contributes to the public discourse on the selective schools, which appears to have been dominated by negative attention and immigrant culture in the media. E.g. hyperbolic commentary suggested that students of immigrant background were gaining entry to the schools unfairly through excessive tutoring.

From a research perspective, this thesis is the first to analyse the Australian selective schools using more sophisticated statistical methods (roughly contemporaneous with Zen, 2016), and the first to examine the issue in-depth; making connections to educational psychology, sociology, and the economics of education more generally. The changing landscape has meant that the selective high schools now appear to fulfill a different role from to the ones held earlier, and it is possible that the narratives associated with specific schools are also outdated. The polarized opinions toward the schools, demonstrated by the high levels of demand for the schools by students and parents contrasted with the negative critiques in the media, indicates that further public discourse is required to formulate and define an evolved or adapted function of the selective schools to ensure their continuing relevance.

Overall, the empirical analyses cast some doubt on the purpose of the schools from the lack of significant positive effects on achievement, particularly given their academic focus. A strong case be made, however, that the narrowly defined measure of assessment, the improvement of university entrance results for individuals, is not the most useful way of looking at the selective schools, especially because of the high levels of implied initial academic ability of their students. The results from the thesis suggested that the advantages of the selective schools' policies were not in improving the medium-term academic performances of their students. However, analysing the educational performance of students attending selective schools is a necessary and important step towards identifying and articulating the advantages of the selective schools (discussed next).

6.2. Policy Challenge

The thesis identified that the key policy challenge for selective schools was articulating their function in the current policy landscape. As a final thought, this thesis attempts to do this by considering a broader range of potential contributions by selective schools. Whether there should be selective schools in Australian education systems is a different question from the one tackled by this thesis of whether the schools improve their students' high school achievement outcomes, where the former is a broader question of which the latter is part.

The thesis showed that there was little positive effect from attending selective schools on university entrance results relative to the schools they would have attended otherwise, and it pointed out a potentially negative indirect consequence from attending these schools. I.e. the possible negative group frame-of-reference effect from having high-achievement peers, which is analogous to being a small fish in a big pond (the reverse description of Marsh's big-fish-little-pond effect).

This observation, along with the local effects produced by comparing marginal selective students and marginal non-selective students (i.e. those who just got in and those who just missed out, respectively) points to the existence of heterogeneous effects. This view is supported by Jonkmann et al. (2012), who showed that the negative effects on academic self-concept from having higher achievement peers were, respectively, smaller for secondary students with higher measures of narcissism, and larger for students with higher levels of neuroticism.

The strongest argument for attending a selective school is provided by Ahmavaara and Houston (2007) who showed that students who attended selective schools tended to have a growth mindset towards achievement, as opposed to a fixed mindset, a distinction developed by Dweck and colleagues (2000) (see Achievement Motivation, section 2.3.5). They suggested that the selective schools could provide confirmation or reinforcement of student's self-identities, and could further increase individuals' aspirations by leading them to aim at higher goals.

That the selective schools form part of an individual's narrative of success bears a strong resemblance to the idea that students can draw confidence from positive narratives associated with the groups they are members of, as discussed by Chua and Rubenfeld (2014) and Lee and Zhou (2015) who analyse immigrant success and Asian American achievement[144]. Due to the large representation of students of immigrant background at selective schools, the potentially close link between individual and group narrative suggests then that one of the current functions of the selective schools is that of helping to fulfill the successful (second generation) immigrant story.

More generally, the relevance of the selective schools can be motivated by the power of narrative in that the schools provide a goal for prospective students to aim at, and help their students strive to emulate or surpass the accomplishments of their alumni. This in turn helps to maintain and enhance the reputations of these schools. The individual schools themselves appear to place value in different ideas which can be hinted at by the traditions that they hold,

[144] See sections 2.5.2 and 2.5.3 from Chapter 2. Both of their writings are influenced by Bourdieu's concept of symbolic capital, defined as intangible resources based on prestige, honour, or social recognition, as described by Lee and Zhou (2015) citing Bourdieu (1984 and 1987) and Wacquant (2013).

and through reviewing their histories and their celebrated alumni (see Chapter 2: A Reputational Explanation, section 2.2.3.1).

In this light, the selective schools may have a positive effect on their students' achievement by generally positively influencing and motivating their applicants, consistent with the possibility that differences in achievement between successful and unsuccessful applicants is not distinguishable. This interpretation is supported by the results from the thesis, which suggested that both groups of students benefited from the existence of the schools by encouraging their aspirations, particularly when students' abilities are latent in that they have yet to be transformed into, grades or performance.

At a more abstract level, the positive achievement and aspirational ideals that the selective schools represent means that questions of their effectiveness on short-term educational performance, and of the optimal number of selective schools, are subordinate to the question of their existence[145]. Further, the positive ideal offered by the existence of the schools is consistent with the values of an achievement-oriented society like Australia's and is also supported by research suggesting that economic growth is positively associated with the achievement motivation of individuals (McClelland 1961).

This is not to say that there are no negative effects produced by the selective school policies. The possible negative effects can be readily identified from the symbolic capital concept in that students who do not fit the positive narratives, roughly approximated as group or cultural stereotypes, are likely to be disproportionately affected by the policies' influences. These could be that they are open to criticism or potentially subject to excessive pressure from unrealistically high expectations from others[146]. It is also possible that students are more susceptible to elitist messages at the schools with counterproductive effects; a subtle critique of elitism can be found in Girard's (1961) analyses of literature, from which he observed that individuals' desires appear to be derived from others' desires.

For completeness, there are several advantages unrelated to improvements in achievement that the selective schools can contribute. First, regardless of value-added, the schools may be considered meaningful outcomes in themselves; as markers of success, although this view does treat the schools as a display of consumption. Second, it has been suggested that the

[145] Note, though, there is no certainty that the removal of an institution through which a phenomenon is channeled, diminishes the phenomenon. In this case the phenomenon being the achievement of aspirational students and the institution being the selective schools.

[146] This theme is well explored in Lee and Zhou (2015). A recent typical example is: "'Severe moral crisis': what I know about selective school students.", anonymous, The Age, 18 January 2018, accessed 18 January 2018.

schools offer an alternative to private schools, increasing equality of opportunity. This is an idea which can be traced to the schools' historical origins[147], though this appears to be an outdated assertion with selective students now typically having very advantaged socioeconomic backgrounds.

There may also be social and professional benefits from attending the selective schools, including the opportunity to join a network of like-minded alumni. Longer term, the asymmetric distribution of creative and scientific output, whereby most of the output is generated by only a disproportionately small number of individuals, suggests that even if benefits from the selective schools accrued to only a small number of individuals, there would be a net positive benefit from a societal perspective[148].

Evidence of little general value-added increases in educational performance from the selective schools is a challenging finding for the economic discipline, which has tended to adopt a utilitarian function of education towards maximizing economic output. The results of the thesis suggest that the importance of these schools may not be for the tangible impacts that they have on short-term achievement outcomes, but instead may be for the narrative impact on individuals' motivations. By creating school and peer environments with an achievement orientation, the selective schools provide a goal to aim at, which also encourages aspirational individuals towards future successes.

[147] Several of the selective schools commenced as the first government alternatives to private secondary schools in the late 1800s.

[148] The asymmetric distribution in creative output is discussed in Chapter 2: Expert Performance and Giftedness (section 2.4.3). Further reading on how only a small number of individuals generate most of the output include: Merton (1988) and Jensen (1995). From Price's law, the square root of the number of individuals generate half the output.

References

Abbas, T. (2007). British South Asians and Pathways Into Selective Schooling: Social Class, Culture and Ethnicity. British Educational Research Journal, 33(1), 75-90.

Abdulkadiroglu, A., Angrist, J., and Parag, P. (2014). The Elite Illusion: Achievement Effects at Boston and New York Exam Schools. Econometrica, 82(1), p137-196.

Ahmavaara, A. and Houston, D. (2007). The effects of selective schooling and self-concept on adolescents' academic aspiration: An examination of Dweck's self-theory. British Journal of Educational Psychology, 77, 613-632.

Amabile, T. (1983). The Social Psychology of Creativity: A Componential Conceptualization. Journal of Personality and Social Psychology, 45(2), 357-376.

Anderson, C., Hildreth, J., and Howland, L. (2015). Is the Desire for Status a Fundamental Human Motive? A Review of the Empirical Literature, 141(3), 574-601.

Angrist, J. and Pischke, J. (2009), Mostly Harmless Econometrics. Princeton University Press.

Angrist, J. (2014). The Perils of Peer Effects. Labour Economics (30), p98-108.

Angrist, J. and Rokkanen, M. (2015). Wanna Get Away? Regression Discontinuity Estimation of Exam School Effects Away from the Cutoff. Journal of American Statistical Association, 110(512), 1331-1344.

Arcidiacono, P., Khan, S., and Vigdor, J. (2011). Representation versus assimilation: How do preferences in college admissions affect social interactions? Journal of Public Economics, 95, 1-15.

Atkinson, J. W. (1966). Motivational determinants of risk-taking behaviour. In Atkinson, J. & Feathers, N. (Eds.), A theory of achievement motivation, 11-30. New York: Wiley.

Becker, K. and Park, K. (2011). Effects of integrative approaches among science, technology, engineering, and mathematics (STEM) subjects on students' learning: A preliminary meta-analysis. Journal of STEM education, 12 (5&6).

Becker, M., Neumann, M., Teztner, J., Bose, S., Knoppick, H., Maaz, K., Baumert, J., and Lehmann, R. (2014). Is Early Ability Grouping Good for High-Achieving Students' Psychological Development? Effects of the Transition into Academically Selective Schools. Journal of Educational Psychology, 106(2), 555-568.

Becker, M. and Neumann, M. (2016). Context-related changes in academic self concept development: On the long-term persistence of big-fish-little-pond effects. Learning and Instruction, 45, 31-39.

Becker, S. and Ichino, A. (2002), Estimation of Average Treatment Effects Based on Propensity Scores, The Stata Journal, 2(4), p358-377.

Beilock, S., Gunderson, E., Ramirez, G., and Levine, S. (2010). Female teachers' math anxiety affects girls math achievement. PNAS, 107(5), 1860-1863.

Blundell, R., Dearden, L., and Sibieta, L. (2010). The demand for private schooling in England: the impact of price and quality. Institute for Fiscal Studies, Working Paper 10/21.

Bonhomme, S., and Sauder, U. (2011). Recovering Distributions in Difference-in-difference Models: A Comparison of Selective and Comprehensive Schooling. The Review of Economics and Statistics, 93(2), 479-494.

Borjas, G. (1992). Ethnic Capital and Intergenerational Mobility. The Quarterly Journal of Economics, 107(1), 123-150.

Bradley, S., Johnes. G., and Millington, J., (2000). The effect of competition on the efficiency of secondary schools in England. European Journal of Operational Research 135, 545-568.

Braggett, E. (1985). Education of Gifted and Talented Children: Australian Provision. Canberra Publishing & Printing Co., Fyshwick, A.C.T.

Braggett, E. and Moltzen, R., (2000). Programs and Practices for Identifying and Nurturing Giftedness and Talent in Australia and New Zealand. In Heller, K, Monks, Sternberg, J., and F., Subotnik, R. (eds.), International Handbook of Giftedness and Talent (2^{nd} edition), Pergamon.

Braithwaite, R. and Kensell, H. (1992). Factors Influencing enrolment at NSW selective high schools. The Australian Educational Researcher, 19(2), 39-49.

Brounstein, Holahan and Dreyden, (1991). Change in Self-Concept and Attributional Styles Among Academically Gifted Adolescents. Journal of Applied Social Psychology, 21(3), 198-218.

Burhans, K. and Dweck, C. (1995). Helplessness in Early Childhood: The Role of Contingent Worth. Child Development, 66 (6), 1719-1738.

Burgess, S., Dickson, M., and Macmillan, L. (2014). Selective Schooling Systems Increase Inequality. IZA Discussion Paper Series, No. 8505.

Butler, R., Ho, C., and Vincent, E. (2017). 'Tutored within an inch of their life': morality and 'old' and 'new' middle class identities in the Australian schools. Journal of Ethnic and Migration Studies, 43 (14), 2408-2422.

Butts, R. (1955). Assumptions underlying Australian Education. Melbourne: Australian Council for Educational Research.

Card, D. and Guilano, L. (2014). Does Gifted Education Work? For Which Students? NBER Working Paper Series, Working Paper 20453.

Carson, S., Peterson, J., and Higgins, D. (2005). Reliability, Validity, and Factor Structure of the Creative Achievement Questionnaire. Creativity Research Journal, 17(1), 37-50.

Chapman, P. (1988). Schools as Sorters: Lewis M. Terman, Applied Psychology, and the Intelligence Testing Movement, 1890-1930. New York University.

Chan, L (1996). Motivational Orientations and Metacognitive Abilities of Intellectually Gifted Students. Gifted Child Quarterly, 40(4), 184-193.

Chiu, M. (2012). The Internal/External Frame of Reference Model, Big-Fish-Little-Pond Effect, and Combined Model for Mathematics and Science. Journal of Educational Psychology, 104(1), 87-107.

Chua, A., and Rubenfeld, J. (2014). The Triple Package: How Three Unlikely Traits Explain the Rise and Fall of Cultural Groups in America. New York: Penguin Press.

Cicala, S., Fryer, Jr R., and Spenkuch, J. (2017). Self-Selection and Comparative Advantage in Social Interactions. Journal of the European Economic Association.

Clark, D. (2010). Selective Schools and Academic Achievement. The B.E. Journal of Economic Analysis & Policy, 10 (1).

Clark, D. and Del Bono (2016). The Long-Run Effects of Attending an Elite School: Evidence from the United Kingdom. American Economic Journal: Applied Economics, 8(1): 150-176.

Cleveland, W. (1979). Robust Locally Weighted Regression and Smoothing Scatterplots. Journal of American Statistical Association, 74 (368), 829-836.

Clynes, T. (2016, September 7). How to raise a genius: lessons from a 45-year study of super-smart children. Nature. Retrieved from https://www.nature.com/news/how-to-raise-a-genius-lessons-from-a-45-year-study-of-super-smart-children-1.20537.

Coe, R., Jones, K., Searle, J., Kokotsaki, D., Mohd Kosnin, A., Skinner, P. (2008). Evidence on the effects of selective educational systems: A report for the Sutton Trust, Centre for Evaluation and Monitoring. Durham University.

Colangelo, N., Assouline, S., and Gross, M. (Eds). (2004). A Nation Deceived: How Schools Hold Back America's Brightest Students (Volume 2). The University of Iowa.

Colangelo, N., and Davis, G. (Eds.). (2003). Handbook of Gifted Education (3rd edition). Allyn and Bacon.

Cole, S. and Barber, E., with Bolyard, M., and Linders, A. (2003). Increasing Faculty Diversity. Cambridge, Massachusetts; London, England: Harvard University Press.

Cullen, J., Jacob, B., and Levitt, S., (2005). The impact of school choice on student outcomes: an analysis of the Chicago Public Schools. Journal of Public Economics (2005), 89, 729-760.

Cunha, F. and Heckman, J. (2008). Formulating, Identifying, and Estimating the Technology of Cognitive and Noncognitive Skill Formation. The Journal of Human Resources, 43 (4), 738-782.

Dandy, J. and Nettelbeck, T. (2002). The Relationship between IQ, Homework, Aspirations and Academic Achievement for Chinese, Vietnamese and Anglo-Celtic Australian School Children. Educational Psychology, 22(3), p267-275.

Davidoff, I. and Leigh, A. (2008). How Much do Public Schools Really Cost? Estimating the Relationship between House Prices and School Quality. The Economic Record 84(264) 193-206.

Davis, G. (2003). Identifying Creative Students, Teaching for Creative Growth. Chapter 24 in Colangelo, N., and Davis, G., Handbook of Gifted Education (3rd edition), Pearson Education.

Davis, J. (1966). The Campus as a Frog Pond: An Application of the Theory of Relative Deprivation to Career Decisions of College Men. American Journal of Sociology, 72(1), 17-31.

De Botton, A., (2004). Status Anxiety. New York: Vintage International.

Deci, E. and Moller, A. (2007). The Concept of Competence. In Elliot, A. and Dweck, C. (Eds.), Handbook of Competence and Motivation, Guildford Publications.

Di Tella, R., New, J.H-D., MacCulloch, R. (2010). Happiness Adaptation to Income and to Status in an Individual Panel. Journal of Economic Behaviour and Organization.

Dobbie. W. and Fryer, Jr R. (2014). 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), p58-75.

Dobson, I. and Skuja, E. (2005). Secondary Schooling, Tertiary Entry Ranks and University Performance. People and Place, 13(1), p53-62.

Duckworth, A (2016). Grit, The Power of Passion and Perseverance. Simon and Schuster.

Dunbar, R. (1992). Neocortex Size as a Constraint on Group Size in Primates. Journal of Human Evolution, 22(6), 469-493.

Duflo E., Dupas P., and Kremer, M., (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya. American Economic Review, 101(5), 1739-1774.

Dustmann, C., Puhani, P., and Schonberg, U. (2012). The Long-term Effects of School Quality on Labor Market Outcomes and Educational Attainment. Center for Research and Analysis and Migration, Discussion Paper Series, 08/12.

Dweck, C. (2000). Self-Theories: Their Role in Motivation, Personality, and Development. Psychology Press. Taylor & Francis Group.

Education and Training Committee, Parliament of Victoria (2012), Inquiry into the Education of Gifted and Talented Students, Parliamentary Paper of Victorian Government, No. 108 Session 2010-2012.

Edwards, D (2005). Change, competition and specialisation: the demise of the comprehensive secondary school and its implications. TASA 2005 Conference Proceedings, University of Tasmania, 6-8 December 2005.

Elliot, R., Strenta, C., Adair, R., Maiter, M., and Scott, J. (1996). The Role of Ethnicity in Choosing and Leaving Science in Highly Selective Institutions. Research in Higher Education, 37(6), 681-709.

Elsner, B., and Isphording, I. (2015). Rank, Sex, Drugs and Crime. IZA Discussion Papers, No. 9478.

Elsner, B. and Isphording, I. (2017). A big fish in a small pond: Ability rank and human capital investment. Journal of Labor Economics, 35(3), 787-828.

Ericsson, A., Nandagopal, K., and Roring, R. (2009). An Expert Performance Approach to the Study of Giftedness. In Shavanina, L. (Ed.), International Handbook of Giftedness, Springer Netherlands.

Estrada and Gignoux (2017). Benefits to Elite Schools and the Expected Returns to Education: Evidence from Mexico City. Development Bank of Latin America, Working Paper 2017/02.

Eysenck, H. J., (1992). The Definition and Measurement of Psychoticism. Personality and Individual Differences, 13(7). 757-785.

Feld, J. and Zolitz, U (2017). Understanding Peer Effects: On the Nature, Estimation and Channels of Peer Effects. Journal of Labour Economics 35(2), 387-428.

Feldhusen, J. (1985). Toward Excellence in Gifted Education. In Feldhusen, J. (Ed.), Love Publishing Company.

Feldman, D. (1982). Developmental approaches to giftedness and creativity. San Francisco: Jossey-Bass.

Flynn, J. (1999). Searching for Justice: The discovery of IQ gains over time. American Psychologist 54(1), 5-20.

Frank, R (1985). Choosing the right pond: human behaviour and the quest for status. New York; Oxford: Oxford University Press.

Frank, R (2007). Falling Behind: How Rising Inequality Harms the Middle Class (2nd Edition). Berkeley: University of California Press.

Fogelman, K. (1984). Problems in Comparing Examination Attainment in Selective and Comprehensive Schools. Oxford Review of Education, 10(1), 33-43.

Fox, L. (1981). Identification of the Academically Gifted. American Psychologist, 36(10), 1103-1111.

Fox, J., and Weisberg, S. (2010). Nonparametric Regression in R: An Appendix to an R Companion to Applied Regression, 2nd edition. Revised December 2010.

Gagne, F. (1985). Giftedness and Talent: Reexamining a Reexamination of the Definitions. Gifted Child Quarterly, 29(3), 103-112.

Gagne, F. (2009). Debating Giftedness: Pronat vs. Antinat. In Shavanina, L (Ed.). International Handbook of Giftedness, Springer Netherlands.

Gee, C (2012). The educational 'arms race': all for one, loss for all, September 27, 2012. https://www.ipscommons.sg/the-educational-arms-race-all-for-one-loss-for-all-part-1/, accessed 11 November 2017.

Gemici, S., Bednarz, A., Karmel, T., and Lim, P. (2014). The factors affecting the educational and occupational aspirations of young Australians. Longitudinal Surveys of Australian Youth, Research Report 66.

Gibbons, S. and Telhaj, S. (2008). Peers and Achievement in England's Secondary Schools. Spatial Economics Research Centre Discussion Paper 1.

Gill, D., Kissova, Z., Lee, J., and Prowse, V. (2015), First-Place Loving and Last-Place Loathing: How Rank in the Distribution of Performance Affects Effort Provision. IZA Discussion Papers, No. 9286.

Girard, R. (1965). Deceit, desire, and the novel: self and other in literary structure / Translated by Yvonne Freccero. Baltimore: John Hopkins Press.

Gladwell, M. (2013). David and Goliath: Underdogs, Misfits, and the Art of Battling Giants, New York: Little Brown and Company.

Gottfredson, L. (1981). Circumscription and Compromise: A Developmental Theory of Occupational Aspirations. Journal of Counseling Psychology Monograph, 28(6), 545-579.

Gradstein, M. and Justman, M. (2000). Human capital, social capital, and public schooling, European Economic Review, 44(4-6), 879-890.

Gross, M (2005). Exceptionally Gifted Children (2nd edition). Taylor and Francis.

Gruber, H. E., (1982). On the hypothesized relation between giftedness and creativity. In D. H. Feldman (Ed.), Developmental approaches to giftedness and creativity, San Francisco: Jossey-Bass.

Guyon, N., Machin, E., and McNally, S. (2012). The Effect of Tracking Students by Ability into Different Schools. The Journal of Human Resources, 47(3), 684-721.

Hastings, J., and Weinstein, J. (2008). Information, School Choice, and Academic Achievement: Evidence from Two Experiments. The Quarterly Journal of Economics, 123(4), 1373-1414.

Handel, M., Vialle, W., and Ziegler, A. (2013). Student perceptions of high-achieving classmates. High Ability Studies, 24(2), 99-114.

Hanushek and Woessman (2005). Does Educational Tracking Affect Performance and Inequality? Differences-in-Differences Evidence Across Countries. NBER Working Paper Series, Working Paper 11124.

Harrod, R. (1958). The Possibility of Economic Satiety, in Problems of United States Economic Development, Vol. I. Committee for Economic Development, New York.

Heckman, J. and Rubinstein, Y. (2001). The Importance of Noncognitive Skills: Lessons from the GED Testing Program. AEA Papers and Proceedings, 91(2), 145-149.

Heaven, P., Ciarrochi, J., and Vialle, W. (2007). Conscientiousness and Eysenckian psychoticism as predictors of school grades: A one-year longitudinal study. Personality and Individual Differences, 42, 535-546.

Hirsch (1976). Social limits to growth. Cambridge, MA: Harvard University Press.

Ho, C. (2011). 'My School' and others: Segregation and white flight. Australian Review of Public Affairs.

Ho, C. (2015). 'People like us': School choice, multiculturalism and segregation in Sydney. Australian Review of Public Affairs. https://www.australianreview.net/digest/2015/08/ho.html, accessed 19 February 2018.

Ho, C. (2017). Angry Anglos and the aspirational Asians: everyday multiculturalism in the selective school system in Sydney. Discourse: Studies in the Cultural Politics of Education, 1-16.

Hoekstra, M., Mouganie, P., and Wang, Y. (2016). Peer Quality and Academic Benefits to Attending Better Schools. NBER Working Paper No. w22337.

Hollingworth, L (1942). Children Above 180 IQ. New York.

Homel, J., Mavisakalyan, A., Nguyen, HT., Ryan, C. (2012). School completion: what we learn from different measures of family background. Longitudinal Surveys of Australian Youth, Research Report 59.

Hong, S. and Lee, J. (2017). Who is sitting next to you? Peer effects inside the classroom. Quantitative Economics, 8, 239-275.

Houng, B. and Justman, M. (2013). Comparing Least-Squares Value-Added Analysis and Student Growth Percentile Analysis for Evaluating Student progress and Estimating School Effects. Melbourne Institute Working Paper 07/13, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

Houng, B. and Justman, M. (2014). NAPLAN Scores as Predictors of Access to Higher Education. Melbourne Institute Working Paper 22/14, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

Hsin, A. and Xie, Y. (2014). Explaining Asian Americans' academic advantage of whites. PNAS, 111 (2), 8415-8421.

Hunt, I. and Merrotsy, P. (2010). Adding value and value adding: the concept of 'value added' in partially and fully selective schools. Australasian Journal of Gifted Education, 19, 37-48.

Husen, T. (1960). Loss of Talent in Selective School Systems: The Case of Sweden. Comparative Education Review, 4(2), 70-74.

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

Jacob, R. Zhu, P, Somers, M-A, and Bloom, H.. (2012). A Practical Guide to Regression Discontinuity. MDRC (www.mdrc.org) working paper.

Jauk, E., Benedek, M., Dunst, B., and Neubauer, A. (2013). The relationship between intelligence and creativity: New support for the threshold hypothesis by means of empirical breakpoint detection. Intelligence, 41, 212-221.

Jensen, A. (1996). Giftedness and genius: Crucial differences. In C. P. Benbow & D. J. Lubinski (Eds.), Intellectual talent: Psychometric and social issues (pp.393-411). Baltimore, MD, US: Johns Hopkins University Press.

Jerrim, J., (2015). Why do East Asian children perform so well in PISA? An investigation of Western-born children of East Asian descent. Oxford Review of Education, 41(3), 310-333.

Jolly, J. (2009). A Resuscitation of Gifted Education. American Educational History Journal, 36 (1/2), 37-52.

Jonkmann. K., Becker, M., Marsh, H., Ludtke, O., and Trautwein, U. (2012). Personality traits moderate the Big-Fish-Little-Pond Effect of academic self-concept. Learning and Individual Differences. 22, 736-746.

Jonsson, J. and Mood, C. (2008), Choice by Contrast in Swedish Schools: How Peers' Achievement Affects Educational Choice. Social Forces 87(2), 741-765.

Kagan, J. (2013). The Human Spark. Basic Books, Member of the Perseus Books Group.

Kagan, J., Reznick, S., and Snidman, N. (1988). Biological bases of childhood shyness. Science, 167.

Kariya, T. (2011). Japanese Solutions to Equity and Efficiency Dilemma? Secondary schools, inequity, and the arrival of 'universal' higher education. Oxford Review of Education, 37(2), 241-266.

Karolyi, Ramos-Ford and Gardner (2003). Multiple Intelligences: A Perspective on Giftedness. Chapter 8 in Colangelo, N. and Davis, G (Eds.), Handbook of Gifted Education (3rd edition). Pearson Education.

Kim, K., (2005). Can Only Intelligent People Be Creative? A Meta-Analysis, The Journal of Secondary Gifted Education, 16(2/3), 57-66.

Kolenikov, S., and Angeles, G. (2004). The Use of Discrete Data in Principal Component Analysis with Application to Socio-Economic Indices. CPC/MEASURE Working paper. No. WP-04-85.

Lamb, S (2007). School Reform and Inequality in Urban Australia - A Case of Residualising the Poor, from Teese, R., Lamb, S., and Duru-Bellat, M. (Eds.). International Studies in Educational Inequality, Theory and Policy Volume 3: Inequality: Educational Theory and Public Policy, 1-38.

Lamb, S (2008). Selective-entry schools: Time for a re-think. Professional Voice, 5(3), 17-22.

Lamb, S., Jackson, J., Walstab, A., and Huo, S. (2015). Educational Opportunity in Australia 2015: Who succeeds and who misses out. Centre for International Research on Education Systems, Victoria University, for the Mitchell Institute, Melbourne: Mitchell Institute.

Lavy, V., Paserman, D., and Schlosser, A. (2007). Inside the black box of ability peer effects: Evidence from variation in high and low achievers in the classroom. Unpublished manuscript.

Lavy, V., Paserman, D., and Schlosser, A. (2012), Inside the Black Box of Ability Peer Effects: Evidence from Variation in the Proportion of Low Achievers in the Classroom, The Economic Journal, 122 (March), 208-237.

Lavy, V., Silva, O., and Weinhardt, F. (2012). The Good, the Bad, and the Average: Evidence on Ability Peer Effects in Schools. Journal of Labour Economics, 30(2), 367-414.

Lee, J., and Zhou, M., (2015). The Asian American Paradox, New York: Russell Sage Foundation.

Lee, J., and Zhou, M., (2017). Why class matters less for Asian-American Academic Achievement. Journal of Ethnic and Migration Studies. 43(14), 2316-2330.

Leslie, S., Cimpian, A., Meyer, M., and Freeland, E., (2015). Expectations of brilliances underlie gender distributions across academic disciplines. Science, 347(6219), 263.

Leung, W. and Denton-Schneider (2017). Peer Effects in Elite Swimmers. Leung, W. PhD Dissertation, University of Minnesota.

Long, J., Lynch, J., Machiran, N., Thomas, S., and Malinow, K. (1982). The Effect of Status on Blood Pressure During Verbal Communication. Journal of Behavioral Medicine, 5(2), 165-172.

Lovecky, D. (1994). Exceptionally gifted children: Different minds, Roeper review, 17(2), 116-120.

Lu, L., and Rickard, K. (2014). Value added models for NSW government schools. Centre for Education Statistics and Evaluation. Technical Paper / July 2014. Education and Communities, Office of Education, NSW government.

Lucas, R. (1988). On the Mechanics of Economic Development, Journal of Monetary Economics, 22(1), 3-42.

Lucas, A. and Mbiti, I. (2014). Effects of School Quality on Student Achievement: Discontinuity Evidence from Kenya. American Economic Journal: Applied Economics, 6(3), 234-263.

Lubinski, D. (2004). Chapter 3 in Colangelo, N., Assouline, S., and Gross, M (Eds.), A Nation Deceived: How Schools Hold Back America's Brightest Students, Volume 2, The University of Iowa.

Lubinski, D., Webb, R., Morelock, M., and Benbow, C. (2001). Top 1 in 10,000: A 10-year follow-up of the profoundly gifted. Journal of Applied Psychology, 86, 718-729.

Macleod, W. B., and Urquiola, M. (2011). Anti-Lemons: School Reputation, Relative Diversity, and Educational Quality. unpublished manuscript.

Macnamara, Hambrick and Oswald. (2014). Deliberate Practice and Performance in Music, Games, Sports, Education, and Professions: A Meta-Analysis. Psychological Science, 25(8), 1608-1618.

Manski, C. (1993). Identification of Endogenous Social Effects: The Reflection Problem. The Review of Economic Studies, 60(3), 531-542.

Marsh, H. and Parker, J. W. (1984). Determinants of Student Self-Concept: Is it better to be a relatively large fish in a small pond even if you don't learn to swim as well? Journal of Personality and Social Psychology, 74(1), p213-231.

Marsh, H. (1987). The Big-Fish-Little-Pond Effect on Academic Self-Concept. Journal of Educational Psychology, 79(3), 280-295.

Marsh, H., Byrne, B., and Shavelson, R. (1988). A Multifaceted Academic Self-Concept: Its Hierarchical Structure and Its Relation to Academic Achievement. Journal of Educational Psychology, 80(3), 366-380.

Marsh, H. (1989). Age and Sex in Multiple Dimensions of Self-Concept: Preadolescence to Early Adulthood. Journal of Psychology, 81(3), 417-430.

Marsh, H., Chessor, D., Craven, R., and Roche, L. (1995). The Effects of Gifted and Talented Programs on Academic Self-Concept: The Big Fish Strikes Again. American Educational Research Journal, 32(2), 285-319.

Marsh, H., Kong, C-K., and Hau, K-T., (2000). Longitudinal Multilevel Models of the Big Fish Little Pond Effect on Academic Self-Concept: Counterbalancing Contrast and Reflected Glory Effects in Hong Kong Schools. Journal of Personality and Social Psychology, 78(2), 337-349.

Marsh, H. and Hau, K. (2003). Big-Fish-Little-Pond Effect of Academic Self-Concept. American Psychologist, 58(5), 364-376.

Marsh, H., Trautwein, U., Ludtke, O., Koller, O., and Baumert, J. (2005). Academic Self-Concept, Interest, Grades, and Standardized Test Scores: Reciprocal Effects Models of Causal Ordering. Child Development, 76(2), 397-416.

Marsh, H., Trautwein, U., Ludtke, O., Baumert, J., and Koller, O, (2007). The Big-Fish-Little-Pond Effect: Persistent Negative Effects of Selective High Schools on Self-Concept after Graduation. American Educational Research Journal, 44(3), 631-669.

Marsh, H., Seaton, M., Trautwein, U., Ludtke, O., Hau, K., O'Mara, A., and Craven, R. (2008). The Big-fish-little-pond-effect Stands Up to Critical Scrutiny: Implications for Theory, Methodology, and Future Research. Educational Psychology Review, 20(3), 319-350.

Masden, C., Leung, O., Shore, B., Schneider, B., and Udvari, S. (2015). Social-perspective coordination and gifted adolescents' friendship quality. High Ability Studies. 26(1), 3-38.

Masters, G. (2016). Five Challenges in Australian School Education, Policy Insights, Issue 5. https://research.acer.edu.au/policyinsights/5/, accessed 19 February 2018.

Mavisakalyan, A (2012). Immigration and School Choice in Australia. The Australian Economic Review 45 (1), 29-49.

McClain and Pfeiffer (2012). Identification of Gifted Students in the United States Today: A Look at State Definitions, Policies, and Practices. Journal of Applied School Psychology, 28, 59-88.

McClelland, D. (1961). The achieving society. New York: Free Press; London: Collier-Macmillan.

McClelland, D., Koestner, R., and Weinberger, J. (1989). How do Self-Attributed and Implicit Motives Differ? Psychological Review, 96(4), 690-702.

McGuire, M., Raleigh, M., and Brammer, G. (1982). Sociopharmacology. Annual Review of Pharmacology and Toxicology, 22(1), 643-661.

McVicar, D., Moschion, J., and Ryan, C. (2013). Right Peer, Right Now? Endogenous Peer Effects and Achievement in Victorian Primary Schools. Melbourne Institute Working Paper 22/13, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

McVicar, D., Moschion, J., and Ryan, C. (2016). Achievement Effects from New Peers: Who Matters to Whom? Melbourne Institute Working Paper 17/16, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

Meadmore, P. (2001). 'Free, compulsory and secular'?, The re-invention of Australian public education. Journal of Education Policy, 16(2), 113-125.

Merrotsy, P. (2013). Tolerance of Ambiguity: A Trait of the Creative Personality? Creativity Research Journal, 25(2), 232-237.

Merton, R. (1988). The Matthew Effect in Science, II: Cumulative Advantage and Symbolism of Intellectual Property. Isis, 79 (4), 606-623.

Milgram and Milgram (1976). Personality Characteristics of Gifted Israeli Children. The Journal of Genetic Psychology, 129, 185-194.

Miller, P. (1999). Immigration Policy and Immigrant Quality: The Australian Points System. The American Economic Review, 89 (2), 192-197.

Mouganie, P., 2015. Three Essays in the Economics of Education. PhD Dissertation, Texas A&M University, College Station.

Mouganie, P. and Wang, Y. (2017). High Performing Peers and Female STEM Choices in School. MPRA Paper No. 81860.

Murphy, R. and Weinhardt, F. (2014). The Importance of Rank Position. CEP Discussion Paper No 1241.

Murphy, K., Schleifer, A., and Vishny, R. (1991). The Allocation of Talent: Implications for Growth. The Quarterly Journal of Economics, 106(2), 503-530.

Niepel, C., Brunner, M., and Preckel, F. (2014). The Longitudinal Interplay of Students' Academic Self-Concepts and Achievements Within and Across Domains. Journal of Educational Psychology, 106(4), 1170-1191.

Oakes, J. (1987). Tracking in Secondary Schools: A Contextual Perspective. Educational Psychologist, 22(2), 129-153.

OECD (2013), PISA 2012 Results: What Makes Schools Successful? Resources, Policies and Practices (Volume IV), PISA, OECD Publishing.

Park, A., Shi, X., Hsieh, C. and An, X. (2015), Magnet high schools and academic performance in China: A regression discontinuity design. Journal of Comparative Economics, 43, p824-843.

Piechowski, M. (1997). Emotional Giftedness: The Measure of Intrapersonal Intelligence. In Colangelo, N. and Davis, G. (eds.), Handbook of Gifted Education (Second edition), Allyn and Bacon.

Piechowski, M. and Colangelo, N. (1984). Developmental Potential of the Gifted. Gifted Child Quarterly, 28(2), 80-88.

Piechowski, Silverman, and Falk, (1985). Comparison of Intellectually and Artistically Gifted on Five Dimensions of Mental Functioning. Perceptual and Motor Skills, 60, 539-549.

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

Powdthavee, N. (2009). How important is rank to individual perception of economic standing? A within-community analysis. The Journal of Economic Inequality, 7(3), 225-248.

Renzulli, J. (1978), What Makes Giftedness? Reexamining a Definition. The Phi Delta Kappan, 60(3), 180-184, 261.

Rich, J. and De Vitis, J. (1992). Competition in Education. Springfield, Illinois, USA: Charles C Thomas.

Robertson, K., Smeets, S., Lubinski, D., and Benbow, C. (2010). Beyond the Threshold Hypothesis: Even among the Gifted and top Math/Science graduate students, cognitive abilities, vocational interests, and lifestyle preferences matter for career choice, performance and persistence. Current Directions in Psychological Science, 19(6), 346-351.

Robinson, P. (1988). Root-N-Consistent Semiparametric Regression. Econometrica, 56(4), 931-954.

Roe, A. (1953). The making of a scientist. New York: Dodd, Mead.

Rosen, S. (1981). The Economics of Superstars. The American Economic Review, 71(5), 845-858.

Royston, P. and Cox, N. (2005). A multivariate scatterplot smoother. The State Journal, 5(3), 405-412.

Runco, M. (1993). Divergent Thinking, Creativity, and Giftedness. Gifted Child Quarterly, 37(1), 16-22.

Ryan, C. (2017). Measurement of Peer Effects. Australian Economic Review, 50(1), pp 121-9.

Sander, R. and Taylor, S. (2012). Mismatch: How Affirmative Action Hurts Students It's Intended to Help, and Why Universities Won't Admit It. Basic Books.

Sacerdote, B. (2011). Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far? in Hanushek, E., Machin, S., Woessman, L. (Eds.), Handbook of the Economics of Education, Volume 3, Elsevier.

Salchegger, S. (2016). Selective School Systems and Academic Self-Concept: How Explicit and Implicit School-Level Tracking Relate to the Big-Fish-Little-Pond Effect Across Cultures. Journal of Educational Psychology, 108(3), 405-423.

Savin-Williams, R. (1979). Dominance Hierarchies in Groups of Early Adolescents. Child Development, 50(4), 923-935.

Schagen, I. and Schagen, S. (2002). A fair comparison: selective v. comprehensive education. Education Journal, 60, 26-27.

Schelling, T. (1969). Models of Segregation. Rand Corporation, Memorandum RM-6014-RC.

Schultheiss and Brunstein, (2007). An Implicit Motive Perspective on Competence. In Chapter 3 in Elliot, A. and Dweck, C. (Eds.), Handbook of Competence and Motivation, New York: Guildford Press.

Shavelson, R., Hubner, J., and Stanton, G. (1976). Self-Concept: Validation of Construct Interpretations. Review of Educational Research, 46(3), 407-441.

Shavelson, R. and Bolus, R. (1982). Self-Concept: The Interplay of Theory and Methods. Journal of Educational Psychology, 74(1), 3-17.

Shea, D., Lubinski, D., and Benbow, C. (2001). Importance of Assessing Spatial Ability in Intellectually Talented Young Adolescents: A 20-Year Longitudinal Study. Journal of Educational Psychology, 93(3), 604-614.

Shi, Y. (2017). Who Benefits from Selective Schools? Evidence from High-Achieving Math and Science Students in North Carolina. Working Paper, Society of Labor Economists Conference, May5-6, Raleigh, North Carolina.

Silverman, L. (1994). The Moral Sensitivity of Gifted Children and The Evolution of Society. Roeper Review, 17(2), 110-116.

Silverman, L. (1997). What we have learned about Gifted Children 1979-1997. Gifted Development Center, Denver, CO.

Silverman, L. (2000). Identifying visual-spatial and auditory-sequential learners: A validation study. In N. Colangelo & S. G. Assouline (Eds.), Talent development V: Proceedings from the 2000 Henry B. and Jocelyn Wallace National Research Symposium on Talent Development. Scottsdale, AZ: Gifted Psychology Press.

Silverman, L. (2002). Upside-Down Brilliance, DeLeon Publishing, Denver, Colorado.

Silverman, L. (2008). Theory of Positive Disintegration in the Field of Gifted Education. Chapter 8: in Mendaglio, S. (Ed.), Dabrowski's theory of positive disintegration, Scottsdale, AZ: Great Potential Press.

Simonton, D. (1985). Intelligence and Personal Influence in Groups: Four Nonlinear Models. Psychological Review. 92 (4), 532-547.

Simonton, D. (1991a). Career landmarks in science: Individual differences and interdisciplinary contrasts. Developmental Psychology 27, 119-130.

Simonton, D. (1991b). Emergence and Realization of Genius: The Lives and Works of 120 Classical Composers. Journal of Personality and Social Psychology. 61 (5), 829-840.

Simonton, D. (2003). When Does Giftedness Become Genius? And When Not? Chapter 28 in Colangelo, N. and Davis, G. (Eds.), Handbook of Gifted Education (3rd edition), Pearson Education.

Smyth, F. and McArdle, J. (2004). Ethnic and Gender Differences in Science Graduation at Selective Colleges with Implications for Admission Policy and College Choice. Research in Higher Education, 45 (4), 354-381.

Snijders, T. and Bosker, R. (2012). Multilevel analysis: an introduction to basic and advanced multilevel modeling. Los Angeles: SAGE.

Spence, M. (1973). Job Market Signaling. The Quarterly Journal of Economics, 87(3), 355-374.

Steele, C. and Aronson, J. (1995). Stereotype Threat and the Intellectual Test Performance of African Americans. Journal of Personality and Social Psychology, 69(5), 797-811.

Steele, C. and Aronson, J. (2007). Stereotypes and Fragility of Academic Competence, Motivation, and Self-Concept. Chapter 24 in Elliot, A. and Dweck, C. (Eds.), Handbook of Competence and Motivation, New York: Guilford Press.

Sternberg, R. (2000). Giftedness as Developing Expertise. In Heller, K, Monks, F., Sternberg, R., and

Sternberg, R. (2007). Intelligence, Competence, and Expertise. Chapter 2 in Elliot, A. and Dweck, C. (Eds.), Handbook of Competence and Motivation, Guildford Publications.

Subotnik, R. Pergamon (Eds.), International Handbook of Giftedness and Talent (2nd edition).

Sulloway, F. (1996). Born to rebel: birth order, family dynamics, and creative lives. New York: Pantheon Books.

Tanenbaum, A. (2003). Nature and Nurture of Giftedness. Chapter 4 in Colangelo, N. and Davis, G. (Eds.), Handbook of Gifted Education (3rd edition), Pearson Education, p45.

Tay, L. and Diener, E. (2011). Needs and Subjective Well-Being Around the World. Personality Processes and Individual Differences, 101(2), 354-365.

Todd, P. and Wolpin, K. (2003). On the Specification and Estimation of the Production Function for Cognitive Achievement. The Economic Journal, 113 (February), F3-F33.

Tcitovsky, T. (1992). The Joyless Economy: The Psychology of Human Satisfaction. New York, Oxford, Oxford University Press.

Tcitovsky, T. (1996). My own criticism of the joyless economy. Critical Review, 10(4), 595-605.

Teachers Federation & Federation of P & C Associates of NSW. (2002). Inquiry into the provision of public education in NSW.

Tiebout, C. (1956). A pure theory of local expenditures. Journal of Political Economy, 64 (5), 416-424.

Tincani, M. (2015). Heterogenous Peer Effects and Rank Concerns: Theory and Evidence. PhD Dissertation, University of Pennsylvania.

Towers, G. (1987). The Outsiders. Gift of Fire, 22.

Trautwein, U., Ludtke, O., Marsh, H., and Nagy, G. (2009). Within-School Social Comparison: How Students Perceive the Standing of Their Class Predicts Academic Self-Concept. Journal of Educational Psychology, 101(4), 853-866.

VanTessel-Baska (1985). Appropriate Curriculum for the Gifted. In J. Feldhusen (Ed.), Toward Excellence in Gifted Education, Love Publishing Company.

Veblen, T. (1899). The Theory of the Leisure Class. New York: MacMillan.

Verardi, V. and Debarsy, N. (2012). Robinson's square root of N consistent semiparametric regression estimator in Stata. Stata Journal, 12(4), 726-735.

Webb, J., Meckstroth, E., and Tolan, S. (1982). Guiding the Gifted Child: A Practical Source for Parents and Teachers. Hawker Brownlow Education.

Wise, A (2017). Behind Singapore's PISA rankings success – and why other countries may not want to join the race. 8 December 2017, The Conversation. https://theconversation.com/behind-singapores-pisa-rankings-success-and-why-other-countries-may-not-want-to-join-the-race-70057, accessed 22 November, 2017.

Woessman and Hanushek (2007). The Role of Education Quality in Economic Growth. World Bank Policy Research Working Paper 4122.

Wooldrigde, J. (2002). Econometric Analysis of Cross Section and Panel Data. The MIT Press. Cambridge, Massachusetts; London, England.

Yamane, S. and Hayashi, R., (2015). Peer Effects among Swimmers. The Scandinavian Journal of Economics, 117(4), 1230-1255.

Zen, K. (2016). The Impact of Selective High Schools on Student Achievement. Honours Research Essay in Economics (not accessible prior to 13 March 2018), University of New South Wales. https://www.business.unsw.edu.au/about/schools/economics/degrees/honours/current-past-honours-students.

Zhang, H. (2013). The Mirage of Elite Schools: Evidence from Lottery-based School Admission in China. unpublished manuscript, dated April 2013.