I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14168

Age of Starting School, Academic Performance, and the Impact of Non-Compliance: An Experiment within an Experiment, Evidence from Australia

Tony Beatton
Michael P. Kidd
Anthony Niu
Francis Vella

MARCH 2021

I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14168

Age of Starting School, Academic Performance, and the Impact of Non-Compliance: An Experiment within an Experiment, Evidence from Australia

Tony Beatton
University of Queensland

Michael P. Kidd
RMIT University and University of Queensland

Anthony Niu
RMIT University

Francis Vella
Georgetown University and IZA

MARCH 2021

ISSN: 2365-9793

IZA – Institute of Labor Economics

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

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

www.iza.org

ABSTRACT

Age of Starting School, Academic Performance, and the Impact of Non-Compliance: An Experiment within an Experiment, Evidence from Australia

This paper estimates the difference in academic performance of the oldest and youngest students in a given grade. We employ Queensland Department of Education school administration panel data for the population of state school students for the years 2008-2016. Academic performance is measured by National standard test scores (NAPLAN) and teacher assessed measures of performance and effort for individuals in grades 3, 5 and 7. The empirical analysis employs a regression discontinuity design (RDD) based on administrative rules on age of school enrolment. The class assigning mechanism operates via a known cut-off date and results in the oldest child in the grade being almost a year older than the youngest. However, as parents may anticipate a disadvantage in their child being the youngest in grade they may choose to delay the timing of initial enrolment. This lack of compliance potentially creates difficulties for the RDD identification strategy, in particular the assumption of exchangeability around the cut-off. We exploit a change in the cut-off rule from a 2008 reform which postponed the school starting age by 6 months and produced a large increase in the compliance rate. This enables one to gauge the importance of non-compliance in estimating the treatment effect of being older versus younger in cohort. We find that the pre-reform treatment effect is small and generally statistically insignificant. Post-reform there is a sizeable and statistically significant treatment effect which diminishes as the sample proceeds through school grades, 3, 5 and 7.

JEL Classification: I29

Keywords: national standard achievement tests, compliance rates, fuzzy regression discontinuity

Corresponding author:
Michael Kidd
Economics Finance and Marketing RMIT
Building 80
443 Swanston Street
Melbourne Vic, 3000
Australia

E-mail: Michael.kidd@rmit.edu.au

1. Introduction

Grades in school systems with initial enrolment determined by an administratively legislated cut-off birthdate comprise students born one day either side of the cut-off. Thus students within grade may differ by almost a year in age. A substantial empirical literature documents that the older members of a grade generally perform better on standardised cognitive tests than their younger counterparts. The differences may be large at entry, but tend to diminish as the child progresses through the school system (see, for example, Datar (2006), Cascio and Lewis (2006), Elder and Lubotsky (2009), Aliprantis (2014), Crawford et al. (2007, 2011, 2013, 2014), Bedard and Dhuey (2006) and Suziedelyte and Zhu (2015).[1] The existence, and persistence, of a positive "oldest in grade" treatment effect is important from a parental perspective, particularly for those with children born just before the cut-off date. Parents may consider it advantageous to hold back (i.e. redshirt) their child and consequently alter their relative position in the grade-age distribution**.** Persistence is also important from a policy perspective. Cunha et al. (2010) suggest that early childhood learning may have significant long-term impacts on the pattern of human capital acquisition, lifetime earnings, and, the economic wellbeing of nations (Cunha & Heckman, 2008, Heckman, 2011; Heckman, Pinto & Savelyev, 2013).

Parental discretion in complying with the administrative school admission criteria creates a difficulty in assessing the impact of being the "oldest in class" from observational data. With perfect compliance it is straightforward to compare the performance of the youngest with the oldest. The exercise appears ideal as a regression discontinuity design (RDD) (see, for example, Crawford et al 2014) as the point of the discontinuity is known and successful manipulation requires precise timing of the date of birth. However, without perfect compliance the parental decision to enroll a child in school may at least in part reflect a perception of the child's academic capacity. The enrolment outcome can then no longer be seen as an experiment in which the child's birthdate falls randomly on either side of the cut-off.

This paper provides important new evidence on the impact of being the eldest in grade for the population of students progressing through the Queensland State school system for the years 2008-2016. The data are uniquely appropriate for analysis of a number of policy issues as students are allocated to grades and initial enrollment on the basis of their birth dates. For school admissions in the earlier part of the data period, non-compliance with the administrative rule was prevalent. In 2007-8 Queensland instituted a set of educational reforms which included the introduction of a non-compulsory preparatory year in 2007

---

[1] Most countries admit children into school at the start of the academic year once they turn a certain age by a cut-off date, but the time window of academic year varies by different countries. In Australia it runs exactly in accordance with normal calendar year, while in the U.K an academic year runs from 1st September to 31st August. Relative age advantage (RAA) effects in this context refer to different academic performances between the youngest members and the oldest members in the same academic cohort. This study interchangeably uses the terms "RAA effects", "academic impacts of being amongst the oldest", and "treatment effects".

(prior to grade 1) combined with a 6-month delay in the cut-off age admission rule for grade 1, one year later. This altered the legislated starting age of school (grade 1) from the year in which the child turns 6 years of age (by December 31st) to a cut-off date of June 30th. The reforms increased the legislated starting age by 6 months which led to a substantial increase in compliance with the school administrative enrolment rule.

We view the reform as a natural experiment in that the remarkable increase in compliance was not its primary objective. Nevertheless, it creates variation in the data which may be exploited to address various key questions of interest. First, to what extent does the reform alter the pattern of non-compliance with the enrolment admission policy? Second, is compliance in the post reform period sufficiently high that RDD is an appropriate method and what do the results suggest about the oldest/youngest premium? Finally, what do the pre- and post-reform period results suggest about the objectives of parental non-compliance? We address this by estimating the difference in academic performance of the youngest and oldest in grade using national standardised test scores at Years 3, 5 and 7.[2] [3]

We acknowledge that in assessing the impact of relative age on performance that, due to their linear dependence, it is generally not possible to isolate the potential roles of age of starting school, age within grade (i.e. relative age) and age at time of test. The literature suggests that all three age related measures determine maturity level which, in turn, affects academic performance, see, for example, Crawford et al., 2013. The manner in which age at starting school and age at test date affect maturity are self explanatory. Relative age may operate through relatively younger students in grade having lower levels of cognitive and language development (National Research Council and Institute of Medicine, 2000), lower self-esteem and confidence (Thompson, Barnsley, & Battle, 2004; Thompson, Barnsley, and Dyck, 1999), worse social skills and a lower chance of being a leader in school activities (Dhuey & Lipscomb, 2008). The youngest in cohort may also be more likely to suffer from psychological disorders (Goodman, Gledhill, & Ford, 2003) and school victimisation (Mühlenweg, 2010). These considerations may impede academic achievement leading to reduced lifetime employment and earnings, and increase the cost to society of later human and economic remediation (Campbell, Pungello, Miller-Johnson, Burchinal, and Ramey, 2001; Heckman, Pinto and Savelyev, 2013; Heckman and Masterov, 2007; Heckman, 2011; Schweinhart, Montie, Xiang, Barnett, Belfield, and Nores, 2004)). In spite of the difficulty in assigning a clear interpretation to the

[2] As the focus is upon the impact of the change of the old/young premium pre/post reform- it is not possible to analyse NAPLAN grade 9.

[3] The most relevant study to ours is (Crawford et al., 2007). Using administrative data held by the U.K Department of Education, they show that September-born children (normally "the oldest" in a given school grade in the U.K) score, on average, over a standard deviation higher than those born in August ("the youngest") in national achievement test at age 7. Although the academic advantage of September-born in the cohort has diverged over time, it remains statistically significant at age 16- i.e., a critical age when the youth make decisions on whether to stay in education after compulsory schooling phase.

oldest/youngest premium it remains important to examine whether it exists and persists throughout the student's schooling process.

The following section describes our data and provides definitions of the key variables. Section 3 describes the experimental design and Section 4 presents our empirical results. We provide some concluding comments and discussion in Section 5.

2. Data and Definitions

We employ administrative records from the Department of Education (DoE), Queensland, which is the third most populous Australian state. The individual-level population data records all students enrolled in the state school sector from 2002 onwards through grade 1 to final grade upon leaving school. [4] We examine each child's performance in the National Assessments Program in Numeracy and Literacy (NAPLAN). The NAPLAN test assesses all Australian students in Years 3, 5, 7 and 9 and has been conducted yearly on the same days in May since 2008. The test consists of five dimensions designed to measure cognitive skills including maths and literacy, the latter based on reading, writing, grammar and spelling. The scores range from 0 to 1000.

The two scores we study are: 1) NAPLAN numeracy; and 2) NAPLAN literacy, calculated as the average of three separate test results for reading, spelling and grammar (following Au, Johnston, & Peeters (2014)).[5] Each score is standardised to have mean 0 and a standard deviation of 1 for each grade and year. An advantage of NAPLAN is that it is a nationwide test and an objective measure. Previous international studies have used similar national standard ability test scores (Black et al., 2011; Crawford, Dearden, & Greaves, 2011; Crawford et al., 2013; Crawford, Dearden, & Greaves, 2014; Crawford et al., 2007; Datar, 2006). Previous Australian research (see Suziedelyte and Zhu 2015) examined cognitive test scores measured by the Longitudinal Study of Australian Children (LSAC) i.e., the Peabody Picture Vocabulary Test (PPVT) and Matrix Reasoning test as well as non-cognitive skills. Jha (2015) uses NAPLAN data aggregated to school level. Our micro analysis uses individual student scores which allows us to control for heterogeneity in the student population.

We also analyse teacher-reported 5-scale grades which have been recorded by DoE since 2009 in three key learning areas (KLA), including English, Maths and Science, for all Year 1 to 12 students. Teachers/schools assess each discipline using grades from A to E to evaluate a student's achievement and effort.[6] These grades are inherently more subjective than NAPLAN but are curriculum-based and more

[4] Class IDs are only available across students enrolled at Year 1 to 8.

[5] Given a significant change in the evaluation of writing within NAPLAN over time we omit this measure.

[6] There is missing information for the behavioural measure so we do not analyse this dimension.

closely aligned with progression through the school system. The KLA results may capture features of a student's academic performance such as motivation, effort, and level of engagement which are not captured by objective tests.

Variables capturing date of birth and year of starting school are used to allocate students into grades although parental discretion may disturb this correspondence. DoE data also includes demographic information for each child (i.e., gender, ethnicity together with a list of basic school characteristics).

An important feature of the data period is the 2007-8 education reforms. In 2007, Queensland introduced a preparatory (Prep) year, a non-compulsory year prior to grade 1. The Queensland education system admits children in the year they turn six years of age by a specified cut-off date.[7] Prior to 2008 this date was 31st December, thereafter it was changed to 30th June. Our focus is on this change in the minimum enrolment age into grade 1, although it is not possible to separately identify the role of Prep.[8] Comparisons pre- and post- reform capture both parts of the 2007-8 reforms, see Jha (2015). The change to the minimum enrolment age implies an increase in the lowest enrolment starting age from 5 years to 5 and ½ years. Both pre- and post-reform, children following the enrolment admission rule in a given grade may vary in age by up to a year. If parents comply with the school entry rules (pre-reform) a child born on 31st December 1996, for example, will enter school at the start of calendar year in 2002 whereas a child born one day later (i.e., 1st Jan 1997) will begin in 2003. Post reform the cut-off date is June 30th, with children born on July 1st being one year older when commencing school.

The difference in starting age for grade 1 students is particularly important; post reform a child born at the beginning of the school enrolment window (1 July in the year prior to the year of enrolment) will be a year older than a child born on the last day of the school enrolment window (30 June of the year of enrolment). Assuming linear child development, a one-year difference represents a 17% increase in time for cognitive and language development for a 6-year-old child entering year one. For the youngest child in a class, this is equivalent to being one year developmentally behind the oldest student in their class.

Table 1 presents the contrasts facilitated by our data due to admission policies and the 2007-8 reform. The top half of the table documents school grade by calendar year and date of birth pre-reform and the bottom half presents the equivalent for post-reform. Similar to Jha (2015) we exclude those born in 2002 as that cohort is in the transitional year and represent a half cohort. NAPLAN assessment begins in 2008 and our focus is on grades 3, 5 and 7. Thus, pre-reform with a cut-off of January 1st and a window or bandwidth of 30 days we compare those born in the month of December 1996 who are in grade 7 in 2008 with those

---

[7] All Queensland children of eligible age are able to attend a full-time Preparatory Year before starting Year 1. The Preparatory Year attendance, however, is only mandatory from 2016. Given the inconsistency of this policy over the study period (2008 to 2016), this analysis uses the cut-off date for Year 1 entry to construct control and treatment groups.

[8] Our data fails to adequately capture prep enrolment.

born January 1997 in grade 7 in 2009. Adoption of a 30-day bandwidth either side of the cut-off follows much of the existing literature. Individuals born 1 day apart are in grade 7, 1 year apart, i.e., 2008 versus 2009. The table suggests we have insufficient data for both young and old pre- and post-reform for grade 9.

<Table 1>

The table also illustrates that for pre-reform grade 7 there are 5 waves of comparison between old and young. The number of equivalent waves for grade 5 is 3, while for grade 3 there is only 1. For post-reform, grade 3 involves 5 waves of comparisons, for grade 5 there are 3 waves, and for grade 7 we have 1 wave. The main analysis uses all waves for a given grade but as we also compare the pre- and post-reform periods, we estimate a symmetric version, i.e., where the number of waves is equal in both. The symmetric version for grade 3 will be a single wave as pre-reform is limited to 1 wave. Grade 5 is already symmetric, and grade 7 pre-reform will be a single wave. In the symmetric case the single wave is chosen to minimise the distance between pre- and post-reform. Finally, we also examine teacher ratings for achievement and effort for Math, English and Science. The ratings are available starting in 2009 which is one year later than the initiation of NAPLAN. This implies that we can no longer compare pre- and post-reform for grade 3. Grade 5 pre-reform will involve 2 waves rather than 3, and grade 7, 4 waves rather than 5. Post-reform waves are unaffected.

When parents comply with the age of starting school administrative rule the assignment of treatment is determined solely by birthdate. Prior to 2008, December-born and January-born are the youngest and the oldest in a given grade, respectively. Post-reform, June-born and July-born are the youngest and the oldest. However, with the school's permission, parents may fail to comply with the rules, by either holding back or advancing their child in terms of year of enrolment. As the vast majority of non-compliers are held back, we ignore those who start school early.

The sample we employ in our analysis depends on how we define the window around the exact cut-off date. Following earlier work (see, for example, Crawford et al. 2014) we include those born 30 days either side of the cut-off as our benchmark. This produces a sample of 87,988 (87,805) child-year observations for grades 3, 5 and 7 combined for strict compliers numeracy (literacy). For this sample, birthdate relative to the cut-off coincides precisely with the definition of younger versus older. Including non-compliers, the respective sample sizes are 104,140 for numeracy and 103,887 for literacy. This includes children starting school on time as advised by administrative rules (compliers) and children who are held back a year. The distinction between compliers and non-compliers is as follows:

1) **Compliers**: Children born either before or after the cut-off date who comply with the school entry rule to start grade 1 on time. Those born up to 30 days prior to the cut-off will be the youngest in grade and labelled as control, whereas those born up to 30 days after the cut-off will be oldest in grade and labelled the "treated".

2) **Non-compliers:** Children born before the cut-off (i.e. assigned to the control group), who fail to comply. These are assigned to the "youngest" but given non-compliance become members of "the oldest".

Within the context of RDD, there is a trade-off between defining a narrow window around the cut-off to minimise mis-specification of the local polynomial and a wider window so as to decrease variance of the estimated coefficients. To enable comparison of our results with the existing literature we choose a fixed baseline bandwidth of 30 days as our major focus. As estimates may be sensitive to the somewhat ad hoc choice of bandwidth we also explore an alternative optimal bandwidth that explicitly balances the trade-off between bias and variance (see Cattaneo et al 2019). This allows one to gauge sensitivity of our empirical results.

3. Education Reform in 2008 and Impacts on Compliance

A preparatory school year (Prep) was introduced in 2007 which provided children with a non-compulsory full-time year of early education prior to grade 1. Children had to be aged 5 years by June 30th in the year they began Prep. It was offered by all Queensland state schools and learning was carried out on school premises. The introduction of Prep was accompanied by an accomodating 6 month increase in the legislated starting school age for grade 1 in 2008 as noted above. The Queensland government also invested heavily in kindergarten services and facilities. The proportion of children attending a childhood program (pre-Prep) increased from 29% to 40% between 2008 and 2010. Although enrolment was not made compulsory until 2017, state-wide attendance rates for Prep were only slightly lower than for other primary year levels (QGOV, 2013).

The collective changes in the Queensland early childhood education landscape, including investment in kindergarden, introduction of Prep and increase in age of starting grade 1 represent a unified response to the overall goal of better preparing children for school. Thompson et al (2010) examine PISA assessments and conclude that in 2009 Queensland ranked 3rd among Australian states and territories in reading and maths and 4th in science.

Figure 1 provides the compliance rates in terms of grade 1 enrolment for our data. If everyone complied, those to the left of the cut-off i.e., born prior would begin grade 1 in sync with enrollment rules and thus would be younger in class. Those to the right of the cut-off, would be born too late in the year and would begin grade 1 a year later. The first panel aggregates the pre- and post-reform period and the 2nd and 3rd panels present the pre- and post- reform separately. For the pooled period, the compliance rate for around 60 days to the left of the cut-off date is around 80%. Non-compliance is monotonically increasing as the birth date approaches the cut-off, with non-compliance as high as 40% for birth dates close to the cut-off. For the pre-reform period, those born 2 months prior to the cut-off have around a 65% chance of complying although at dates close to the cut-off almost 60% are non-compliers. For the post-reform pattern, those born around 2 months prior to the cut-off exhibit a 90% compliance rate. The profile is upward sloping but relatively flat at around 20% close to the cut-off. This represents a dramatic change in degree of non-compliance between pre- (60%) and post- (20%) reform.

<Figure 1>

While we acknowledge the reform, seemingly responsible for these changes, would invalidate a comparison over the whole sample, it does not appear to contaminate examination of the oldest/youngest contrast in the post-reform period. Even if Prep improved the performance of all students the oldest/youngest comparison is still valid. While the lower compliance rates pre-reform suggest an RDD is not likely to identify a causal effect, post-reform the required assumptions appear more likely to be satisfied.

3.1 Regression Discontinuity Design

To identify the causal impact of the treatment effect of being oldest in cohort we employ an RDD (see, for example, Imbens and Lemieux, 2008). This estimates the mean treatment effect for subgroups of the population close to the discontinuity under minimal assumptions (Hahn, Todd and Van der Klaauw, 2001). The identifying assumption is that individuals do not select or manipulate entry into treatment based on expected gains. This may be violated if parents manipulate birth timing in relation to the school enrolment cut-off date or fail to comply with school entry enrolment rules. The former appears unlikely given the imprecision of planning of date of birth. However, we noted above substantial non-compliance with school administrative enrolment rules. This is particularly problematic if the propensity to comply is associated with some characteristics that also affect the child's academic performance. In addition, we require that the individual's other demographic characteristics do not vary substantially around the cut-off point used to assign changes in performance to the treatment. We examine this below.

The RDD requirement that the probability of treatment varies discontinuously at the respective cut-off point is clearly satisfied for both the pre- and post-reform periods. However, the probability does not increase from zero to one at the respective cut-offs. We address this below. With perfect compliance the strict RDD estimate of the "being oldest" treatment effect on test score is:

$$\boldsymbol{\beta_{STRICT} = E[(Test\,Score_i * (D=1) - Test\,Score_i * (D=0))|X, Z_i = \overline{Z_i}]} \quad \textbf{(1)}$$

where *Test Score* represents a child's NAPLAN numeracy or literacy in grade 3, 5 or 7; *D* is an indicator denoting whether or not the child is "oldest"; *Z* is the running variable capturing the time between date of birth and the cut-off date; $\overline{Z_i}$ denotes Z=0, *X* is a set of covariates including gender, indigenous status, language other than English spoken at home and cohort fixed effects.

With imperfect compliance we employ a fuzzy RDD approach (see Cattaneo, Idrobo and Titiunik, 2017):

$$\beta_{FUZZY} = \frac{E[(Test\ Score_i * D_i\ (1) - Test\ Score_i(0) * D_i\ (0))|X, Z_i = \overline{Z_i}\ ]}{E[(D_i\ (1) - D_i(0))|X, Z_i = \overline{Z_i}\ ]} \qquad \textbf{(2)}$$

The denominator of (2) accounts for the probability of compliance. β_{FUZZY} is the ratio between the strict RDD estimate and the average effect of the treatment-on-treatment take-up, both evaluated at $\overline{Z_\iota}$. The fuzzy RDD accounts for imperfect compliance but not for any violation of the identifying assumption. If this assumption is not satisfied then fuzzy RDD will produce inconsistent estimates of the treatment effect. We explore this below.

4. Results

Table 2 presents a set of balancing tests examining the distribution of personal characteristics either side of the cut-off. The variables of interest are an index of available school resources (sdindex); an indicator capturing whether the mother tongue is not English (mlote); a female indicator; an indigenous indicator; and a Community Socio-Educational Advantage index (icsea) which enables comparisons across schools by the socio-economic and educational advantage of their catchment areas (ACARA, 2017).

To check for continuity around the cut-off each of the variables described above is treated as the dependent variable in an RDD specification estimated over a 30-day window either side of the cut-off for pre- and post-reform. The results demonstrate that for the vast majority of personal characteristics the treatment is not statistically significantly different from zero. For the pre-reform period the only statistically significant treatment effect is for the female indicator which is negative and statistically significant at the 10% level for grade 7 only. This raises questions regarding the randomness of gender around the cut-off. The raw data reveal that 57% of males born in a 30-day window prior to the cut-off are held back compared to 40% for females. Dhuey et al (2019) also find higher levels of redshirting for males using US data. Once again, for the post-reform period the vast majority of treatment effects are not statistically significant at 5 percent, the only exceptions are the mlote and female variables for grade 3. To further explore the balancing tests, we treat the set of separate estimated equations as a seemingly unrelated regression and test the joint hypothesis that all treatment effects are jointly equal to zero. Results are recorded at the bottom of Table 2, estimates are pooled across dependent variables with separate estimates by grade and pre/post reform. Treatment effects are significantly different from zero at the 5% only pre-reform grade 7. Overall results are generally supportive of treating the covariates as balanced across the cut-off.

Figure 2 illustrates the NAPLAN numeracy scores at grades 3, 5 and 7 for pre- and post-reform respectively for the pooled sample of compliers and non-compliers. Figure 3 presents the corresponding figures for NAPLAN literacy. The horizontal axis represents the number of days before and after the legislated cut-off for school enrolment. The vertical line represents the eligibility cut-off date, normalised to zero. Figure 2 suggests small negative returns in test scores at the cut-off for the pre-reform period for all grades. Post-reform there is clear evidence of a large positive treatment effect. For grade 3 the “oldest” score is approximately 0.3 standard deviations higher in numeracy than the “youngest”. The scores in grade 5 exhibit a similar difference with some convergence by grade 7 with a gap of around 0.2 standard deviations. Figure 2 denotes the pattern for the combination of complier and non-compliers. This distinction is only relevant to the left of the cut-off, as we ignore the small number of children who enroll early in school. To the right of the cut-off are students born after the cut-off date, begin school the following year and are defined as older and consequently labelled as compliers. To the left of the cut-off there are a combination of two groups, younger compliers and older non-compliers. Both groups are born prior to the cut-off and according to the enrolment rules should enroll in the current year. Compliers are younger in cohort and in the control group. Non-compliers are held back for a year and are older in cohort, i.e., treated. In Figure 2, a comparison of the gap between NAPLAN scores to the left and right of the cut-off provides an initial estimate of the intention to treat (ITT).

<Figure 2 >

Figure 2 provides clear evidence of a larger empirical gap when comparing the left and right of the cutoff for the post-reform sample relative to that for the pre-reform. The ITT estimate of being older in cohort is much larger post-reform versus pre-reform. This may in part be due to the larger proportion of non-compliers pre versus post-reform. Intuitively, if the non-compliers do well by being held back, this would increase the average score to the left of the cut-off and the gap would be reduced. A similar set of conclusions hold for the literacy scores in Figure 3.

<Figure 3>

The results in Table 3 illustrate the estimated treatment effects for both numeracy and literacy pooled over pre- and post-reform periods. The table reports strict RDD results which employ a symmetric 30-day window. For grade 3 the ITT strict RDD treatment effect is 0.32; for grade 5 it is 0.23; and for grade 7 it is 0.08 The latter is not statistically significant at the 10% level. The literacy results display a very similar pattern to those for numeracy. The estimated treatment effect for grade 3 is 0.29; 0.21 for grade 5 and 0.08 (statistically insignificant) for grade 7.

< Table 3>

While these results are suggestive of a large “treatment” effect the key identifying assumption seems contentious. The high proportion of “held back non-compliers” in the pre-reform period suggests parents are concerned that their children are not adequately prepared for school. Accordingly, we examine the pre and post-reform periods separately. Our goal is not necessarily to compare the effects across periods as this would be contaminated by the introduction of the preparatory year. Our objective is to obtain estimates from the sample which best satisfies the required identification assumptions.

< Table 4 and Table 5>

Tables 4 and 5 present the results pre and post-reform. The first 2 columns are strict RDD estimates for pre and post-reform with a bandwidth of 30 days either side of the cut-off. Columns 3 and 4 allow the bandwidth to be chosen optimally i.e., to explicitly trade-off bias and variance using a MSE criteria. The optimal bandwidth pre-reform is 30 days, but post-reform it is 60 days. The advantage of comparing the fixed bandwidth with optimal choice is we can gauge the sensitivity of the estimates with respect to the bandwidth choice. The disadvantage of using an optimal bandwidth is that in comparing pre and post-reform results we are not holding the bandwidth constant. The remaining columns of Table 4 (and 5) present equivalent results i.e., fixed and optimal bandwidth for the fuzzy RDD. Finally, the strict RDD results represent intention to treat estimates, i.e., the impact of being assigned to treatment, rather than the impact of actually being older, i.e., treated. The fuzzy RDD estimates represent the impact of actual treatment for those that are treated close to the cut-off.

Most existing studies that estimate the treatment effect of being older in cohort use a strict rather than fuzzy approach. For example, Crawford et al (2014) note that non-compliance is a minor issue with around 99% of children in England located in the correct academic year. Clearly in the current context, Table 4, non-compliance, particularly pre-reform is an important issue. The first four columns present strict RDD estimates which may be given an ITT interpretation. The overall pattern is similar regardless of bandwidth. Consistent with Figure 2, we see small and generally statistically insignificant treatment effects pre-reform, with much larger effects, in terms of magnitude and statistical significance, post-reform. The overall pattern of treatment effects across grade illustrates a monotonic decline in magnitude post-reform, pre-reform it appears to be inverted-U shaped across grades. Focussing on pre-reform with the fixed 30-day bandwidth, ITT results are statistically insignificant at the 5 percent level for grades 3, 5 and 7. With the optimal bandwidth, treatment effects are of a similar magnitude but for grade 5 it is statistically significant at the 1% level. In contrast, the equivalent results for post-reform are large in magnitude and highly statistically significant at the 1% level for grades 3 and 5 with the 30-day bandwidth and for all grades with the optimal bandwidth. The magnitude declines from 0.37 for grade 3 to 0.29 grade 5 and to 0.21 for grade 7. The equivalent results for the optimal bandwidth are 0.36, to 0.27 to 0.15.

With the relatively high compliance rates post-reform our results are comparable to the existing literature for the UK. Crawford et al. (2014) report a treatment effect of 0.5 of a standard deviation lower for those born in August relative to September for children age 7. Bedard and Dhuey (2006) report similar effects for multiple countries with a range of patterns for non-compliance. For the UK for example, the treatment effect is 0.36. Our results also illustrate the declining importance of being relatively older as the grade level rises. This is consistent with U.K studies which use a similar approach (Crawford et al., 2007, 2011, 2014).

The pattern for fuzzy results post-reform is decreasing with grade, i.e., 0.435, 0.354 and 0.260 with grades 3 and 5 statistically significant at the 1 percent level and grade 7 at the 5% level. The overall conclusion of Table 4 is that treatment i.e., being "older" has a strong and significant impact on NAPLAN numeracy scores post-reform and a statistically insignificant impact pre-reform. Dhuey et al (2006) report results for grade 4 numeracy for countries with prominent non-compliance which generally reveal smaller treatment effects. This is similar to our pre-reform sample. Dhuey et al (2019) employ a fuzzy RDD approach using US administrative data and find a remarkably stable 0.2 standard deviation for the August /September treatment effect for pooled test results over grades 3 to 8. Their results are very stable over samples separated by socioeconomic status, race and gender. They also demonstrate that the extent of red-shirting varies substantially by these categories and argue that this parental intervention tends to reduce the treatment effect of "older" on test scores.

A final point related to the fuzzy RDD is that it estimates the LATE (i.e., local average treatment effect) for the sub-population of compliers defined at the cut-off. This local interpretation and lack of generalisability is due to the expectation that compliers and non-compliers differ. Bertanha and Imbens (2019) examine the conditions for external validity under which fuzzy RDD estimates may be used to extrapolate beyond compliers at the cut-off, both in terms compliance status and at points further than a local area around the cut-off. They provide a test for external validity which amounts to testing 2 restrictions jointly; (1) The expected value of Y for treated conditional on being at the cut-off and being a complier is equal to the same for always takers, and (2) The expected value of Y for non-treated (control) conditional on being at the cut-off and being a complier is equal to the same for never takers. In our one-sided non-complier context, (i.e., to the right of the cut-off we have sharp RDD), we just need to test condition (1).[9] We test equality of treatment effects for treated compliers and always takers. We do not reject the null hypothesis of equality and conclude that our fuzzy estimates are defined more widely than the usual LATE. The p-value for the test with and without covariates is 0.522 and 0.675 respectively.

The equivalent results for literacy are provided in Table 5. Overall, the results are similar. Post-reform results are generally much larger than pre-reform. The influence of bandwidth, i.e., the comparison between 30 days and the optimal choice of bandwidth, provides similar magnitude of effects, although the statistical significance is often higher in the optimal versus the fixed bandwidth.

Tables 4 and 5 indicate that the large and positive treatment effects reported in Table 3 for the pooled pre- and post-reform sample are largely driven by the post reform period. The post reform treatment

[9] We utilize the stata utility rdexo.

effects are around 2 to 3 times those of the pre-reform. Given the high compliance rate, it is not surprising that for the post-reform period the fuzzy and sharp results are generally similar. The results are highly suggestive of a large treatment effect and although it is decreasing as the child progresses through the schooling process there is evidence that at year 7 it is still quite large.[10]

The difference in the compliance rates across the regimes reflects the parents' decisions. Parents appear more comfortable in complying with the legislated school starting age in the post-reform period than the pre-reform period. Comparing the pre-reform ITT strict RDD results with the fuzzy RDD results reveals that the latter estimated treatment effects for all grades are much larger in magnitude. This suggests parents are keeping back children who would not do well, and this is masking the presence of a treatment effect. As anticipated, the gap between fuzzy and strict is smaller post versus pre-reform, reflecting the difference in non-compliance.

There is a significant change in pattern of non-compliance with school enrolment rules pre- versus post-reform. It is interesting to ask why parents appear to have such a strong reaction, in terms of compliance, to a small increase in the school starting age. However as discussed earlier, the change in the minimum starting age was part of a sequence of reforms including the introduction of the preparatory year. It suggests that parents believe that the reforms 2007-8 imply the children are better prepared to begin school on time. The lower compliance in the pre-reform period is an indication of concerns regarding the child's preparedness rather than age differences per se within the class. The strong parent reaction to enrolment compliance may also in part due to the introduction of 'compulsory schooling' by Queensland Education in 2007, accompanied by legal obligations that enforced enrolment of a child that is considered to be of compulsory school age from 6 years and 6 months until they turn 16 (QGOV2, 2007). To not enroll a child in grade one by six years and six months required special consideration by the Principal of the student's school, with the list of acceptable excuses for non-enrolment very limited.

A limitation of our results is that the treatment effect is based on a comparison of the oldest and youngest in grade who are born one day apart. While this is likely to be informative of the performance of those with birthdates close to the cut-off it does not capture larger age differences. To extend the policy relevance of our results we expand the control and treatment effects in a number of distinct ways. First, we aggregate the birthdates into 2 day and one-week intervals and redo our empirical work. Second, we delete the very youngest from the sample and re-estimate the treatment effects. Finally, we also simultaneously delete the very youngest and very oldest from the sample and examine results.

---

[10] To check for sensitivity of results we also estimate symmetric pre- and post-reform specifications i.e. grade 3 with 2 years, grade 5 as above 5 years and grade 7, 2 years. The results are extremely similar.

A series of panels marked Table 6 examine the robustness of our earlier results for NAPLAN numeracy grades 3, 5 and 7 to the above redefinitions of treatment and control groups. The layout of Table 6 replicates that of Tables 4 and 5. Each panel displays pre- and post-reform estimates. In section (i) the results begin with aggregation across 2 days followed by those for 7 days. For the aggregated groups we set the window symmetrically to 15 days for the 2-day aggregation, and 5 for the 7-day aggregation. Starting with the strict RDD results for the pre-reform period, the overall pattern for the two-day aggregated results is very similar to Table 4. Magnitudes are very close; the only minor aberration is the treatment effect for grade 5 is now statistically significant at the 1% rather than 10% level as in Table 4. The post-reform treatment effect continues to be much larger in magnitude and is statistically significant at 1% level for all grades 3, 5 and 7. and the aggregated fuzzy patterns both pre- and post-reform are also similar. Thus, results for 2-day aggregation closely replicate our initial results.

<Table 6>

Results aggregating over 7 days are presented in the next panel of Table 6. The overall pattern for strict RDD treatment effects continues to hold, with no pre-reform estimates statistically significant even at the 10% level and all effects are statistically significant at the 1% level post-reform. In terms of post-reform, the 7-day aggregated results for grades 3, 5 and 7 are 0.38, 0.32 and 0.26; these are very close to those reported in Table 4, i.e., 0.37, 0.29 and 0.21. Turning to fuzzy RDD results, for pre-reform 7 day aggregated results the only statistically significant treatment effect is at the 10% for grade 5. Once again, all post-reform grades are statistically significant at the one percent level. The magnitude of aggregated results versus original Table 4 results are similar. Thus, the overall conclusion is that results in terms of overall pattern are not sensitive to this level of aggregation. This suggests that an “older treatment effect” exists for comparisons beyond the very youngest and oldest in class.

Section (ii) of the table deletes the youngest individuals. We first delete those born one day prior to the cut-off and then extend this to 7- and 14-day deletions. These deletions are not symmetric i.e., only those located to the left of the cut-off are deleted. Finally, we replicate the analysis deleting symmetrically across the cut-off. Results are not sensitive to any of these robustness checks.

Given our earlier finding that gender differences exist in compliance rates, Table 7 presents the main results separately by gender. We focus on the post-reform results given the pre-reform estimates are likely biased although we report them here for completeness. The effects are generally larger for boys although large effects also exist for girls. We also find that the fuzzy RDD results are closer to the strict RDD results for girls than for boys due to the former’s much higher compliance rates. Boys are “red-shirted” twice as frequently as girls, and their test scores are lower. Recent work by Dhuey et al (2019) also demonstrates a much higher degree of red-shirting for boys.

< Table 7>

It is useful to contrast our results with earlier evidence from Australia. Suziedelyte and Zhu (2015) use the Longitudinal survey of Australian Children (LSAC) which includes a series of cognitive test results (e.g., Peabody Picture Vocabulary test PPVT)) in addition to non-cognitive skills captured by the Strengths and Difficulties Questionnaire. They focus on the impact of age starting school on performance at age 6-7 and thus unlike our study focus on a single early grade and use data from 4 distinct Australian states and territories. Their data incorporates students from both public and private schools, whereas our study only tracks those through the public system. They use a similar approach to that employed by the current study, i.e., a fuzzy RDD. Their results suggest that being younger in cohort improves cognitive skills but has a negative impact on non-cognitive skills. Their major focus is on the impact of being younger versus older in cohort for students coming from advantaged versus disadvantaged backgrounds. They suggest that starting school early narrows the gap between the two sets of households for cognitive skills, but there is no equivalent catch-up in non-cognitive skills. Our results are not directly comparable given the difference in adopted measures. However, the overall conclusion is clearly different, as our results suggest that there is a premium to being older rather than younger in cohort.

The second Australian study, by Page et al., 2017 uses a totally different approach and focus to that of the current study. Their paper devises an interesting experimental design, examining whether being younger in cohort shapes student preferences for competition. They conclude that older students, particularly males, have a stronger preference for competition than younger in cohort. Interestingly, their evidence does not support stronger preference for competition arising via an increase in self-confidence or willingness to take risks associated with being older. Thus, the overall tenet of their results suggests an advantage associated with being oldest in class and thus is in line with our results.

Finally, Jha (2015) uses a similar approach to the current study, using NAPLAN data for both Queensland and West Australia. The important distinction is that their analysis adopts a difference in difference approach and uses aggregated school level data rather than individual level. Their analysis examines the impact of being oldest in cohort pre- and post-reform with grade 3 as the treated group, and grade 5 as the control. This is quite distinct in terms of control versus treatment and the focus on aggregate school level data rather than individuals. This makes it difficult to compare results across studies. However, in general the results show a positive impact on NAPLAN scores of the reforms i.e., the increase in age of starting school and the introduction of prep.

Table 8 reports the estimates from an analysis of the KLA scores which are an alternative to NAPLAN. As the scores are reported as letter grades A-E we code the highest score for achievement and effort as 5, with the lowest coded as 1. The format of Table 8 follows a similar structure to the previous tables. The rows refer to teacher assessed achievement and effort for English, maths and science for grades 5 and 7. Results for grade

3 are omitted as teacher assessed scores only began in 2009 and we do not have data for the pre-reform period. Results are presented separately by pre- and post-reform periods. The results appear consistent with the earlier analysis of NAPLAN data. We continue to view only the post reform evidence as informative. As the outcome variable is now measured in different units the treatment effects are no longer directly measured in fractions of a standard deviation and the magnitude of the estimates are not directly comparable to those for the NAPLAN data. The treatment effects in the case of math achievement and effort for grade 5 post reform are 0.36 and 0.173 respectively, both significant at the 1% level. The associated raw means are around 3.4 and 4.1 and accompanying standard deviations 0.96 and 0.89 respectively. Thus, roughly being older in cohort increases math achievement (effort) by around 0.4 (0.2) standard deviations respectively. The results once again reaffirm the benefits of being older in cohort in both achievement and effort levels. Results are also compatible in general with a declining treatment effect by grade. The overall evidence from the analysis of the KLA scores supports the earlier conclusion that there is a large advantage in being the oldest over the youngest.

4. Conclusion

This paper examines the achievement premium for being the oldest rather than the youngest student in the grade. We do this by estimating the impact of day of birth on national standard test scores in Queensland. A challenge to addressing this issue is the role of parents in complying with legislated school starting dates. We overcome this hurdle by exploiting a policy reform which delayed the school starting age by only 6 months but which drastically increased compliance rates.

By estimating the achievement premium separately over the pre- and post-reform periods we find evidence that the premium is substantial at younger ages but declines with progress through school. Nevertheless, for the time period we examine the effects still exist when we last observe the children, grade 7. Moreover, the nature of the reform also allows us to examine the behaviour of the parents. We find that at younger school starting ages the parents appear to be assessing their child’s likelihood of success and delaying school enrolment where necessary. This appears to be reducing the observed treatment premium. Most interestingly, the parents appear to drastically change their behaviour in response to a minor increase in their child’s age at the time they need to enrol.

REFERENCES:

Aliprantis, D. (2014). When should children start school? *Journal of Human Capital, 8*(4), 481-536.

Black, N., Johnston, D. W., and Peeters, A. (2015). Childhood obesity and cognitive achievement. *Health economics*, *24*(9), 1082-1100.

Bedard, K., Graand Dhuey, E. (2006). The persistence of early childhood maturity: International evidence of long-run age effects. *The Quarterly Journal of Economics*, 1437-1472.

Bertanha, M and Guido W Imbens, (2019) External validity in fuzzy regression discontinuity designs, Journal of Business and Economics Statistics, 38 (3), 593-612.

Black, S. E., Devereux, P. J., and Salvanes, K. G. (2011). Too young to leave the nest? The effects of school starting age. *The Review of Economics and Statistics, 93*(2), 455-467.

Campbell, Frances A., Elizabeth Pungello, Shari Miller-Johnson, Margaret Burchinal, and Craig T. Ramey. (2001).The Development of Cognitive and Academic Abilities: Growth Curves From an Early Childhood Educational Experiment. *Developmental Psychology*, 37(2), 231–242. https://doi.org/10.1037/0012-1649.37.2.231

Cascio, E. U., and Lewis, E. G. (2006). Schooling and the Armed Forces Qualifying Test Evidence from School-Entry Laws. *Journal of human resources, 41*(2), 294-318.

Crawford, C., Dearden, L., and Greaves, E. (2011). Does when you are born matter? The impact of month of birth on children's cognitive and non-cognitive skills in England. *Month, 1*, 11.

Crawford, C., Dearden, L., and Greaves, E. (2013). Identifying the drivers of month of birth differences in educational attainment. *Institute for Fiscal Studies (IFS), Working Paper*(13/09).

Crawford, C., Dearden, L., and Greaves, E. (2014). The drivers of month-of-birth differences in children's cognitive and non-cognitive skills. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 177*(4), 829-860.

Crawford, C., Dearden, L., and Meghir, C. (2007). *When you are born matters: the impact of date of birth on child cognitive outcomes in England*: Centre for the Economics of Education, London School of Economics and Political Science.

Cunha, F. and Heckman, J. (2008). Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Journal of Human Resources*, 43(4), 738-782.

Datar, A. (2006). Does delaying kindergarten entrance give children a head start? *Economics of Education Review*, *25*(1), 43-62.

Dhuey, E., and Lipscomb, S. (2008). What makes a leader? Relative age and high school leadership. *Economics of Education Review, 27*(2), 173-183.

Dhuey, E., Figlio, D., Karbownik , K. and Roth, J. (2019). School Starting Age and Cognitive Development. *Journal of Policy Analysis and Management* 38(3), 538-578.

Elder, T. E., and Lubotsky, D. H. (2009). Kindergarten entrance age and children's achievement impacts of state policies, family background, and peers. *Journal of human resources, 44*(3), 641-683.

Fredriksson, P., and Ockert, B. (2005). Is early learning really more productive? The effect of school starting age on school and labor market performance.

Goodman, R., Gledhill, J., and Ford, T. (2003). Child psychiatric disorder and relative age within school year: cross sectional survey of large population sample. *Bmj*, *327*(7413), 472.

Hahn, J., Todd, P., and Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, *69*(1), 201-209.

Heckman, J. J., and Masterov, D. V. (2007). The productivity argument for investing in young children. *Applied Economic Perspectives and Policy*, 29(3), 446-493.

Heckman, J. J. (2011). The economics of inequality: The value of early childhood education. *American Educator*, 35(1), 31.

Heckman, J. Pinto, R, and Savelyev, P. (2013). Understanding the Mechanisms Through Which an Influential Early Childhood Program. Boosted Adult Outcomes. *American Economic Review*, 103(6), 2052-2086.

Imbens, G. and Lemieux, T. (2008) Regression Discontinuity Designs: A Guide to Practice. *Journal of Econometrics*, 142, 615-635.

Jha, N. (2015). Late Start with Extra Schooling: The Effect of Increase in School Entry Age and Preschool Provision. *Economic Record* 91(S1), 54-77.

Mühlenweg, A. (2010). Young and innocent: international evidence on age effects within grades on victimization in elementary school. *Economics Letters*, *109*(3), 157-160.

Mühlenweg, A., Blomeyer, D., Stichnoth, H., and Laucht, M. (2012). Effects of age at school entry (ASE) on the development of non-cognitive skills: Evidence from psychometric data. *Economics of Education Review*, *31*(3), 68-76.

Mühlenweg, A., and Puhani, P. (2010). The evolution of the school-entry age effect in a school tracking system. *Journal of human resources, 45*(2), 407-438.

Musch, J., and Grondin, S. (2001). Unequal competition as an impediment to personal development: A review of the relative age effect in sport. *Developmental review*, *21*(2), 147-167.

National Research Council and Institute of Medicine, (2000). *From Neurons to Neighborhoods: The Science of Early Childhood Development*. Committee on Integrating the Science of Early Childhood Development. Jack P. Shonkoff and Deborah A. Phillips, eds. Board on Children, Youth, and Families, Commission on Behavioral and Social Sciences and Education. Washington DC.: National Academy Press.

Page, L., Sarkar, D., and Silva-Goncalves, J. (2017). *The older the bolder: Does relative age among peers influence children's preference for competition?* Journal of Economic Psychology 63, 43-81.

Pellizzari, M., and Billari, F. C. (2012). The younger, the better? Age-related differences in academic performance at university. *Journal of Population Economics*, *25*(2), 697-739.

Puhani, P. A., and Weber, A. M. (2007). Does the early bird catch the worm? *Empirical Economics*, *2*(32), 359-386.

Smith, J. (2009). Can regression discontinuity help answer an age-old question in education? The effect of age on elementary and secondary school achievement. *The BE Journal of Economic Analysis & Policy*, *9*(1).

QGOV, (2013). Performance Insights: School Attendance. Queensland Government. accessed 8 January 2021 https://education.qld.gov.au/initiativesstrategies/Documents/performance-insights-report.pdf

QGOV2, (2007). Enrolment in state primary, secondary and special schools. Queensland Department of Education. Queensland Government. Brisbane , Australia. accessed 12 January 2012 https://ppr.qed.qld.gov.au/education/management/Procedure%20Attachments/Enrolment%20in%20State%20Primary,%20Secondary%20and%20Special%20Schools/enrolment-in-state-primary-secondary-special-schools.pdf

Suziedelyte, A., and Zhu, A. (2015). Does early schooling narrow outcome gaps for advantaged and disadvantaged children? *Economics of Education Review*, *45*, 76-88.

Schweinhart, Montie, Xiang, Barnett, Belfield, and Nores, (2004). The High/Scope Perry Preschool Study Through Age 40 Summary, Conclusions, and Frequently Asked Questions. Highscope Press.

Thompson, A. H., Barnsley, R. H., and Battle, J. (2004). The relative age effect and the development of self-esteem. *Educational Research*, *46*(3), 313-320.

Thompson, A. H., Barnsley, R. H., and Dyck, R. J. (1999). A new factor in youth suicide: The relative age effect. *The Canadian Journal of Psychiatry*, *44*(1), 82-85.

Figure 1: Probability of treatment: Pooled pre-reform and post-reform

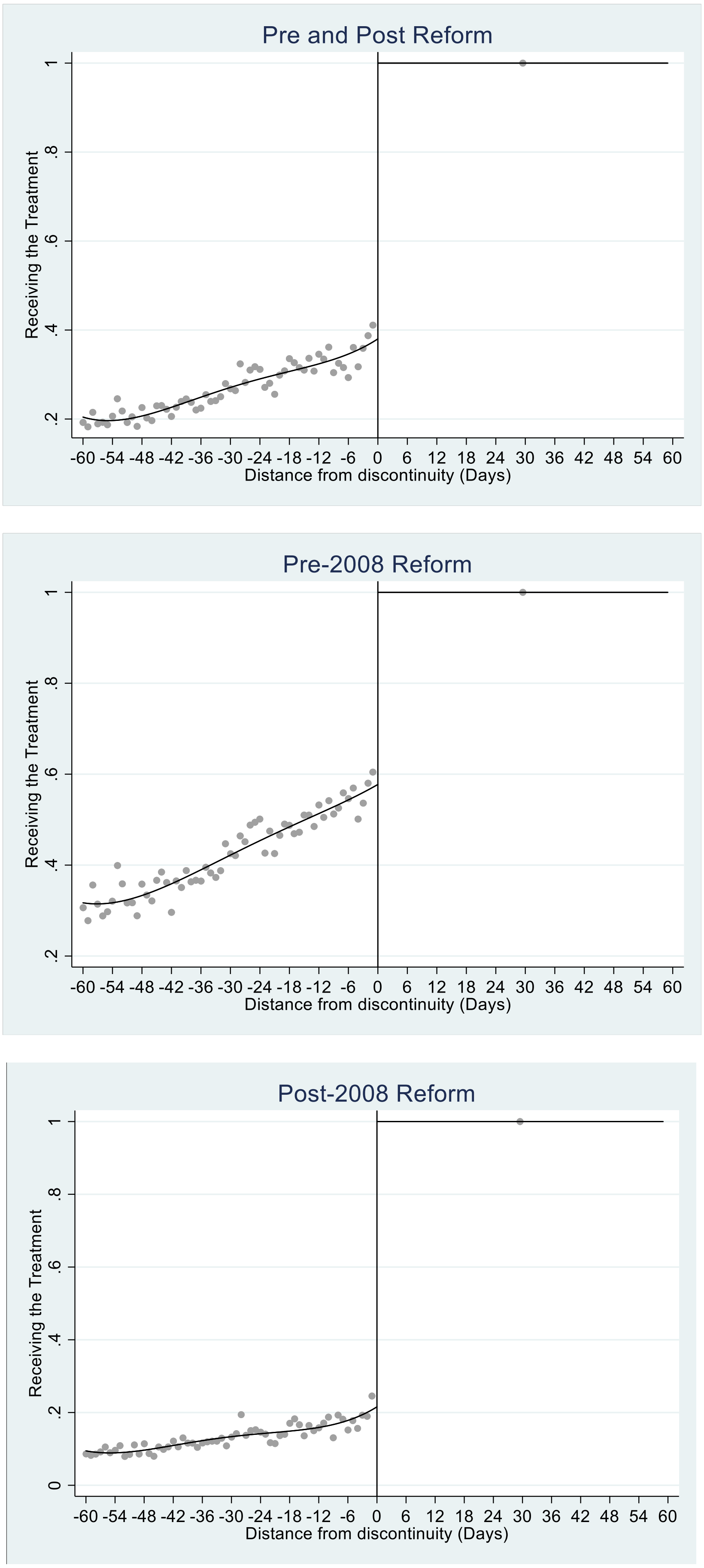

Figure 2: Numeracy Naplan scores pre post reform: Complier and Non-complier

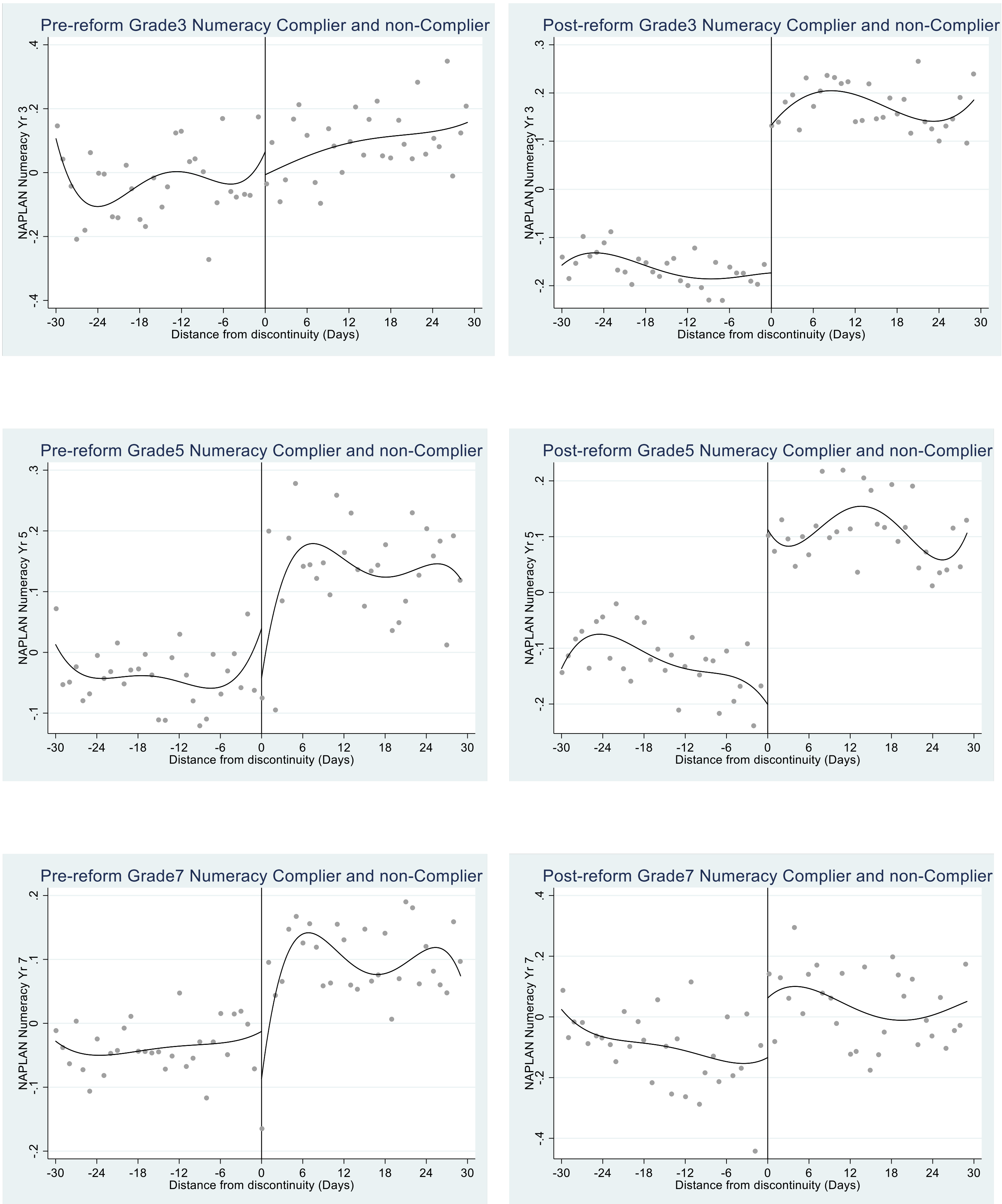

Figure 3: Graphs for literacy pooled

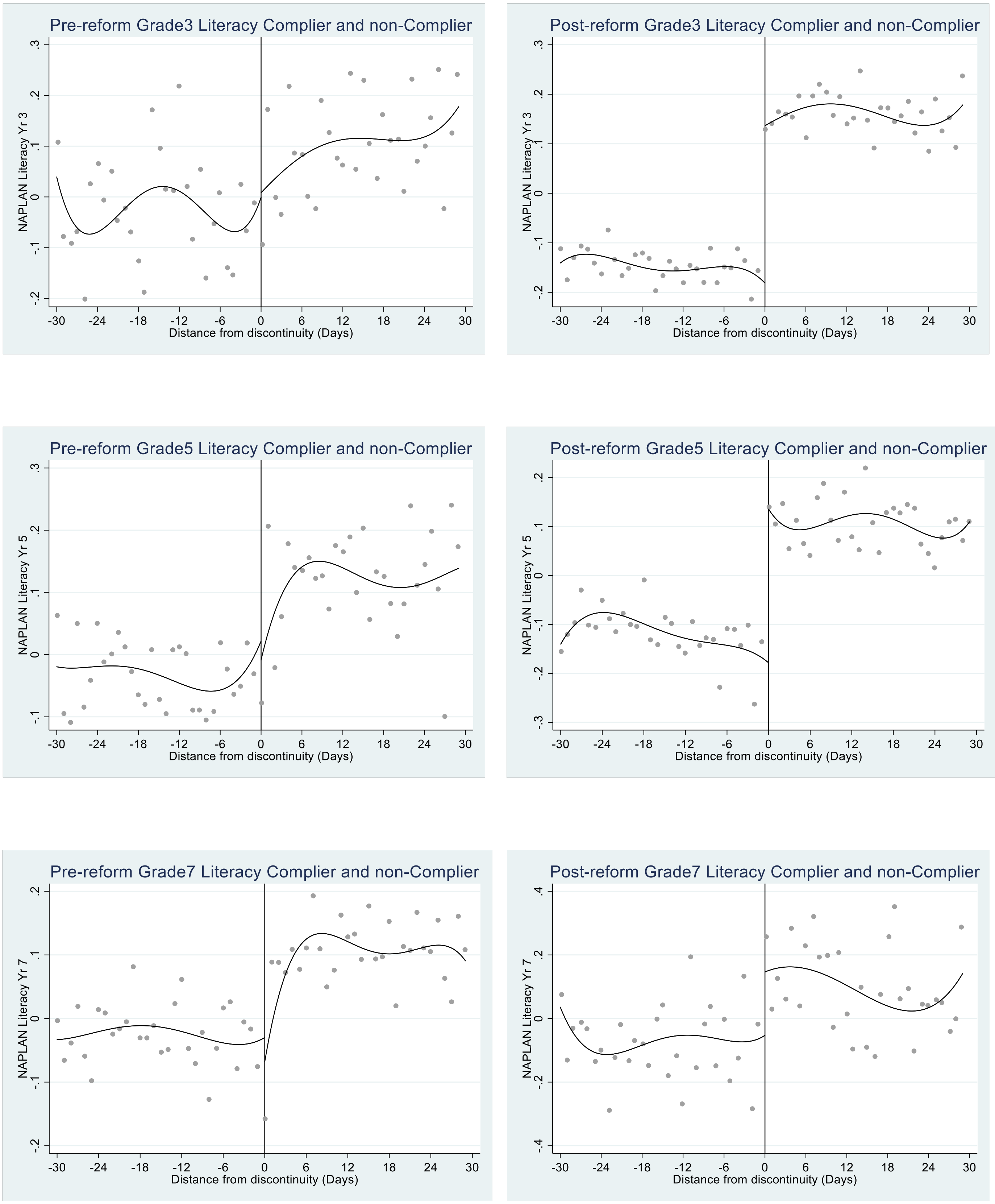

Table 1: Definition of young versus old pre- and post-reform

	Month born	Year born	2008	2009	2010	2011	2012	2013	2014	2015	2016
Pre-reform	June	1996									
	Dec	1996	Year 7 - Young								
	June	1997		Year 7 - Old							
	Dec	1997		Year 7 - Young							
	June	1998			Year 7 - Old						
	Dec	1998	Year 5 - Young		Year 7 - Young						
	June	1999		Year 5 - Old		Year 7 - Old					
	Dec	1999		Year 5 - Young		Year 7 - Young					
	June	2000			Year 5 - Old		Year 7 - Old				
	Dec	2000	Year 3 - Young		Year 5 - Young		Year 7 - Young				
	June	2001		Year 3 - Old		Year 5 - Old		Year 7 - Old			
Post-reform	June	2003				Year 3 - Young		Year 5 - Young		Year 7 - Young	
	Dec	2003					Year 3 - Old		Year 5 - Old		Year 7 - Old
	June	2004					Year 3 - Young		Year 5 - Young		
	Dec	2004						Year 3 - Old		Year 5 - Old	
	June	2005						Year 3 - Young		Year 5 - Young	
	Dec	2005							Year 3 - Old		Year 5 - Old
	Jun	2006							Year 3 - Young		
	Dec	2006								Year 3 - Old	
	June	2007								Year 3 - Young	
	Dec	2007									Year 3 - Old

Table 2: Balance check: strict RDD

	Compliers and non-compliers	
	Pre-reform	**Post reform**
sdindex		
Grade 3	0.775	-0.245
Robust 95% CI	[-9.22; 11.23]	[-3.120; 2.402]
Eff. N of obs	2,573/2,791	15,491/16,241
N of waves	1	5
Grade 5	-2.362	-0.599
Robust 95% CI	[-8.837; 0.988]	[-2.927; 1.857]
Eff. N of obs	7,891/8,684	8,876/9,232
N of waves	3	4
Grade 7	-1.454	-1.376
Robust 95% CI	[-13.583; 10.148]	[-3.495; 5.894]
Eff. N of obs	13,129/14,176	2,504 2,552
N of waves	5	2
mlote		
Grade 3	0.003	-0.015***
Robust 95% CI	[-0.089; 0.115]	[-0.041; -0.009]
Eff. N of obs	2,573/2,791	15,491/16,241
N of waves	1	5
Grade 5	0.005	-0.009*
Robust 95% CI	[-0.054; 0.086]	[-0.046; 0.002]
Eff. N of obs	7,891/8,684	8,876/9,232
N of waves	3	4
Grade 7	0.035	0.0004
Robust 95% CI	[-0.020; 0.139]	[-0.043; 0.050]
Eff. N of obs	13,129/14,176	2,504/2,552
N of waves	5	2
female		
Grade 3	-0.007	0.005**
Robust 95% CI	[-0.058; 0.051]	[0.001; 0.052]
Eff. N of obs	2,573/2,791	15,491/16,241
N of waves	1	5
Grade 5	-0.008	0.021**
Robust 95% CI	[-0.025; 0.061]	[0.009; 0.071]
Eff. N of obs	7,891/8,684	8,876/9,232
N of waves	3	4
Grade 7	-0.032*	-0.039
Robust 95% CI	[-0.069; 0.006]	[-0.159; 0.031]
Eff. N of obs	13,129 14,176	2,054/2,552
N of waves	5	2
indigenous		
Grade 3	0.012	0.013*
Robust 95% CI	[-0.026; 0.058]	[-0.002; 0.036]
Eff. N of obs	2,573/2,791	15,491/16,241
N of waves	1	5

Table 2. Continued

	Compliers and non-compliers	
	Pre-reform	**Post reform**
Indigenous (continued)		
Grade 5	0.007	0.002
Robust 95% CI	[-0.007; 0.043]	[-0.015; 0.032]
Eff. N of obs	7,891/8,684	8,876/9,232
N of waves	3	4
Grade 7	0.005	0.013*
Robust 95% CI	[-0.008; 0.027]	[-0.003; 0.054]
Eff. N of obs	13,129/14,176	2,504/2,552
N of waves	5	2
icsea		
Grade 3	-0.788	1.270
Robust 95% CI	[-14.031; 11.322]	[-1.812; 3.655]
Eff. N of obs	2,573/2,791	15,491/16,241
N of waves	1	5
Grade 5	1.965	0.478
Robust 95% CI	[-1.943; 8.169]	[-2.396; 2.873]
Eff. N of obs	7,891/8,684	8,876/9,232
N of waves	3	4
Grade 7	1.605	4.998*
Robust 95% CI	[-2.258; 5.660]	[-0.727; 9.064]
Eff. N of obs	13,129/14,176	2,504/2,552
N of waves	5	2
SURE joint test of treatment effects=0		
Grade 3 Chi squared (5 df)	2.40	10.12*
P value	0.7916	0.0719
Grade 5 Chi squared (5 df)	3.73	1.80
P value	0.5889	0.8757
Grade 7 Chi squared (5 df)	24.47***	7.99
P value	0.0002	0.1570

Table 3: Pooled Pre and Post reform Strict RDD: Numeracy and Literacy

	30 days
	Pre and Post
Numeracy	
Grade 3	0.323***
Robust 95% CI	[0.269; 0.325]
Eff. N of obs	18,064/19,032
N of waves	6
Grade 5	0.231***
Robust 95% CI	[0.142; 0.235]
Eff. N of obs	16,767/17,916
N of waves	7
Grade 7	0.077
Robust 95% CI	[-0.042; 0.116]
Eff. N of obs	15,633/16,728
N of waves	7
Literacy	
Grade 3	0.287***
Robust 95% CI	[0.235; 0.315]
Eff. N of obs	18,016/19,006
N of waves	6
Grade 5	0.209***
Robust 95% CI	[0.128; 0.220]
Eff. N of obs	16,746/17,904
N of waves	7
Grade 7	0.080*
Robust 95% CI	[-0.010; 0.114]
Eff. N of obs	15,572/16,643
N of waves	7

Table 4: Strict and Fuzzy RDD: Numeracy

	Strict RDD fixed bandwidth		Strict RDD optimal bandwidth		Fuzzy RDD fixed bandwidth		Fuzzy RDD optimal bandwidth	
	Pre-reform	**Post reform**	**Pre-reform**	**Post reform**	**Pre-reform**	**Post reform**	**Pre-reform**	**Post reform**
First-Stage Grade 3					0.435	0.853	0.435	0.849
z-value (robust)					8.08	37.598	13.054	91.839
Robust 95% CI					[0.318; 0.522]	[0.803; 0.892]	[0.369; 0.500]	[0.828; 0.864]
Grade 3	0.023	0.371***	0.023	0.362***	0.053	0.435***	0.052	0.426***
Robust 95% CI	[-0.082; 0.169]	[0.308; 0.365]	[-0.101; 0.099]	[0.330; 0.390]	[-0.179; 0.383]	[0.359; 0.436]	[-0.228; 0.225]	[0.391; 0.461]
Eff. N of obs	2,573/2,791	15,491/16,241	2,573/2,791	31,807/32,399	2,573/2,791	15,491/16,241	2,573/2,791	31,807/32,399
N of waves	1	5	1	5	1	5	1	5
Bandwidth	30	30	30	60	30	30	30	60
First-Stage Grade 5					0.571	0.829	0.559	0.84
z-value (robust)					43.347	42.434	53.438	73.195
Robust 95% CI					[0.543; 0.595]	[0.766; 0.840]	[0.539; 0.580]	[0.814; 0.858]
Grade 5	0.162*	0.293***	0.184***	0.267***	0.284*	0.354***	0.329***	0.318***
Robust 95% CI	[-0.003; 0.215]	[0.199; 0.344]	[0.104; 0.252]	[0.230; 0.317]	[-0.009; 0.383]	[0.249; 0.428]	[0.181; 0.455]	[0.274; 0.380]
Eff. N of obs	7,891/8,684	8,876/9,232	12,716/13,730	18,480/18,657	7,891/8,684	8,876/9,232	12,716/13,730	18,480/18,657
N of waves	3	4	3	4	3	4	3	4
Bandwidth	30	30	47	61	30	30	47	61
First-Stage Grade 7					0.643	0.801	0.645	0.812
z-value (robust)					31.867	29.218	53.658	54.077
Robust 95% CI					[0.596; 0.674]	[0.722; 0.826]	[0.620; 0.667]	[0.781; 0.840]
Grade 7	0.055	0.208*	0.072*	0.145***	0.086	0.260**	0.112*	0.178***
Robust 95% CI	[-0.087; 0.107]	[-0.007; 0.404]	[-0.002; 0.128]	[0.058; 0.253]	[-0.138; 0.171]	[0.011; 0.504]	[-0.004; 0.199]	[0.075; 0.309]
Eff. N of obs	13,129/14,176	2,504/2,552	20,984/22,270	5,815/5,931	13,129/14,176	2,504/2,552	20,984/22,270	5,815/5,931
N of waves	5	2	5	2	5	2	5	2
Bandwidth	30	30	46	69	30	30	46	69

Table 5: Strict and Fuzzy RDD: Literacy

	Strict RDD fixed bandwidth		**Strict RDD optimal bandwidth**		**Fuzzy RDD fixed bandwidth**		**Fuzzy RDD optimal bandwidth**	
	Pre-reform	**Post reform**	**Pre-reform**	**Post reform**	**Pre-reform**	**Post reform**	**Pre-reform**	**Post reform**
First-Stage Grade 3					0.436	0.855	0.441	0.853
z-value (robust)					7.749	37.923	15.961	85.802
Robust 95% CI					[0.316; 0.530]	[0.806; 0.894]	[0.387; 0.495]	[0.831; 0.870]
Grade 3	0.103**	0.315***	0.101*	0.311***	0.235**	0.368***	0.230*	0.364***
Robust 95% CI	[0.022; 0.261]	[0.259; 0.333]	[-0.007; 0.185]	[0.278; 0.333]	[0.063; 0.602]	[0.294; 0.402]	[-0.014; 0.418]	[0.324; 0.394]
Eff. N of obs	2,564/2,790	15,452/16,216	3,781/4,059	26,775/27,389	2,564/2,790	15,452/16,216	3,781/4,059	26,775/27,389
N of waves	1	5	1	5	1	5	1	5
Bandwidth	30	30	43	50	30	30	43	50
First-Stage Grade 5					0.571	0.829	0.561	0.842
z-value (robust)					47.138	44.686	53.893	81.367
Robust 95% CI					[0.546; 0.594]	[0.770; 0.840]	[0.542; 0.583]	[0.817; 0.857]
Grade 5	0.144**	0.277***	0.150***	0.241***	0.252**	0.334***	0.268***	0.286***
Robust 95% CI	[0.004; 0.172]	[0.193; 0.351]	[0.085; 0.210]	[0.198; 0.285]	[0.006; 0.302]	[0.240; 0.436]	[0.148; 0.376]	[0.235; 0.342]
Eff. N of obs	7,880/8,684	8,866/9,220	11,521/12,613	22,655/22,823	7,880/8,684	8,866/9,220	11,521/12,613	22,655/22,823
N of waves	3	4	3	4	3	4	3	4
Bandwidth	30	30	43	73	30	30	43	73
First-Stage Grade 7					0.642	0.808	0.644	0.815
z-value (robust)					32.117	30.415	49.919	49.272
Robust 95% CI					[0.595; 0.673]	[0.735; 0.836]	[0.616; 0.667]	[0.784; 0.849]
Grade 7	0.062	0.191**	0.065*	0.1910**	0.097	0.237**	0.101*	0.208***
Robust 95% CI	[-0.053; 0.107]	[0.024; 0.383]	[-0.002; 0.122]	[0.066; 0.289]	[-0.085; 0.171]	[0.045; 0.472]	[-0.004; 0.191]	[0.084; 0.351]
Eff. N of obs	13,090/14,124	2,482/2,519	20,022/21,299	4,876/4,980	13,090/14,124	2,482/2,519	20,022/21,299	4,876/4,980
N of waves	5	2	5	2	5	2	5	2
Bandwidth	30	30	44	59	30	30	44	59

Table 6: Numeracy Robustness check

	Strict RDD		Fuzzy RDD	
	Pre-reform	**Post reform**	**Pre-reform**	**Post reform**
(i) Aggregation				
2 days aggregated				
First-Stage Grade 3			0.434	0.853
z-value (robust)			11.01	37.404
Robust 95% CI			[0.350; 0.502]	[0.806; 0.895]
Grade 3	0.019	0.372***	0.045	0.435***
Robust 95% CI	[-0.062; 0.164]	[0.301; 0.356]	[-0.137; 0.372]	[0.345; 0.427]
Eff. N of obs	2,514/ 2,690	14,953/15,730	2,514/2,690	14,953/15,730
N of waves	1	5	1	5
Window width =15				
First-Stage Grade 5			0.571	0.826
z-value (robust)			32.739	35.46
Robust 95% CI			[0.535; 0.603]	[0.756; 0.844]
Grade 5	0.156***	0.297***	0.274**	0.359***
Robust 95% CI	[0.019; 0.140]	[0.196; 0.358]	[0.032; 0.248]	[0.233; 0.460]
Eff. N of obs	7,648/8,444	8,544/8,908	7,648/8,444	8,544/8,908
N of waves	3	4	3	4
Window width =15				
First-Stage Grade 7			0.641	0.8
z-value (robust)			22.358	47.104
Robust 95% CI			[0.578; 0.689]	[0.740; 0.804]
Grade 7	0.05	0.222***	0.078	0.278***
Robust 95% CI	[-0.112; 0.099]	[0.066; 0.378]	[-0.178; 0.161]	[0.091; 0.483]
Eff. N of obs	12,697/13,736	2,430/2,442	12,697/13,736	2,430/2,442
N of waves	5	2	5	2
Window width =15				
7 days aggregated				
First-Stage Grade 3			0.422	0.847
z-value (robust)			42.117	49.317
Robust 95% CI			[0.391; 0.429]	[0.817; 0.918]
Grade 3	-0.003	0.381***	-0.008	0.445***
Robust 95% CI	[-0.088; 0.243]	[0.295; 0.340]	[-0.208; 0.574]	[0.334; 0.396]
Eff. N of obs	2,514/2,690	14,953/15,730	2,514/2,690	14,953/15,730
N of waves	1	5	1	5
Window width =15				
First-Stage Grade 5			0.556	0.832
z-value (robust)			44.852	53.372
Robust 95% CI			[0.513; 0.560]	[0.767; 0.827]
Grade 5	0.156	0.316***	0.281*	0.379***
Robust 95% CI	[-0.013; 0.141]	[0.182; 0.385]	[-0.022; 0.272]	[0.225; 0.500]
Eff. N of obs	7,648/8,444	8,544/8,908	7,648/8,444	8,544/8,908
N of waves	3	4	3	4
Window width =15				

Table 6. Continued

	STRICT - Compliers and non-compliers		FUZZY - Compliers and non-compliers	
	Pre-reform	**Post reform**	**Pre-reform**	**Post reform**
7 days aggregated (continued)				
First-Stage Grade 7			0.644	0.805
z-value (robust)			52.651	125.303
Robust 95% CI			[0.610; 0.657]	[0.775; 0.800]
Grade 7	0.035	0.257***	0.054	0.320***
Robust 95% CI	[-0.191; 0.060]	[0.222; 0.391]	[-0.298; 0.096]	[0.287; 0.488]
Eff. N of obs	12,697/13,736	2,430/2,442	12,697/13,736	2,430/2,442
N of waves	5	2	5	2
Window width =15				
(ii) Non-symmetric Deletion of youngest to Left of cut off (1 day)				
First-Stage Grade 3			0.46	0.861
z-value (robust)			19.087	54.106
Robust 95% CI			[0.428; 0.526]	[0.835; 0.898]
Grade 3	0.041*	0.377***	0.089*	0.438***
Robust 95% CI	[-0.001; 0.190]	[0.316; 0.375]	[-0.06; 0.411]	[0.357; 0.439]
Eff. N of obs	2,520/2,791	14,951/16,241	2,520/2,791	14,951/16,241
N of waves	1	5	1	5
Window width =30				
First-Stage Grade 5			0.572	0.839
z-value (robust)			36.697	40.654
Robust 95% CI			[0.542; 0.603]	[0.783; 0.862]
Grade 5	0.153	0.295***	0.268	0.351***
Robust 95% CI	[-0.038; 0.190]	[0.175; 0.378]	[-0.067; 0.332]	[0.213; 0.460]
Eff. N of obs	7,605/8,684	8,554/9,232	7,605/8,684	8,554/9,232
N of waves	3	4	3	4
Window width =30				
First-Stage Grade 7			0.643	0.819
z-value (robust)			28.98	20.479
Robust 95% CI			[0.599; 0.686]	[0.725; 0.879]
Grade 7	0.045	0.241*	0.07	0.294**
Robust 95% CI	[-0.112; 0.075]	[-0.006; 0.549]	[-0.176; 0.119]	[0.013; 0.61]
Eff. N of obs	12,631/14,176	2,388/2,552	12,631/14,176	2,388/2,552
N of waves	5	2	5	2
Window width =30				
(ii) Non-symmetric Deletion of youngest (7 days) to Left of cut off				
First-Stage Grade 3			0.457	0.845
z-value (robust)			5.745	22.814
Robust 95% CI			[0.324; 0.660]	[0.774; 0.919]
Grade 3	-0.047	0.384***	-0.102	0.455***
Robust 95% CI	[-0.652; 0.505]	[0.182; 0.532]	[-1.402; 1.095]	[0.225; 0.619]
Eff. N of obs	2,013/2,791	11,734/16,241	2,013/2,791	11,734/16,241
N of waves	1	5	1	5
Window width =5				

Table 6. Continued

	STRICT - Compliers and non-compliers		FUZZY - Compliers and non-compliers	
	Pre-reform	**Post reform**	**Pre-reform**	**Post reform**
(ii) Non-symmetric Deletion of youngest (7 days) to Left of cut off (continued)				
First-Stage Grade 5			0.545	0.828
z-value (robust)			12.447	11.632
Robust 95% CI			[0.435; 0.598]	[0.637; 0.896]
Grade 5	0.216***	0.272***	0.397***	0.328***
Robust 95% CI	[0.114; 0.373]	[0.112; 0.329]	[0.206; 0.730]	[0.141; 0.440]
Eff. N of obs	6,130/8,684	6,751/9,232	6,130/8,684	6,751/9,232
N of waves	3	4	3	4
Window width =5				
First-Stage Grade 7			0.598	0.827
z-value (robust)			22.966	9.005
Robust 95% CI			[0.498; 0.591]	[0.645; 1.004]
Grade 7	0.096	0.232*	0.16	0.281**
Robust 95% CI	[-0.035; 0.250]	[-0.013; 0.551]	[-0.047; 0.434]	[0.010; 0.642]
Eff. N of obs	10,210/14,176	1,876/2,552	10,210/14,176	1,876/2,552
N of waves	5	2	5	2
Window width =5				
(ii) Non-symmetric Deletion of youngest (14 days) to Left of cut off				
First-Stage Grade 3			0.41	0.847
z-value (robust)			58.894	49.289
Robust 95% CI			[0.415; 0.444]	[0.847; 0.917]
Grade 3	0.017***	0.379***	0.042***	0.447***
Robust 95% CI	[-0.310; -0.082]	[0.347; 0.476]	[-0.756; -0.203]	[0.373; 0.562]
Eff. N of obs	3,799/4,058	22,530/23,203	3,799/4,058	22,530/23,203
N of waves	1	5	1	5
Window width =15				
First-Stage Grade 5			0.524	0.83
z-value (robust)			227.834	58.545
Robust 95% CI			[0.586; 0.596]	[0.780; 0.834]
Grade 5	0.198***	0.296***	0.377***	0.356***
Robust 95% CI	[0.093; 0.230]	[0.407; 0.419]	[0.134; 0.387]	[0.490; 0.527]
Eff. N of obs	11,548/12,621	12,874/13,184	11,548/12,621	12,874/13,184
N of waves	3	4	3	4
Window width =15				
First-Stage Grade 7			0.628	0.805
z-value (robust)			68.954	1100.00
Robust 95% CI			[0.634; 0.671]	[0.807; 0.810]
Grade 7	0.05	0.250***	0.08	0.310***
Robust 95% CI	[-0.139; 0.058]	[0.335; 0.344]	[-0.228; 0.092]	[0.415; 0.425]
Eff. N of obs	19,145/20,458	3,625/3,619	19,145/20,458	3,625/3,619
N of waves	5	2	5	2
Window width =15				

Table 6. Continued

	STRICT - Compliers and non-compliers		FUZZY - Compliers and non-compliers	
	Pre-reform	**Post reform**	**Pre-reform**	**Post reform**
(iii) Symmetric Deletion of youngest to Left of cut off (1 day)				
First-Stage Grade 3			0.461	0.867
z-value (robust)			19.089	50.979
Robust 95% CI			[0.430; 0.529]	[0.844; 0.911]
Grade 3	0.038	0.381***	0.083	0.440***
Robust 95% CI	[-0.022; 0.198]	[0.313; 0.383]	[-0.050; 0.426]	[0.345; 0.447]
Eff. N of obs	2,520/2,707	14,951/15,696	2,520/2,707	14,951/15,696
N of waves	1	5	1	5
First-Stage Grade 5			0.575	0.841
z-value (robust)			35.408	38.015
Robust 95% CI			[0.545; 0.609]	[0.782; 0.867]
Grade 5	0.149***	0.300***	0.259	0.356***
Robust 95% CI	[-0.055; 0.172]	[0.170; 0.385]	[-0.099; 0.298]	[0.205; 0.469]
Eff. N of obs	7,605/8,447	8,554/8,902	7,605/8,447	8,554/8,902
N of waves	3	4	3	4
First-Stage Grade 7			0.65	0.819
z-value (robust)			27.462	20.454
Robust 95% CI			[0.604; 0.697]	[0.725; 0.879]
Grade 7	0.046	0.276**	0.07	0.337**
Robust 95% CI	[-0.133; 0.082]	[0.074; 0.597]	[-0.208; 0.129]	[0.116; 0.718]
Eff. N of obs	12,631/13,740	2,388/2,449	12,631/13,740	2,388/2,449
N of waves	5	2	5	2
(iii) Symmetric Deletion of youngest to Left of cut off (1 for 7 days)				
First-Stage Grade 3			0.464	0.882
z-value (robust)			5.872	26.127
Robust 95% CI			[0.335; 0.670]	[0.839; 0.976]
Grade 3	-0.046	0.398***	-0.1	0.451***
Robust 95% CI	[-0.646; 0.513]	[0.178; 0.540]	[-1.364; 1.094]	[0.196; 0.592]
Eff. N of obs	2,013/2,132	11,734/12,332	2,013/2,132	11,734/12,332
N of waves	1	5	1	5
First-Stage Grade 5			0.59	0.821
z-value (robust)			13.518	11.984
Robust 95% CI			[0.486; 0.650]	[0.627; 0.873]
Grade 5	0.213***	0.328***	0.360***	0.399***
Robust 95% CI	[0.110; 0.331]	[0.169; 0.383]	[0.182; 0.593]	[0.221; 0.520]
Eff. N of obs	6,130/6,639	6,751/6,968	6,130/6,639	6,751/6,968
N of waves	3	4	3	4
First-Stage Grade 7			0.648	0.828
z-value (robust)			26.004	8.935
Robust 95% CI			[0.559; 0.650]	[0.644; 1.006]
Grade 7	0.034	0.262***	0.052	0.317***
Robust 95% CI	[-0.101; 0.139]	[0.091; 0.607]	[-0.153; 0.219]	[0.142; 0.703]
Eff. N of obs	10,210/10,822	1,876/1,980	10,210/10,822	1,876/1,980
N of waves	5	2	5	2

Table 6. Continued

	STRICT - Compliers and non-compliers		FUZZY - Compliers and non-compliers	
	Pre-reform	**Post reform**	**Pre-reform**	**Post reform**
(iii) Symmetric Deletion of youngest to Left of cut off (1 for 14 days)				
First-Stage Grade 3			0.474	0.896
z-value (robust)			1.076	4.129
Robust 95% CI			[-0.375; 1.289]	[0.404; 1.134]
Grade 3	0.214	0.384***	0.452	0.429
Robust 95% CI	[-0.901; 1.420]	[-0.115; 0.943]	[-2.025; 3.151]	[-0.173; 1.219]
Eff. N of obs	1,374/1,454	7,928/8,440	1,374/1,454	7,928/8,440
N of waves	1	5	1	5
First-Stage Grade 5			0.631	0.822
z-value (robust)			2.706	2.993
Robust 95% CI			[0.166; 1.036]	[0.210; 1.007]
Grade 5	0.131**	0.380**	0.208*	0.462***
Robust 95% CI	[0.022; 0.868]	[0.048; 1.453]	[-0.034; 1.464]	[0.249; 1.818]
Eff. N of obs	4,199/4,612	4,558/4,814	4,199/4,612	4,558/4,814
N of waves	3	4	3	4
First-Stage Grade 7			0.669	0.827
z-value (robust)			2.501	2.284
Robust 95% CI			[0.076; 0.627]	[0.091; 1.196]
Grade 7	-0.039	0.290**	-0.058	0.351***
Robust 95% CI	[-0.368; 0.620]	[0.154; 1.805]	[-0.583; 0.906]	[0.336; 2.189]
Eff. N of obs	6,869/7,561	1,287/1,401	6,869/7,561	1,287/1,401
N of waves	5	2	5	2

Table 7: Numeracy by gender

	Strict RDD		Fuzzy RDD	
	Pre-reform	**Post reform**	**Pre-reform**	**Post reform**
Male				
First-Stage Grade 3			0.342	0.823
z-value (robust)			7.125	28.373
Robust 95% CI			[0.240; 0.422]	[0.755; 0.867]
Grade 3	0.061	0.334***	0.177	0.406***
Robust 95% CI	[0.226; 0.290]	[0.242; 0.357]	[-0.627; 0.827]	[0.305; 0.435]
Eff. N of obs	1,280/1,417	7,968/8,321	1,280/1,417	7,968/8,321
N of waves	1	5	1	5
First-Stage Grade 5			0.496	0.786
z-value (robust)			32.932	25.268
Robust 95% CI			[0.462; 0.521]	[0.695; 0.812]
Grade 5	0.073	0.215***	0.148	0.273***
Robust 95% CI	[-0.154; 0.183]	[0.062; 0.291]	[-0.312; 0.372]	[0.080; 0.393]
Eff. N of obs	4,024/4,382	4,555/4,655	4,024/4,382	4,555/4,655
N of waves	3	4	3	4
First-Stage Grade 7			0.575	0.784
z-value (robust)			31.433	15.742
Robust 95% CI			[0.528; 0.598]	[0.660; 0.848]
Grade 7	-0.027	0.155***	-0.047	0.198
Robust 95% CI	[-0.201; 0.035]	[-0.109; 0.469]	[-0.348; 0.058]	[-0.124; 0.598]
Eff. N of obs	6,636/7,071	1,268/1,284	6,636/7,071	1,268/1,284
N of waves	5	2	5	2
Female				
First-Stage Grade 3			0.527	0.889
z-value (robust)			6.387	50.473
Robust 95% CI			[0.350; 0.659]	[0.856; 0.925]
Grade 3	-0.011	0.407***	-0.02	0.459***
Robust 95% CI	[-0.078; 0.185]	[0.300; 0.447]	[-0.147; 0.347]	[0.330; 0.508]
Eff. N of obs	1,293/1,374	7,523/7,920	1,293/1,374	7,523/7,920
N of waves	1	5	1	5
First-Stage Grade 5			0.643	0.873
z-value (robust)			42.705	58.17
Robust 95% CI			[0.612; 0.671]	[0.827; 0.885]
Grade 5	0.232***	0.373***	0.361***	0.428***
Robust 95% CI	[0.074; 0.280]	[0.273; 0.463]	[0.115; 0.437]	[0.326; 0.534]
Eff. N of obs	3,867/4,302	4,321/4,577	3,867/4,302	4,321/4,577
N of waves	3	4	3	4
First-Stage Grade 7			0.708	0.82
z-value (robust)			30.746	46.99
Robust 95% CI			[0.658; 0.747]	[0.761; 0.828]
Grade 7	0.123	0.254*	0.174	0.310**
Robust 95% CI	[-0.017; 0.185]	[-0.002; 0.408]	[-0.027; 0.267]	[0.008; 0.506]
Eff. N of obs	6,493/7,105	1,236/1,268	6,493/7,105	1,236/1,268
N of waves	5	2	5	2

Table 8: KLA by subjects

Subject		Strict - RDD		Fuzzy - RDD	
		Pre-reform	Post reform	Pre-reform	Post reform
Math achieve	First-Stage Grade 5			0.495	0.831
	z-value			22.675	37.828
	Robust 95% CI			[0.452; 0.537]	[0.765; 0.849]
	Grade 5	0.122	0.362***	0.247	0.436***
	Robust 95% CI	[-0.043; 0.248]	[0.283; 0.432]	[-0.100; 0.516]	[0.350; 0.536]
	Eff. N of obs	5,238/5,697	8,736/9,124	5,238/5,697	8,736/9,124
	N of waves	3	4	3	4
Math effort	First-Stage Grade 5			0.496	0.831
	z-value			22.51	37.743
	Robust 95% CI			[0.452; 0.538]	[0.765; 0.849]
	Grade 5	0.071	0.173***	0.143	0.208***
	Robust 95% CI	[-0.025; 0.105]	[0.124; 0.237]	[-0.054; 0.215]	[0.148; 0.298]
	Eff. N of obs	5,235/5,694	8,738/9,129	5,235/5,694	8,738/9,129
	N of waves	3	4	3	4
English achieve	First-Stage Grade 5			0.494	0.833
	z-value			23.419	38.122
	Robust 95% CI			[0.452; 0.534]	[0.768; 0.851]
	Grade 5	0.143**	0.328***	0.290**	0.394***
	Robust 95% CI	[0.012; 0.220]	[0.232; 0.388]	[0.011; 0.460]	[0.286; 0.481]
	Eff. N of obs	5,225/5,689	8,722/9,114	5,225/5,689	8,722/9,114
	N of waves	3	4	3	4
English effort	First-Stage Grade 5			0.495	0.832
	z-value			23.236	37.145
	Robust 95% CI			[0.452; 0.535]	[0.766; 0.851]
	Grade 5	0.082	0.158***	0.166	0.190***
	Robust 95% CI	[-0.030; 0.105]	[0.100; 0.234]	[-0.064; 0.216]	[0.119; 0.293]
	Eff. N of obs	5,222/5,685	8,725/9,121	5,222/5,685	8,725/9,121
	N of waves	3	4	3	4

Table 8. Continued

Subject		Strict - RDD		Fuzzy - RDD	
		Pre-reform	Post reform	Pre-reform	Post reform
Science achieve	First-Stage Grade 5			0.494	0.83
	z-value			23.231	42.835
	Robust 95% CI			[0.453; 0.537]	[0.769; 0.842]
	Grade 5	0.101	0.28***	0.203	0.338***
	Robust 95% CI	[-0.036; 0.193]	[0.197; 0.347]	[-0.076; 0.393]	[0.245; 0.431]
	Eff. N of obs	5,237/5,695	8,808/9,184	5,237/5,695	8,808/9,184
	N of waves	3	4	3	4
Science effort	First-Stage Grade 5			0.495	0.83
	z-value			23.136	41.826
	Robust 95% CI			[0.454; 0.538]	[0.766; 0.842]
	Grade 5	0.092*	0.169***	0.186*	0.204***
	Robust 95% CI	[-0.012; 0.214]	[0.109; 0.219]	[-0.035; 0.442]	[0.134; 0.274]
	Eff. N of obs	5,245/5,704	8,813/9,186	5,245/5,704	8,813/9,186
	N of waves	3	4	3	4
Math achieve	First-Stage Grade 7			0.604	0.801
	z-value			26.683	29.909
	Robust 95% CI			[0.550; 0.637]	[0.725; 0.826]
	Grade 7	0.079	0.248***	0.13	0.310***
	Robust 95% CI	[-0.080; 0.146]	[0.104; 0.379]	[-0.135; 0.249]	[0.150; 0.473]
	Eff. N of obs	10,278/11,315	2,457/2,514	10,278/11,315	2,457/2,514
	N of waves	5	2	5	2
Math effort	First-Stage Grade 7			0.604	0.802
	z-value			26.499	29.953
	Robust 95% CI			[0.549; 0.637]	[0.724; 0.826]
	Grade 7	-0.027	0.101	-0.02	0.126
	Robust 95% CI	[-0.152; 0.047]	[-0.041; 0.241]	[-0.154; 0.018]	[-0.041; 0.298]
	Eff. N of obs	5,106/11,298	2,459/2,519	10,281/11,298	2,459/2,519
	N of waves	5	2	5	2
English achieve	First-Stage Grade 7			0.604	0.802
	z-value			26.78	32.158
	Robust 95% CI			[0.550; 0.637]	[0.727; 0.821]
	Grade 7	0.087**	0.240***	0.144**	0.300***
	Robust 95% CI	[0.002; 0.163]	[0.090; 0.458]	[0.001; 0.277]	[0.132; 0.572]
	Eff. N of obs	10,273/11,326	2,455/2,513	10,273/11,326	2,455/2,513
	N of waves	5	2	5	2

Table 8. Continued

Subject		Strict - RDD		Fuzzy - RDD	
		Pre-reform	**Post reform**	**Pre-reform**	**Post reform**
English effort	First-Stage Grade 7			0.604	0.802
	z-value			26.449	32.523
	Robust 95% CI			[0.549; 0.637]	[0.728; 0.822]
	Grade 7	0.026	0.147*	0.044	0.183**
	Robust 95% CI	[-0.012; 0.094]	[-0.006; 0.328]	[-0.019; 0.156]	[0.005; 0.410]
	Eff. N of obs	10,268/11,301	2,459/2,518	10,268/11,301	2,459/2,518
	N of waves	5	2	5	2
Science achieve	First-Stage Grade 7			0.605	0.8
	z-value			26.61	29.412
	Robust 95% CI			[0.551; 0.639]	[0.724; 0.827]
	Grade 7	0.048	0.263***	0.08	0.329***
	Robust 95% CI	[-0.073; 0.133]	[0.124; 0.511]	[-.0122; 0.224]	[0.179; 0.634]
	Eff. N of obs	10,307/11,320	2,473/2,519	10,307/11,320	2,473/2,519
	N of waves	5	2	5	2
Science effort	First-Stage Grade 7			0.605	0.801
	z-value			26.426	29.954
	Robust 95% CI			[0.551; 0.639]	[0.726; 0.827]
	Grade 7	0.032	0.111	0.053	0.139
	Robust 95% CI	[-0.072; 0.125]	[-0.039; 0.328]	[-0.119; 0.209]	[-0.038; 0.408]
	Eff. N of obs	10,316/11,326	2,476/2,521	10,316/11,326	2,476/2,521
	N of waves	5	2	5	2