AFEP | Africa Fellows in Education Program

GLOBAL EDUCATION ANALYTICS INSTITUTE
Putting evidence at the heart of education policy

AFEP WORKING PAPER

Does Class Size Matter for Learning? For Whom and How? Evidence from Public Primary Schools in Rwanda

Nneka Esther Osadolor

University of Benin, Nigeria

April 2026

geai.io

A GEAI program, implemented in partnership with pep partnership for economic policy

Acknowledgements

This work was carried out with financial and scientific support from the Partnership for Economic Policy (PEP) and the Global Education Analytics Institute (GEAI) through funding provided by the Yidan Foundation. I am deeply grateful to Anthony Wambugu and Daniel Rodriguez-Segura for their mentorship and intellectual guidance, and to Ramaele Moshoeshoe for his leadership and support. I gratefully acknowledge NewGlobe for access to the RwandaEQUIP data. I also thank the fellows and mentors who gave feedback during the AFEP brown-bag sessions, and participants at the PEP Annual Conference (Nairobi, 9–13 June 2025) and the PEP Foundational Learning Symposium (19 August 2025).

Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Partnership for Economic Policy (PEP) or the Global Education Analytics Institute (GEAI).

Does Class Size Matter for Learning? For Whom and How? Evidence from Public Primary Schools in Rwanda

Nneka Esther Osadolor*

University of Benin, Nigeria

April 2026

Abstract

Developing countries have achieved substantial gains in school enrolment, yet improvements in learning outcomes have lagged behind. Overcrowded classrooms are a key barrier, as they tend to compromise the conditions for quality and effective instruction. Understanding the role of class size - one of the most debated and policy-relevant inputs in education - is therefore critical, particularly in resource-constrained systems where reductions are costly. This paper estimates the causal effect of class size on student learning outcomes using administrative data from the RwandaEQUIP programme across 761 public primary schools in Rwanda. Exploiting enrolment-based class size thresholds, we implement a fuzzy regression discontinuity design to identify exogenous variation in class size. The results show that reducing class size by approximately 16 pupils per stream - moving from roughly 52 to 36 pupils - improves short-run learning outcomes by 0.23 standard deviations (SD) in English, 0.13 SD in Kinyarwanda, 0.22 SD in Science, and 0.20 SD in Social Studies. Mathematics gains are negligible in the short run but grow to 0.06 SD by year-end, consistent with the cumulative nature of numeracy skill acquisition. Effects are largest in early primary grades and single-shift schools, and dissipates by Term 3, when estimates remain positive but statistically insignificant across all subjects. The findings highlight the importance of targeting class size reductions to early grades, where marginal returns are highest, and contribute to the limited causal evidence on class size and learning in Sub-Saharan Africa.

Keywords: class size; learning outcomes; regression discontinuity; Sub-Saharan Africa

* University of Benin, Nigeria; Africa Fellows in Education Program (AFEP). Email: nneka.okoh@uniben.edu

1 Introduction

Globally, especially among developing countries, there has been remarkable progress towards the reduction in the number of out-of-school children, with significant improvements in school enrolments (World Bank, 2022). However, learning outcomes have remained poor (learning poverty – defined as being unable to read and understand, write or perform basic numeracy activities by age 10), more so among students of government owned schools. According to the World Bank (2022), prior to the Covid-19 pandemic, the learning poverty rate was estimated to be about 57 percent in Low- and Middle-Income countries and 86 percent in Sub-Saharan Africa (SSA). This has further been exacerbated by the advent of the Covid-19 pandemic following school closures with the learning poverty rate in Low- and Middle-Income countries surging to 70 percent (Herbert, Saavedra, Marr and Jenkins, 2021; World Bank, 2022). These trends have spurred increased attention towards identifying strategies to improve learning outcomes especially in resource-constrained educational environments.

Among the various inputs into education systems, class size remains one of the most debated and policy-relevant factors potentially influencing student achievement. Although, class size reduction is often viewed as a means of enhancing the quality of instruction by enabling more individualized attention and classroom management (Altinok & Kingdon, 2012; Krueger, 2003; Hanushek, 1999), empirical evidence on its effectiveness remains mixed, particularly in LMICs. While some studies have found that smaller classes improve student outcomes (Moshoeshoe, 2015; Mueller, 2013; Altinok and Kingdon, 2012; Asadullah, 2005; Jepsen and Rivkin, 2009; Krueger, 1999), others report negative, negligible or heterogeneous effects depending on context, grade level, or school quality (Datta and Kingdon, 2023; K hler, 2022; Urquiola and Verhoogen, 2009; Hoxby, 2000; Hanushek, 1986). Besides the role of the study context (country, institutional setting, time frame, etc.) in the class size – learning outcomes relationship (Moshoeshoe, 2015), there are several channels through which class size might improve learning outcomes. As highlighted by Crawfurd and Le Nestour (2022), the starting size of the class matters for the effect of class size on learning. Specifically, a smaller class might improve learning outcomes if it allows teachers to spend significantly more time on individualized support to the pupils. For instance, reducing a class from 80 to 40 pupils may not make much difference with respect to learning improvements, as it may be difficult for the teacher to engage the pupils meaningfully in individualized instruction (Crawfurd and Le Nestour, 2022). However, if the overcrowding is so bad that kids literally cannot see the board, hear the teacher, or focus. Through this channel, the reduction in class size might be effective, not because of more personalized instruction, but because kids get more from existing instruction.

In Rwanda, the government has made concerted efforts to improve the quality and equity of education through a range of interventions—including curriculum reform, teacher professional

development, and digital learning innovations (Rodriguez-Segura et al., 2023). Nonetheless, the challenge of large class sizes remains acute. For instance, despite the addition of 2,222 classrooms between the 2022/23 and 2023/24 school years, the national pupil–classroom ratio (PCR) increased from 50:1 to 52:1, with ratios in primary schools averaging around 60:1 (Ministry of Education (MINEDUC), 2025), indicating that the growth in inputs was merely keeping up with population growth. Similarly, the pupil–teacher ratio rose from 42:1 to 48:1 over the same period, following a 6% increase in public primary school enrolment and 25,293 primary school teachers in 2023/24. In addition to the increasing enrolment pressure, learning outcomes remained poor despite the significant progress in enrolment. Existing data reveals that even with reforms like the Competence-Based Curriculum (CBC) in 2019, 80% of students in lower primary fell below the required reading levels, with only 16% of students meeting the early grade reading benchmarks (UNESCO, 2022). In light of these, in 2022, the government launched the RwandaEQUIP programme, targeting public and government-subsidized schools, to drive improvements in foundational literacy and numeracy. With about 761 schools covered by the programme in 2024, RwandaEQUIP offers standardized delivery of instruction using scripted lesson plans and teacher guides aligned with the national curriculum (Rodriguez-Segura et al., 2023).

This study leverages administrative data from the RwandaEQUIP programme—spanning multiple grades, subjects, and schools—and employs a fuzzy regression discontinuity design (RDD) to estimate the causal effect of class size on student academic performance. The design exploits discontinuities in class size induced by the enforcement of a maximum class size rule, generating exogenous variation suitable for estimating treatment effects. This strategy is informed by Urquiola and Verhoogen (2009), who demonstrated how class-size caps in Chile generated sharp enrolment cut-offs that could be exploited using RDD. However, unlike their study, where sorting of students and parental school choice compromised the validity of the RDD (Kedagni, et al., 2021), these concerns are minimized in the RwandaEQUIP context. Specifically, students attend schools by catchment area, and streaming is based on administrative constraints rather than academic ability. This reduces selection biases (limits parent and student sorting), ensuring that variation in class size is exogenous. This institutional setting substantially limits, though may not entirely eliminate, scope for sorting on unobservables around the class-size threshold. In particular, catchment-based school assignment and administrative class splitting imply that enrolment-driven variation in streams is plausibly exogenous with respect to pupil ability and household characteristics, conditional on observed controls. Nevertheless, following the discrete nature of the running variable with few mass points from our data, we adopt an explicit extrapolation-based IV approach as emphasized by Cattaneo et al. (2024). Based on this, identification does not rely on local randomization at the cutoff, but rather on the assumption that the conditional relationship between enrollment-induced class structure and learning

outcomes is well approximated by a linear function in a neighbourhood covering the available support of the running variable.

This paper contributes to the literature of class size reduction in several ways. First, this paper contributes to the class size literature by providing causal evidence from a setting characterized by substantially higher starting pupil–teacher ratios and more severely overcrowded classrooms than most prior studies. The existing literature spans a range of country contexts — including the United States (Krueger, 1999), Chile (Urquiola and Verhoogen, 2009), India (Datta and Kingdon, 2023), Bangladesh (Asadullah, 2005), and Lesotho (Moshoeshoe, 2015) — yet the majority of these settings involve starting class sizes that are considerably smaller than those observed in Rwanda, where the national pupil–classroom ratio stands at 52:1 and averages above 60:1 in public primary schools. This is important for at least two reasons. Theoretically, the relationship between class size and learning is unlikely to be linear: the marginal benefit of reducing a class from 60 to 44 pupils may differ substantially from reducing one from 30 to 25, particularly through the channel of individualized attention. Empirically, estimates from lower-ratio settings may therefore provide limited guidance for policy in systems where overcrowding is more severe. To put the magnitude of the intervention studied here in perspective, the class size reduction we identify - approximately 16 fewer pupils per stream, moving from roughly 52 to 36 pupils - is considerably larger in absolute terms than reductions studied in, for example, Angrist and Lavy (1999), where the Maimonides rule generates average class size differences of around 8 pupils, or the STAR experiment (Krueger, 1999), where class sizes were reduced from approximately 22 to 15 pupils. The policy-relevant question in Rwanda is therefore whether class size reductions from a high base — achieved at significant fiscal cost — generate learning gains commensurate with that investment. Our findings answers this question directly, showing that a reduction of this magnitude yields gains of 0.23 SD in English and 0.22 SD in Science in the short run, with effects concentrated in early grades and single-shift schools. Rwanda's combination of high starting ratios, a double-shift system, and rapid enrolment growth makes it a distinctive and policy-relevant context, and one that is more representative of the conditions facing many education systems across Sub-Saharan Africa than the settings that dominate the existing literature.

Second, this study provides a more nuanced analysis by assessing how other factors, particularly, student and school-level characteristics, moderate the class size – learning outcomes relationship. While much of the existing work estimates average effects of class size, recent studies have called for greater attention to heterogeneity in impacts across contexts and populations (Altinok & Kingdon, 2012; Datta & Kingdon, 2023; Köhler, 2022). For example, Duflo, Dupas, and Kremer (2015) highlight how student ability grouping may shape the efficacy of class size interventions, while Urquiola and Verhoogen (2009) show how sorting around class size thresholds can influence outcomes. Similarly, Jepsen and Rivkin (2009) and Moshoeshoe (2015) find

that class size effects differ across grade levels and school quality. To build on this, this study shows that class size effects vary with grade level, geographic location of the school, and the shifting model operated by the school (that is, whether the school operates a double or single shift model). We do this by leveraging the rich dataset spanning 761 public primary schools in Rwanda, combining foundational literacy and numeracy (FLN) assessment data with end-of-term examination outcomes across the 2023–2024 academic year. The combined datasets allow for a granular analysis of class size effects across subjects, grades, and regions, offering nuanced insights into the conditions under which class size reductions are most effective. The findings aim to inform education policy design in high student-teacher ratio environments.

Third, this study is among the few studies, which assess how the starting size of a class mediates the effects of class size reduction on learning outcomes. Specifically, we go beyond the question of simply whether class sizes should be reduced, to asking how small they need to be for reductions to translate into meaningful learning gains**.** This is particularly important to ensure that class sizes do not go beyond the maximum threshold necessary to maximize the gains in learning outcomes, especially in resource-constrained settings like Rwanda. Previous studies reveal the existence of nonlinearities in the relationship between class size reductions and learning outcomes ((Crawfurd and Le Nestour, 2022; Datta and Kingdon, 2023). We show that the benefits of reducing class stream sizes are concentrated among moderately sized streams and provide support for a policy threshold of approximately 40 to 45 pupils per stream, hence serving as a reasonable policy benchmark. Our results show that there are diminishing marginal returns to further reductions in stream size once streams fall below approximately 40 pupils. In particular, our estimates for different effective stream sizes speak directly to a key policy question: at what point do further reductions in class size yield diminishing returns? The pattern that gains are concentrated when streams contain roughly 40–45 pupils suggests that, in a system with constrained staffing and infrastructure, setting targets in this range may maximize learning gains per additional teacher, rather than pursuing more ambitious but fiscally costly reductions.

The rest of the paper proceeds as follows. Section 2 presents the context of the study. Section 3 outlines the empirical strategy and data. The main results are presented in section 4 with heterogeneity analyses as well as potential mechanism. Section 5 concludes the paper.

2 Context of the Study

2.1 The Rwanda Educational System

Rwanda's basic education system follows a 3-6-3-3 structure, comprising three years of pre-primary, six years of primary, and six years of secondary schooling. Secondary education is divided into lower secondary (ordinary level, three years) and upper secondary (advanced level,

three years). At the end of primary school, students sit national examinations in five subjects. Although all students advance to lower secondary under the government's Twelve Years Basic Education policy, examination performance determines placement into boarding versus day schools. A second national examination at the end of lower secondary determines entry into upper secondary streams — general education, technical and vocational training, or teacher training colleges — based on student results and preferences (Nsabimana et al., 2024; MINEDUC, 2025).

Rwanda's school system comprises three categories of institutions: public schools, private schools, and government-aided schools — privately managed institutions that receive public funding for teacher salaries and operational costs. Given the historical role of religious organisations in Rwandan education, the majority of private and government-aided schools are faith-based, with Catholic institutions accounting for more than half (Trines, 2019). As of 2024, Rwanda had 1,569 public secondary schools, 2,077 government-aided schools, and 1,340 private schools (MINEDUC, 2025).

Since the introduction of free and compulsory basic education in 2003, Rwanda has significantly expanded access to schooling. Public education is provided at no cost through the end of secondary school. As of the 2023/24 academic year, 35.9% of secondary school students were enrolled in public schools, 57.8% in government-aided institutions, and 6.2% in private schools (MINEDUC, 2025). Despite this expansion, class sizes in public primary schools remain large, with the average pupil-classroom ratio exceeding 60:1. To manage capacity constraints, most schools operate a double-shift system, dividing pupils into morning and afternoon sessions. While this arrangement increases the number of pupils that can be accommodated, it reduces total instructional time per pupil.

2.2 The RwandaEQUIP Programme

In response to persistent poor learning outcomes despite significant investments in school infrastructure and human capital, the Rwandan government launched the Rwanda Education Quality Improvement Programme (RwandaEQUIP) in January 2022 (Rodriguez-Segura et al., 2023). The programme targets public primary schools and pursues three main objectives: improving foundational literacy and numeracy proficiency; increasing the share of Primary 6 pupils achieving combined Division I to III in national examinations; and strengthening instructional quality through teacher training and coaching. RwandaEQUIP was rolled out in three phases. The first phase covered 100 schools across 13 districts in all five provinces, beginning in February 2022. A second phase expanded coverage to 250 schools from September 2022, and a third phase reached a total of 761 schools by 2024 (Table 1).

Figure 1 below shows the geographical spread of the RwandaEQUIP programme as at 2024. The programme operates through five interrelated components. First, it provides teachers with

Table 1: Schedule of RwandaEQUIP Program Rollout

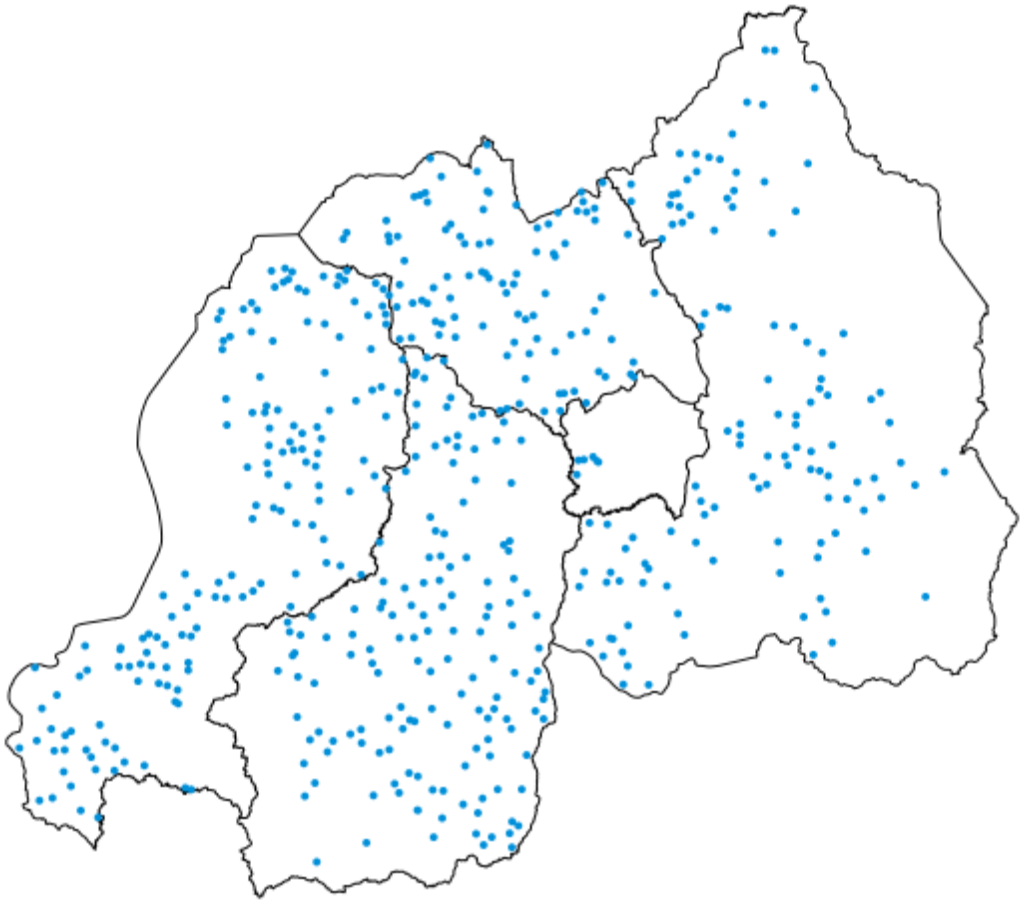
	Feb - June 2022 Phase 1	Sept. - June 2023 Phase 2	Sept. - June 2024 Phase 3
Schools Reached	100	250	761

Notes: **Source**: RwandaEQUIP

structured, curriculum-aligned lesson guides that specify learning objectives, instructional sequences, and formative assessment tasks. Second, each teacher receives a tablet device containing digitised lesson plans organised according to a centrally managed timetable, which is also used to record student progress and generate instructional feedback. Third, the programme includes a ten-day initial training for teachers and head teachers, covering lesson delivery, classroom management, learning assessment, and the use of positive reinforcement. This is supplemented by ongoing school-level coaching from dedicated School Supervisors who conduct weekly visits. Fourth, a school support model integrates operational monitoring — such as teacher deployment and device access — with instructional performance tracking, enabling a rapid-response team to address emerging issues. Fifth, programme data on operational and instructional indicators are compiled in real time and made accessible to school leaders, programme staff, and government stakeholders to support accountability and continuous improvement.

A key structural feature of RwandaEQUIP relevant to this study is its class organisation policy. As part of the programme's strategy to maximise instructional time, schools were encouraged to transition from the double-shift model to a single-shift model, with classes consolidated around a maximum of 69 pupils per stream — approximately the national average at the time of implementation. Where enrolment exceeded this threshold, schools were expected to split classes into additional streams. This policy introduced enrolment-based variation in class structure across schools and grades, forming the basis of the identification strategy employed in this paper.

Figure 1: Aprroximate Locations of phase 3 RwandaEQUIP schools



3 Data and Empirical Strategy

3.1 Data

We leverage multiple datasets collected from 761 Rwandan public primary schools with six grades per school under the RwandaEQUIP programme and a total sample of 4,471 school*grade level observations. This includes data from foundational literacy and numeracy (FLN) assessments (as the baseline measures at the school*grade level), the end-of-term exam results for the 2023–2024 school year (terms 1 and 3) at the pupil level, as well as class information data, which is the source of data for the running variable, at the school*grade level. The data provides information on school characteristics and individual student characteristics including grade level, the phase during which the school joined the RwandaEQUIP programme (i.e. either phase 1, 2 or 3), the region where the school is located. It also provides information on the number of streams, total number of students in each grade, double/Single shift status: whether the grade is under a double shift or single shift status, pupil's gender and age.

Class size is defined as the pupil-to-stream ratio. That is:

$$Class\ size_{sg} = \frac{Number\ of\ pupils_{sg}}{Number\ of\ streams_{sg}}$$

The ratio reflects the average number of pupils per classroom stream, which is equivalent to the number of pupils per teacher in a class.

Measurement of Literacy and Numeracy Exam Scores

At the end of each academic term, students write an end of term exam on multiple subject. The dataset provides assessment score (on a scale from 0-100) on five subjects (Mathematics, English, Kinyarwanda, SRS (Social and Religious Studies) and SET (Science and Elementary Technology) from all three terms.

Pupil literacy level is measured based on established Correct Words Per Minute (CWPM) threshold:

- Level A (< 31 cwpm)
- Level B (>= 31 & < 61 cwpm)
- Level C (>= 61 & < 111 cwpm)
- Level D (>= 111 cwpm)

On the other hand, the numeracy levels are based on an internally designed assessment consisting of 50 items, and the results are mapped onto levels ranging from A-E (10 items for each level, and designed with increasing difficulty across levels). Level A consists of some simplest sub-constructs like number recognition and counting, and level E consists of the hardest sub-constructs in primary education, e.g., signed numbers and probability. Pupils need to score above the internally defined threshold to be on the next level, the thresholds for passing are 60/70/70/70 for level A/B/C/D respectively (highest level is level E, hence, pupils cannot technically ”pass” level E). For example, pupils need to score at or above 60% on level A items (Q1-10) to be at level B, or 70% on level B items (Q11-20) to be leveled at level C.

Summary Statistics

Table 2 reports summary statistics for key student and school characteristics across grade 1 to grade 6 as well as for the outcome variables (exam scores). These include pupil age, gender distribution, average number of pupils per grade, the stream count per grade, among others. We aggregated all covariates and outcomes to the school-grade level. This is to ensure that they match the level at which treatment and the running variable (number of pupils per grade) are defined. In addition, gender, which is a binary variable, was averaged to represent the proportion of female pupils per grade.

Mean scores vary across subjects and grade levels, with the full distribution shown in the table. The mean age of pupils increases progressively with grade level, consistent with Rwanda's

official school-age structure, though slight overage enrolment is evident in the upper grades. For instance, pupils in Primary 1 have an average age of 7.87 years (SD = 0.47), increasing to 12.14 years (SD = 0.59) in Primary 4 and reaching 14.43 years (SD = 0.54) in Primary 6. The proportion of female pupils remains approximately balanced across grades, rising modestly from 46% in Primary 1 to 56% in Primary 6. Average enrolment per grade declines with grade level — from 162.78 pupils in Primary 1 to 65.97 in Primary 6 — likely reflecting dropout and grade repetition patterns common in low- and middle-income education systems.

Pupils are drawn from all five provinces of Rwanda. The largest share of pupils is from the Western Province (27%), followed by the Southern (25%), Eastern (24%), Northern (17%) and Kigali (6%) provinces. This geographic spread ensures that the study captures potential regional heterogeneity in implementation and learning outcomes. Program rollout occurred in three distinct phases. About 13% of the primary 1 pupils are in schools that received the intervention in Phase 1 (2022), 20% in Phase 2 (2023), and the remaining 68% in Phase 3 (2024).

3.2 Empirical Strategy

To estimate the causal effect of class size on student learning outcomes, we exploit the implementation rule of RwandaEQUIP that stipulates a maximum class size of 69 students per stream. According to this rule, schools are to split any class exceeding 69 students into two or more streams, effectively reducing the number of students per stream. While this rule is not strictly enforced, there is evidence (figure 2) of a discrete jump in the probability of having multiple class streams (and average number of streams) at the enrollment threshold of 69 students, which suggests partial compliance with the policy. Specifically, figure 2(a) shows the first stage relationship between number of pupils per grade (enrollment) and the probability of having more than one class stream, while figure 2(b) shows the number of streams in relation to the enrollment threshold. From the figure, there is a sharp increase in the probability of class splitting at the enrollment threshold of 69 pupils. Above the cutoff, the probability of having at least two streams increases by approximately 49 percentage points, indicating strong but imperfect compliance with the class-size rule. The probability continues to rise smoothly with enrollment and reaches nearly one beyond the second threshold. Additionally, figure 2(b) shows that the average number of streams increases from approximately 1.3 to 1.9, implying an average increase of approximately 0.6 additional streams at the cutoff. This (imperfect) discontinuity provides a basis for implementing a fuzzy regression discontinuity design (FRD).

This finding aligns with the implementation design of the RwandaEQUIP programme. Specifically, although Rwanda does not have formal class size threshold, RwandaEQUIP introduced practical class size threshold as a part of its implementation strategy. The goal was to maximize pupils' instructional time by minimizing the number of classes operating under a double-shift system (Rodriguez-Segura et al., 2023). Classes were consolidated to operate

Table 2: Summary Statistics

Variables	N	Primary 1		Primary 2		Primary 3		Primary 4		Primary 5		Primary 6	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
T1 Outcome Variables													
T1 English Score (%)	761	50.93	12.61	61.52	13.27	63.28	13.24	61.61	14.29	55.69	15.99	50.22	16.92
T1 Kinyarwanda Score (%)	761	46.57	12.62	57.17	15.17	66.36	13.41	44.47	12.87	39.23	13.05	50.76	12.45
T1 Mathematics Score (%)	761	42.91	12.61	62.90	13.91	72.53	14.84	76.94	14.96	73.98	18.55	77.73	17.17
T1 SET Score (%)	761	53.96	13.33	50.91	13.76	45.71	13.93	28.79	14.92	26.20	14.64	33.35	12.16
T1 SRS Score (%)	761	55.11	12.91	54.75	13.15	48.53	12.84	33.47	13.23	32.82	12.96	32.90	12.86
T3 Outcome Variables													
T3 English Score (%)	761	39.77	17.23	41.17	17.06	44.95	17.22	31.39	16.24	29.52	14.42	55.52	21.47
T3 Kinyarwanda Score (%)	761	0.26	1.51	0.66	4.85	2.44	7.37	38.00	16.00	41.99	16.67	55.07	18.83
T3 Mathematics Score (%)	761	42.30	15.98	41.21	15.14	40.65	13.83	34.23	14.92	33.48	14.87	51.47	17.23
T3 SET Score (%)	761	40.61	18.41	38.90	16.69	35.27	15.76	31.96	15.37	29.43	14.18	29.14	14.29
T3 SRS Score (%)	761	45.68	19.45	47.59	18.78	54.77	20.35	34.03	15.46	31.52	15.33	31.04	15.87
Pupil-level Characteristics													
Age	761	7.87	0.47	9.53	0.57	10.93	0.58	12.14	0.59	13.64	0.62	14.43	0.54
Gender	761	0.46	0.07	0.47	0.06	0.48	0.06	0.51	0.07	0.52	0.08	0.56	0.09
School-level Characteristics													
Pupils per grade	761	162.78	90.13	159.38	78.44	144.12	76.71	140.71	77.95	104.59	56.31	65.97	38.23
Streams per grade	761	3.20	1.46	3.13	1.36	2.87	1.29	2.63	1.26	2.28	1.03	1.64	0.75
Double shift model (1/yes)	761	0.23	0.42	0.18	0.38	0.12	0.33	0.00	0.06	0.00	0.00	0.00	0.00
Pupil to stream ratio	761	51.52	16.60	51.43	15.25	50.58	14.75	53.79	15.33	46.54	14.69	39.91	12.15
Province													
Eastern	761	0.24	0.43	0.24	0.43	0.24	0.43	0.23	0.42	0.23	0.42	0.23	0.42
Kigali City	761	0.06	0.25	0.07	0.25	0.07	0.25	0.07	0.25	0.07	0.25	0.07	0.25
Northern	761	0.17	0.38	0.17	0.38	0.17	0.38	0.18	0.38	0.18	0.38	0.18	0.38
Southern	761	0.25	0.43	0.26	0.44	0.26	0.44	0.26	0.44	0.26	0.44	0.26	0.44
Western	761	0.27	0.44	0.26	0.44	0.26	0.44	0.26	0.44	0.26	0.44	0.27	0.44
Program phase													
Cohort 1	761	0.13	0.33	0.13	0.34	0.13	0.34	0.13	0.33	0.13	0.33	0.13	0.33
Cohort 2	761	0.20	0.40	0.20	0.40	0.20	0.40	0.20	0.40	0.20	0.40	0.20	0.40
Cohort 3	761	0.68	0.47	0.67	0.47	0.67	0.47	0.68	0.47	0.68	0.47	0.67	0.47

Notes: **Source:** Author's computation using RwandaEQUIP (Endline) data

with a maximum of 69 pupils—approximately the national average at that time. Nonetheless, some school-grades operated either fully or partially under a double-shift model due to capacity constraints and/or strong head teacher preferences. In some cases, there may have been some selection into double shifts beyond the 69-pupil threshold – which was also not fully binding. These non-binding, yet important class size thresholds generated quasi-experimental variation in class size across grades and schools.

Formally, we estimate the following two-stage least squares (2SLS) model:

First Stage:

$$D_{sg} = \propto_0 + \propto_1 .1\left(Z_{sg} \geq c\right) + \propto_2 f\left(Z_{sg} - c\right) + \varepsilon_{sg} \quad (1)$$

Where D_{sg} is a measure of the treatment assignment. We define treatment as a binary indicator equal to 1 if the number of streams at the school-grade level is greater than one, and 0 otherwise. In other words, $D_{sg} = 1\left[Streams_{sg} > 1\right]$. Hence, school grades with enrollment just above 69 are split into more streams, leading to smaller class sizes (lower pupil-to-stream ratio). In addition, since class size is given by the ratio of enrollment to the number of streams, an increase in the number of streams mechanically reduces the number of pupils per classroom. $1\left(Z_{sg} \geq c\right)$ is an indicator for being above the class size threshold, and $f\left(Z_{sg} - c\right)$ is a smooth function of the running variable (number of pupils per grade, a measure of enrolment). For the analysis, the running variable is centered at zero and is measured as the distance from the cutoff, that is, difference between the numbers of pupils per grade. Hence, a negative value of the running variable indicates that no additional class stream was created (control), a positive value indicates that additional class streams were created (treatment) and the cutoff is at zero.

Second Stage:

$$Y_{sg} = \beta_0 + \beta_1 .\widehat{D_{sg}} + \beta_2 f\left(Z_{sg} - c\right) + \mu_{sg} \quad (2)$$

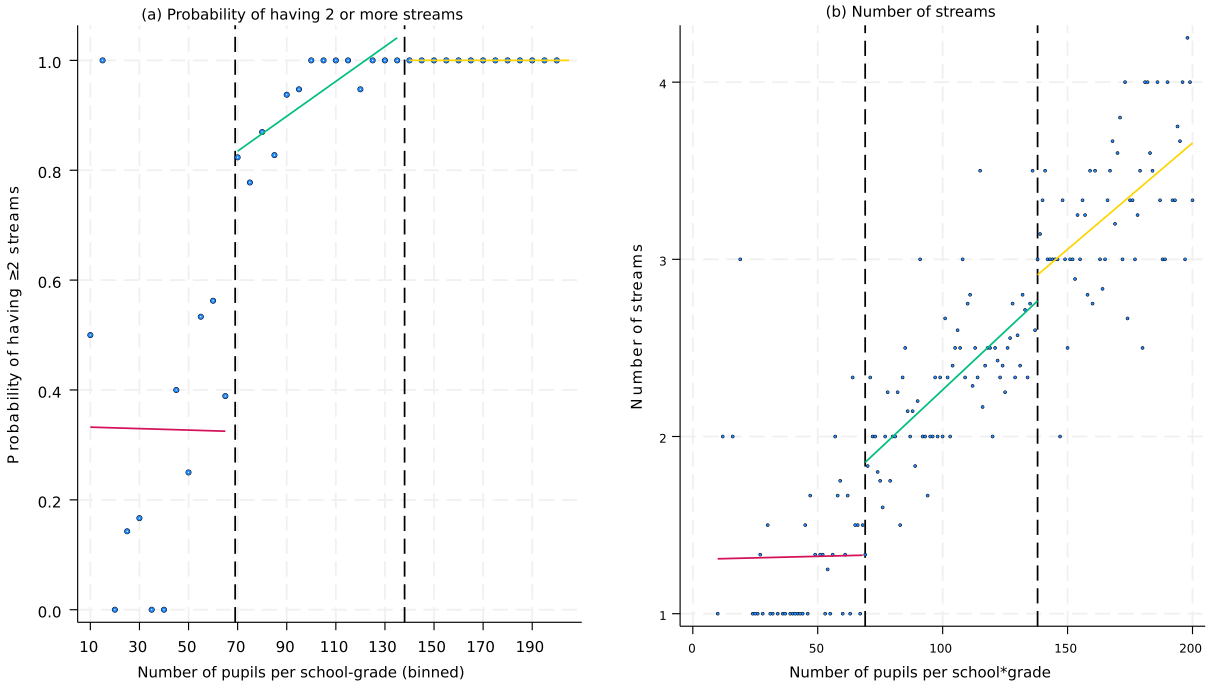
Where Y_{sg} represents the outcomes, that is, the exam scores (English, Mathematics, Science and Elementary Technology (SET), Social and Religious Study (SRS), and Kinyarwanda) for both term one and three, and $\widehat{D_{sg}}$ is the predicted treatment assignment from the first stage. With the inclusion of the controls, the full estimation equations are specified in equations 3 and 4 below:

$$Streams_{sg} = \propto_0 + \propto_1 1\left(enrolment_{sg} \geq 69\right) + f\left(enrolment_{sg} - 69\right) + X'_{sg}\gamma + \varepsilon_{sg} \quad (3)$$

$$Y_{sg} = \beta_0 + \beta_1 streams_{sg} + f\left(\widehat{enrolment_{sg}} - 69\right) + X'_{sg}\delta + \varepsilon_{sg} \quad (4)$$

Where X_{sg} includes baseline exam scores in the relevant subjects, average age, share of female pupils, indicators for province, programme phase, and double-shift status. All regressions include a flexible polynomial in centred enrolment $f\left(enrolment_{sg} - 69\right)$. In addition, all main

Figure 2: Discontinuity in class streams at enrolment threshold



Notes: Source: Author's computation using RwandaEQUIP data (grade 3) (1) In panel a, the data are binned in 5-pupil intervals. (2) The dashed vertical lines indicate enrollment thresholds implied by the class-size rule. The first threshold at 69 pupils is used for identification in the regression discontinuity design, while the second threshold at 138 pupils is shown for institutional context, as it governs the shifts from two to three streams.

outcome and covariate variables used in the 2SLS regressions are averaged to the school–grade level, consistent with the fact that both the running variable (enrolment per grade) and the treatment (class splitting) are defined at that level. Standard errors are clustered at the school–grade level to account for within-cohort correlation.

However, given that the running variable, is only available at endline, we implement the fuzzy RDD cross-sectionally at endline. Based on this, we use the baseline outcomes (exam scores) as controls. This helps absorb any pre-existing variation in learning outcomes and improves precision (Bruhn and McKenzie, 2009; McKenzie, 2012). Consequently, the estimated treatment effect is a Local Average Treatment Effect (LATE) of class size reduction on academic performance for schools near the cutoff. Importantly, the validity of this design relies on the assumption that no other school-level or student-level characteristics change discontinuously at the cutoff, and that there is no manipulation of the running variable around the threshold (Hahn, Todd, and Van der Klaauw, 2001; Imbens and Lemieux, 2008; McCrary, 2008). The availability of data across multiple grades and subjects within the same school offers a significant empirical advantage by enabling the exploitation of within-school variation in class size. This within-school variation is particularly valuable when implementing regression discontinuity designs, as it strengthens causal inference by reducing omitted variable bias associated with cross-school comparisons (Hahn, Todd, and Van der Klaauw, 2001; Angrist & Pischke, 2009).

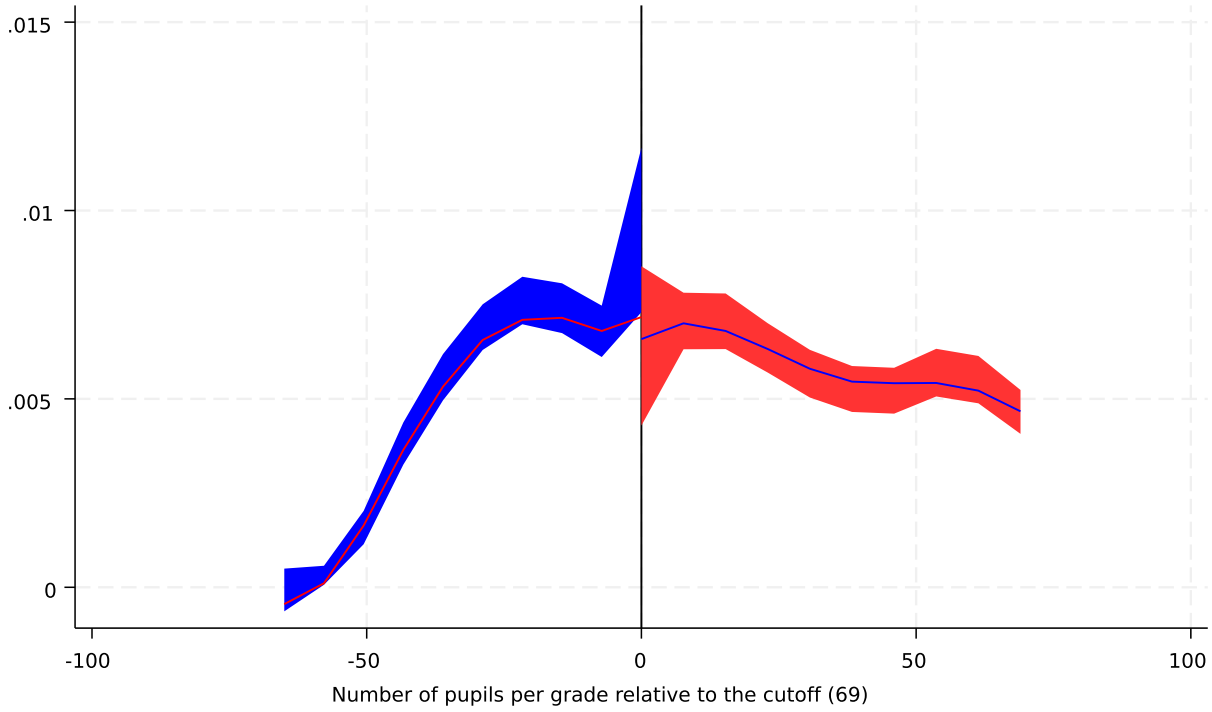
Furthermore, it is important to note that our LATE captures the effect of the particular way RwandaEQUIP induces changes in class structure — through splitting enrolment into additional streams — rather than the effect of an arbitrary or costless reduction in class size. In principle, this organizational change could simultaneously alter teacher workload alongside class size, complicating the causal interpretation. However, this concern is substantially mitigated by the institutional context of our study. Under the RwandaEQUIP programme, the large majority of schools (91%) were transitioned from the traditional double-shift model to a single-shift model as part of the programme's implementation strategy. In these schools, stream splitting at the enrolment threshold operates within an already-consolidated instructional day and does not additionally burden teachers with more shifts. The bundling of class size and workload changes is therefore most relevant for the small minority of schools (9%) that retained the double-shift model, which we examine explicitly in our heterogeneity analysis.

3.2.1 Identification Assumptions

A key identifying assumption in regression discontinuity (RD) designs is that units just below and above the treatment threshold are locally comparable, and that there is no manipulation of the running variable around the cutoff (Lee & Lemieux, 2010). To rule out any manipulation of the running variable—number of pupils per grade— at the cut-off, we employ the McCrary test. The McCrary (2008) density test is commonly employed to assess whether there is a discontinuity in the density of the running variable, which could suggest sorting or strategic manipulation. Figure 3 shows the result of the McCrary test. We find no evidence of bunching or manipulation at the cut-off: the density plot shows smooth densities on both sides of the cut-off with no visible discontinuity. Among other things, one contextual factor that supports the plausibility of the no-manipulation result is centrally managed nature of school enrolment in Rwandan public primary schools. This minimizes the opportunity for household- or school-level manipulation around the threshold (Rodriguez-Segura et al., 2023).

In addition, to rule out possible selection in the number of pupils per grade (enrollment), we conduct balance tests across the cutoff on a range of baseline covariates, including baseline literacy and numeracy levels, age, and gender. We conduct the balance checks using data from Phase 3 schools only. This ensures that all observations were untreated at baseline, given that the RwandaEQUIP programme was rolled out in Phase 3 schools beginning later in 2024, when the baseline assessment was conducted. Restricting to this group allows for a valid assessment of pre-treatment balance in student and school characteristics around the cutoff. The lack of statistically significant discontinuities in these characteristics provides evidence against manipulation (Cattaneo, Idrobo, & Titiunik, 2020). The result as presented in figure 4 below shows that there are no significant or visible jump around the cutoff for any of the baseline variables

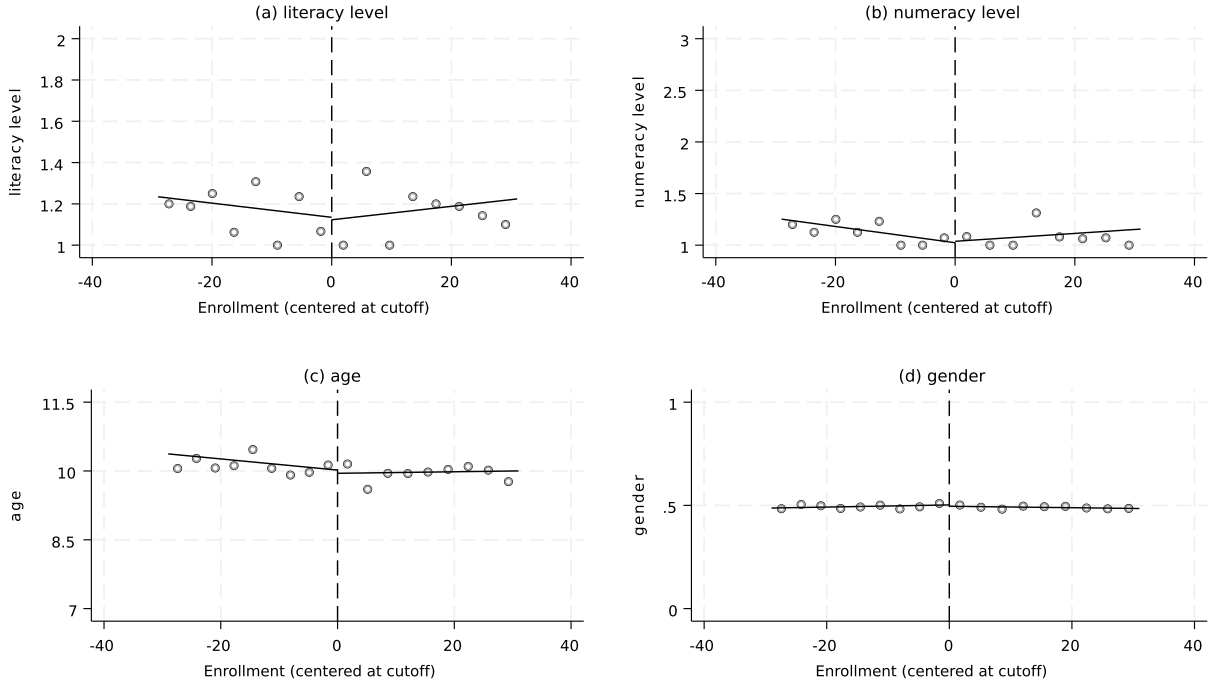
Figure 3: McCrary Test Plot



Notes: Source: Author's computation using RwandaEQUIP data

including age, gender and literacy and numeracy levels. The results are further corroborated with the balance test using the t-test, as shown in appendix one.

Figure 4: Balance Test (Phase 3 schools only)



Notes: Source: Author's computation using RwandaEQUIP data

While the local randomization checks in Appendix 4b point to some differences in predetermined covariates within very narrow windows, our main specifications rely on parametric extrapolation over a wider enrolment range and include extensive controls. The balance evidence should therefore be interpreted as suggestive rather than definitive.

A further concern relates to potential measurement error in the administrative counts of pupil and stream counts, which could attenuate first-stage estimates and bias the 2SLS estimator if mismeasurement is correlated with outcomes. Two features of our variable definitions and data mitigate this concern. First, both the treatment indicator — defined as a binary variable equal to one when the number of streams exceeds one — and the instrument — defined as a binary threshold indicator for enrolment above 69 pupils — are less sensitive to precise count mismeasurement than continuous analogues would be, as minor miscounting does not alter the binary classification of either variable. Second, an inspection of the distribution of the pupil-to-stream ratio, presented in appendix 17, reveals a smooth, unimodal distribution concentrated between approximately 20 and 100 pupils per stream, with no suspicious spikes or bunching at round numbers that would be characteristic of systematic administrative misreporting. The absence of such patterns, combined with the centralized nature of school data collection under the RwandaEQUIP programme, provides evidence that measurement error in the underlying counts is unlikely to be a material source of bias in our estimates.

3.2.2 Discrete Running Variable and the Extrapolation-Based IV Approach

An investigation of the density of the running variable shows the existence of mass points, which arises from the discrete nature of the running variable (number of pupils per grade) in the data. Specifically, there are many school-grade level observations in the data that share the same value of number of pupils per grade. This arises because enrollment is measured as an integer count at the school-grade level; the running variable exhibits substantial mass points near the cutoff, limiting the precision of local RD comparisons. This creates a natural departure from the canonical RD design. Our empirical investigation shows that there are 373 distinct values (mass points) out of the total sample size of 4,471 observations; this implies that there are 12 observations per value. Additionally, in appendix 4a and 4b, we show the number of observations for three mass points closest to the cutoff and the RD effect of the treatment on predetermined covariates within the local randomization RD framework, respectively. In sum, the result shows that there is an insufficient number of mass points to estimate the canonical RD parameter, as bandwidth selection becomes inappropriate and undermines the local randomization assumptions (Cattaneo et al., 2024). As a result, we do not interpret the threshold as generating quasi-experimental variation in the usual RD sense. However, to rule out any violation of the continuity assumption of the RD design given the discrete nature of the running variable and the presence of mass points, we visually examine the distribution of the standardized learning outcomes around the

threshold. The result in appendix 5 shows that there is no evidence of discontinuities in the outcomes distributions at the cutoff, thereby supporting the continuity assumption underlying the study design.

As emphasized by Cattaneo et al. (2024), in the presence of a discrete running variable with few mass points, the estimation of the canonical RD parameter necessarily relies on parametric extrapolation toward the cutoff. Hence, we adopt explicit extrapolation-based IV approach. We use an indicator for enrolment (number of pupils per grade) above the policy threshold as an instrument for class structure (measured by the number of streams), estimating the relationship between class organization and learning outcomes through a linear two-stage least squares (2SLS) framework. This approach can be interpreted as fitting a low-order parametric approximation to the conditional expectation functions on either side of the cutoff and extrapolating these functions to recover an effect at the threshold.

Formally, we assume that for enrolment values within a window [c – h, c + h], the conditional expectations $E\left[Y_{sg} \mid enrolment_{sg},\ above69_{sg}\right]$ *and* $E\left[streams_{sg} \mid enrolment_{sg},\ above69_{sg}\right]$ can be approximated by linear functions in centred enrolment. Under this assumption, and given the observed first-stage discontinuity in streams at 69 pupils, our 2SLS estimator recovers a local average treatment effect for schools near the threshold. In other words, identification does not rely on local randomization at the cutoff, but rather on the assumption that the conditional relationship between enrollment-induced class structure and learning outcomes is well approximated by a linear function in a neighbourhood covering the available support of the running variable.

In canonical regression discontinuity designs, bandwidth selection involves a bias-variance tradeoff governed by local smoothing around the cutoff. However, as discussed above, the discrete nature of our running variable precludes this approach. Under the extrapolation-based IV framework we adopt following Cattaneo et al. (2024), identification rests on the assumption that conditional expectation functions are well-approximated by a linear polynomial over the observed support of the running variable. Consequently, the relevant sample restriction is not a narrow local window around the cutoff, but rather a window that retains sufficient support for the parametric approximation to be credible while excluding extreme outliers likely driven by measurement error or unusual institutional arrangements. Our main specifications therefore restrict the sample to school-grade observations with enrolment within 100 pupils of the class-size threshold (i.e., |running variable| ≤ 100), which retains 74% of observations.

We therefore interpret our estimates as extrapolation-based local effects for schools near the class size threshold, rather than as nonparametric RD estimates. All results are presented alongside extensive diagnostics and robustness checks, and should be interpreted with appropriate caution given the reliance on parametric structure. As expected under the extrapolation-based IV framework, estimates become less precise and statistically insignificant when the sample is

restricted to narrow enrolment windows around the threshold. This reflects the discrete nature of the running variable: with few distinct mass points in a narrow window, the linear functional form cannot be estimated with sufficient precision. This pattern is consistent with the theoretical motivation for the extrapolation approach — namely, that identification in this setting requires leveraging variation across a wider range of the running variable rather than relying on local comparisons at the cutoff.

4 Results

In this section, we present and discuss the results of our analyses. Our findings are organized in the following sections: Section 4.1 presents evidence on the impact of class size reduction on learning outcomes among public primary school pupils in Rwanda. Section 4.2 presents the heterogeneous effects of class size reduction by grade, the shifting model and program phase. In section 4.3, we present evidence on the potential mechanism, specifically, individualized attention, while section 4.4 shows the results of the robustness checks.

5 Effect of Class Size on Learning Outcomes

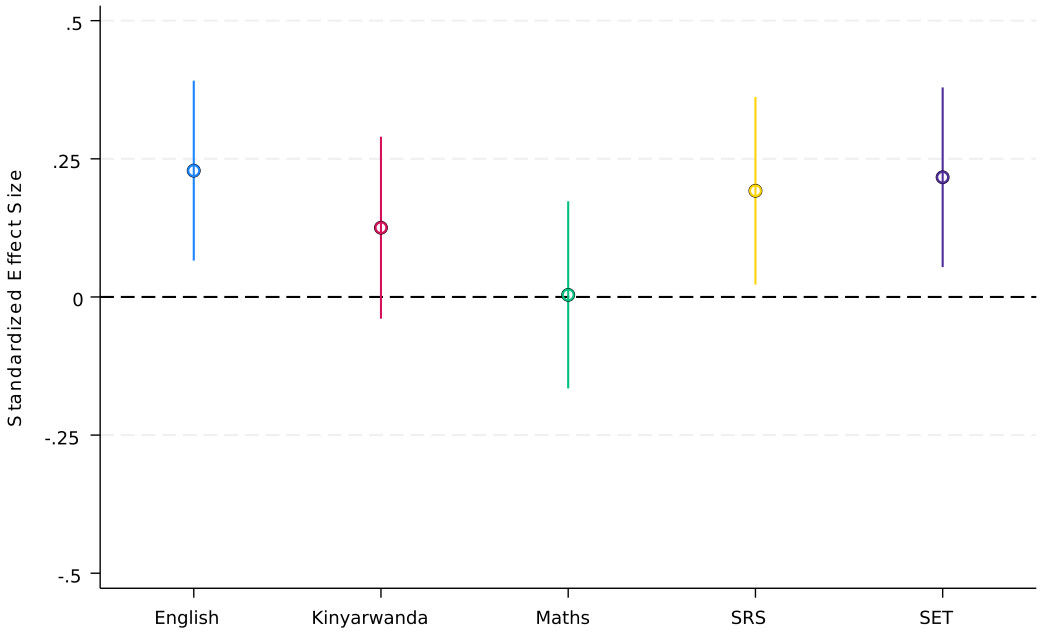
We present instrumental variables (IV) estimates of the effect of enrollment-induced class structure, specifically, class splitting into streams, on pupil learning outcomes. The results are presented separately for short-run (Term 1 - T1) and medium-run (Term 3 - T3) outcomes. For each term, we report specifications without covariates and covariate-adjusted specifications, including pupil and school-level characteristics, to assess the robustness and precision of the estimates. In all the models, an indicator for enrollment above the policy-implied threshold is used as an instrument for class structure, and estimation is restricted to school–grade cohorts operating within a neighborhood of the threshold. It is noteworthy to state that based on the discussion in section three, the reported estimates should be interpreted as extrapolation-based IV effects anchored at the enrollment threshold, rather than as locally randomized regression discontinuity estimates.

5.1 Short-run Effects (Term 1 Outcomes)

The T1 results presented in figure 5 provide evidence on the short-run effects of enrollment-induced changes in class structure, mainly from splitting classes into streams, during the first term of the academic year. The results show that class size reduction through increase in number of class streams have positive effects on learning outcomes across all the subjects except mathematics, which is negligible. However, the effects are only statistically significant for En-

glish, SRS and SET. Specifically, we find that class size reduction by one standard deviation (SD) – equivalent to approximately 16 fewer pupils per stream - leads to a 0.23 SD increase in English scores, 0.13 SD increase in Kinyarwanda scores, 0.22 SD increase in SET scores and 0.20 SD increase in SRS scores. Appendix 6 presents the detailed table of results for the T1 outcomes.

Figure 5: Effect of class size reduction on term 1 (T1) learning outcomes



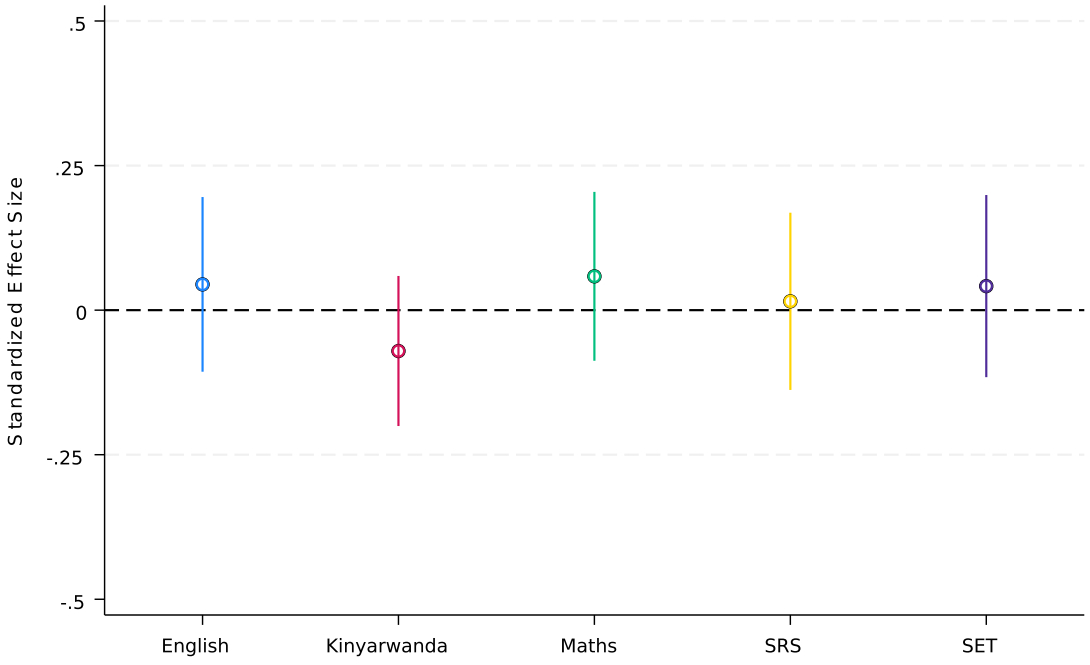
Notes: **Source**: Author's computation using RwandaEQUIP data; *p<0.10; **p<0.05; ***p<0.01 All outcome variables are standardized to have mean zero and unit standard deviation. Term one (T1) outcomes are based on assessments administered in the first school term, while Term 3 (T3) outcomes correspond to examinations administered at the end of the school session/year in the same subjects.

While our findings on other subjects adds to the class size and learning outcomes literature by exploring other subjects besides mathematics and English, our results on the effects on English are consistent with previous evidence of a positive and modest effects of class size on literacy scores (Angrist and Lavy, 1999; Moshoeshoe, 2015) but not for mathematics. However, our evidence of a negligible (0.004 SD) and insignificant effect of class size reduction on mathematics is similar to the findings of Moshoeshoe (2015) who finds no significant effect on mathematics scores in Lesotho. The results are robust to covariate-adjustment of the model through the inclusion of pupil and school-level control variables. However, a plausible explanation for the negligible and insignificant effects we find in mathematics is the fact that unlike English language, students may require less teacher-pupil interaction to build proficiency in Mathematics skills (Moshoeshoe, 2015). In addition, mathematics is a skill-intensive subject, the mastery of which is built cumulatively over time, hence, the T1 period may be too short a time to find significant changes in students' performance in mathematics.

5.1.1 Medium-run Effects (Term 3 Outcomes)

We additionally assess the changes in the pupils' learning outcomes at the end of the school year, which is measured by the term 3 exam scores. In figure 6, we find positive effects of class size reduction on pupils' examination scores across all subjects except for Kinyarwanda. Specifically, the results show that in the medium run, a one standard deviation decrease in class size – equivalent to approximately 16 fewer pupils per stream - results in a 0.04 SD increase in English scores, 0.06 SD increase in mathematics scores, 0.04 SD and 0.02 SD increase in SET and SRS scores, respectively. However, we find that in addition to the reduction in the magnitude of effects between T1 and T3 estimates across all subjects, none of the effects is statistically significant. This is an indication that the gains in learning outcomes from class size reduction through class splitting, fade out over time.

Figure 6: Effect of class size reduction on term 3 (T3) learning outcomes



Notes: **Source**: Author's computation using RwandaEQUIP data; *p<0.10; **p<0.05; ***p<0.01 All outcome variables are standardized to have mean zero and unit standard deviation. Term one (T1) outcomes are based on assessments administered in the first school term, while Term 3 (T3) outcomes correspond to examinations administered at the end of the school session/year in the same subjects.

Furthermore, our result shows that the effect on mathematics score is higher in magnitude (0.06 SD), though not significant, in the medium-run (term 3), changing. The results are also shown in appendix 7. This supports the conjecture from the discussion of the T1 findings that the intensive nature of mathematics, implies that the mastery of it will build up over time. Overall, our findings are consistent with previous evidence of small and statistically insignificant effects of class size reduction on student learning outcomes (Hanushek, 1986; Hoxby, 2000; Asadullah, 2005; Duflo, Dupas, and Kremer, 2015; Crawfurd and Le Nestour, 2022).

5.2 Heterogeneity

In this section, we examine the heterogeneous effect of class size reduction, specifically, we assess the impact by grade level and the shift model operated by the school, that is, whether or not the school operates a double shift system. These analyses follow from the literature, which finds that in some cases, effects of class size varies by pupil, household and school-level characteristics (Angrist and Lavy, 1999; Moshoeshoe, 2015; Crawfurd and Le Nestour, 2022).

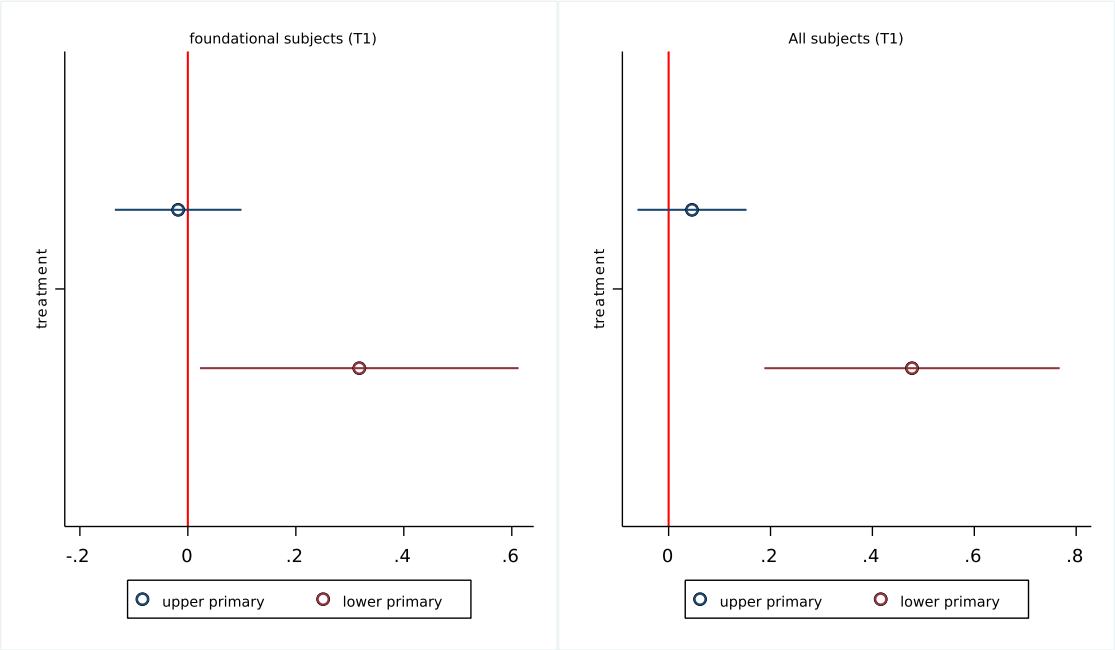
5.2.1 Differences by Grade Level

Figures 7 and 8 shows the effect of class size on the learning outcomes by grade levels for T1 and T3, respectively. We present the results by the individual subjects in appendix 8. For this analysis, we categorize the grades into lower primary and upper primary, where lower primary corresponds to grades one to three and upper primary corresponds to grades four to five. In addition, for ease of comparison and presentation, we categorize the subjects into foundational subjects and all the subjects together. The foundational subjects include English, Kinyarwanda and Mathematics, while the all subjects category additionally includes Set and SRS. The results show that reducing class size by splitting classes into streams leads to a significant improvement in the performance of pupils in lower primary in the foundational subjects but not for those in upper primary. Specifically, from appendix 10, we find that in the short-run, splitting the classes into additional streams significantly improves the learning outcomes of pupils in lower primary grades across all the subjects apart from mathematics, which is not statistically significant. For mathematics, the results reveal that class size reduction – equivalent to approximately 16 fewer pupils per stream - leads to a 0.04 SD increase in the exams score for pupils in lower primary, though the effect is not statistically significant. As previously mentioned, this effect could be attributed to the possible skill-intensive nature of mathematics, for which mastery is built cumulatively over time.

Furthermore, in the right panel of appendix 10, we find a similar trend for pupils in upper primary. In particular, class size reduction – equivalent to approximately 16 fewer pupils per stream - results in an increase in the pupils' exam scores across all the subjects except for mathematics, for which there is a decrease. However, only the effect on the exam score for English language is statistically significant, with a 0.20 SD increase. Overall, the results show that class size reduction through the introduction of additional class streams at the cutoff, has more effect on the academic performance of pupils in early grades (lower primary) than those in later or upper grades.

In the medium- run, our results in appendix 11 show a similar pattern to those of the short-run effects, however, there is a reduction in the magnitude of the effects across all the subjects, except for mathematics, where we find a significant increase in magnitude by 0.40 SD. In addition, we

Figure 7: Effect of class size on T1 learning Outcomes by grade level



Notes: **Source**: Author's computation using RwandaEQUIP data All outcome variables are standardized to have mean zero and unit standard deviation. *Foundational* composite scores are constructed as weighted averages of English, Kinyarwanda, and Mathematics exam scores. *All-subject* composite scores additionally include Social and Religious Studies (SRS) and Science and Elementary Technology (SET). Term one (T1) outcomes are based on assessments administered in the first school term, while Term 3 (T3) outcomes correspond to examinations administered at the end of the school session/year in the same subjects.

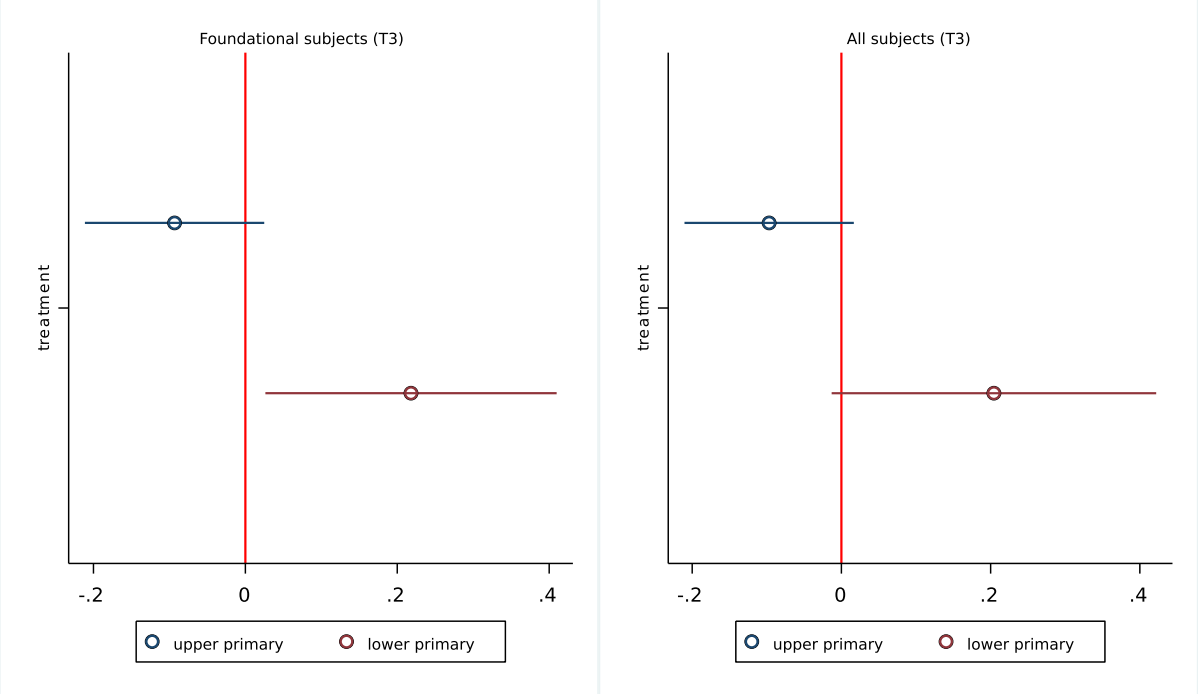
find that contrary to term one, in term 3, the effect on Kinyarwanda, though positive, is no longer significant. For mathematics, it is assumed that the gains in numeracy skills are beginning to set in given the period from programme implementation, hence, supporting our initial assumption that numeracy skill builds up with time. However, for the pupils in upper primary, the effect on mathematics, though statistically significant, is negative, while the effect on the exam scores for all other subjects is negative. This point to the fact that the gains in learning outcomes from the reduction in class size "fade out" over time, more so for children is higher grades.

A plausible explanation for the differences in effects by grade level is that the close interaction that results from class size reduction may be more beneficial for pupils in earlier grade, as they require more guidance during learning compared to those in higher grades (Crawfurd and Le Nestour, 2022). This finding is consistent with the previous evidence on the heterogeneous effects of class size by grade (Angrist and Lavy, 1999; Crawfurd and Le Nestour, 2022; Rodriguez-Segura et al., 2023).

5.2.2 Differences by Shifting System

In this section, we explore the effects of class size reduction by the shifting system adopted by the school, specifically, whether or not the school operates a double shift versus a single shift

Figure 8: Effect of class size on T3 learning Outcomes by grade level

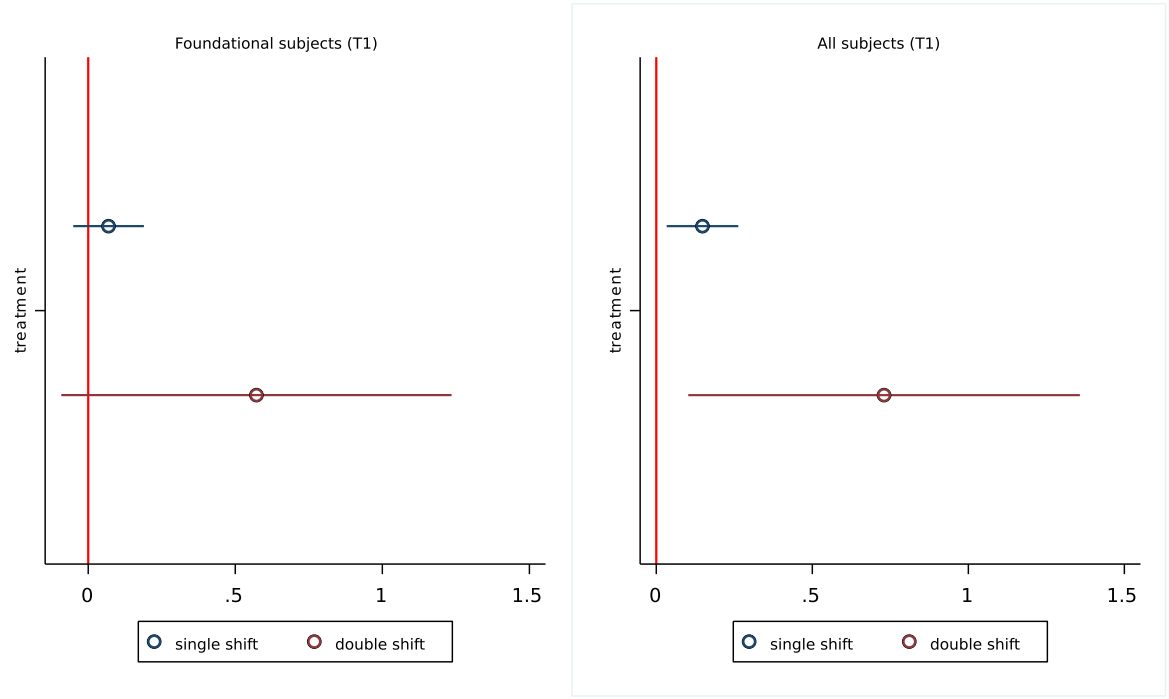


Notes: **Source**: Author's computation using RwandaEQUIP data All outcome variables are standardized to have mean zero and unit standard deviation. *Foundational* composite scores are constructed as weighted averages of English, Kinyarwanda, and Mathematics exam scores. *All-subject* composite scores additionally include Social and Religious Studies (SRS) and Science and Elementary Technology (SET). Term one (T1) outcomes are based on assessments administered in the first school term, while Term 3 (T3) outcomes correspond to examinations administered at the end of the school session/year in the same subjects.

system. Under the double shift model, approximately half the pupils attend school in the morning (shift), while the other half attend in the afternoon. The double shift model is the business-as-usual approach across public primary schools in Rwanda, implemented with the aim of ensuring that pupils receive instruction in smaller classes and to some extent that schools have enough physical capacity to accommodate all their pupils within classrooms (Rodriguez-Segura et al., 2023). However, under the RwandaEQUIP programme, most schools were transitioned into the single shift model, with the introduction of the class size threshold and class streams, as ways of avoiding overcrowding.

Our results from figure 9 show that pupils in the single shift model classes experienced larger, and statistically significant gains in their learning outcomes across all the subjects compared to those in the double shift schools. Specifically, from appendix 12, we find that the in single shift schools, splitting classes into streams at the class size threshold leads to 0.18 SD increase in English scores, 0.25 SD and 0.20 SD increase in SRS and SET scores, respectively, in the short-run (term one). We find no significant effect of class size reduction on the learning outcomes among the schools operating the double shift model. One plausible explanation is that, in double-shift settings, reductions in class size are accompanied by increases in teacher workload, as teachers are required to teach multiple shifts within the same day. This may lead to fatigue,

Figure 9: Effect of class size on T1 learning outcomes by shifting model



Notes: **Source**: Author's computation using RwandaEQUIP data All outcome variables are standardized to have mean zero and unit standard deviation. *Foundational* composite scores are constructed as weighted averages of English, Kinyarwanda, and Mathematics exam scores. *All-subject* composite scores additionally include Social and Religious Studies (SRS) and Science and Elementary Technology (SET). Term one (T1) outcomes are based on assessments administered in the first school term, while Term 3 (T3) outcomes correspond to examinations administered at the end of the school session/year in the same subjects.

reduced preparation time, and lower instructional quality, offsetting the potential benefits of smaller classes. As a result, class size reductions achieved through adjustments in class structure may be less effective in double-shift schools, where teacher effort and instructional intensity are already constrained. Hence, in line with some evidence in the literature, class size reductions are effective, so long as they do not come at the cost of increased instructional burden for the teachers (Duflo, Dupas, and Kremer, 2015). In addition, as before the gains in learning outcomes fade-out over time, with a reduction in the magnitude of the effect size for both the double and single shift schools, as shown in appendix 13.

Nevertheless, it is important to note that since 91% of schools in our sample operate the single-shift model — the main model under RwandaEQUIP — our main estimates primarily reflect the effect of class reorganization in an environment where teacher workload is not additionally burdened by shift multiplication. The double-shift results should therefore be interpreted as evidence of a boundary condition rather than a general qualification of the main findings.

5.2.3 Differences by Programme Phase

We further examine whether the effects of class size reduction vary across the three phases of the RwandaEQUIP programme rollout, which serve as a proxy for time since programme implementation. It is expected that if longer programme exposure facilitates greater consolidation of class size management practices and teacher capacity, then, Phase 1 schools —that had approximately two years of programme exposure at the time of data collection — should exhibit larger class size effects than Phase 2 or Phase 3 schools. Figure 10 presents T1 IV estimates separately by programme phase, with Phase 1 comprising 299 observations, Phase 2 comprising 499 observations, and Phase 3 comprising 2,509 observations within the estimation window. Contrary to this expectation, the results show no clear or consistent difference in class size effects across the three programme phases. Specifically, for English, Phase 2 produces the largest positive point estimate, while for SRS, Phase 1 shows the largest positive estimate. For Mathematics and SET, the phase ordering is inconsistent across terms. Moreover, across all subjects, none of the phase-specific differences is statistically distinguishable from each other at the conventional significance levels.

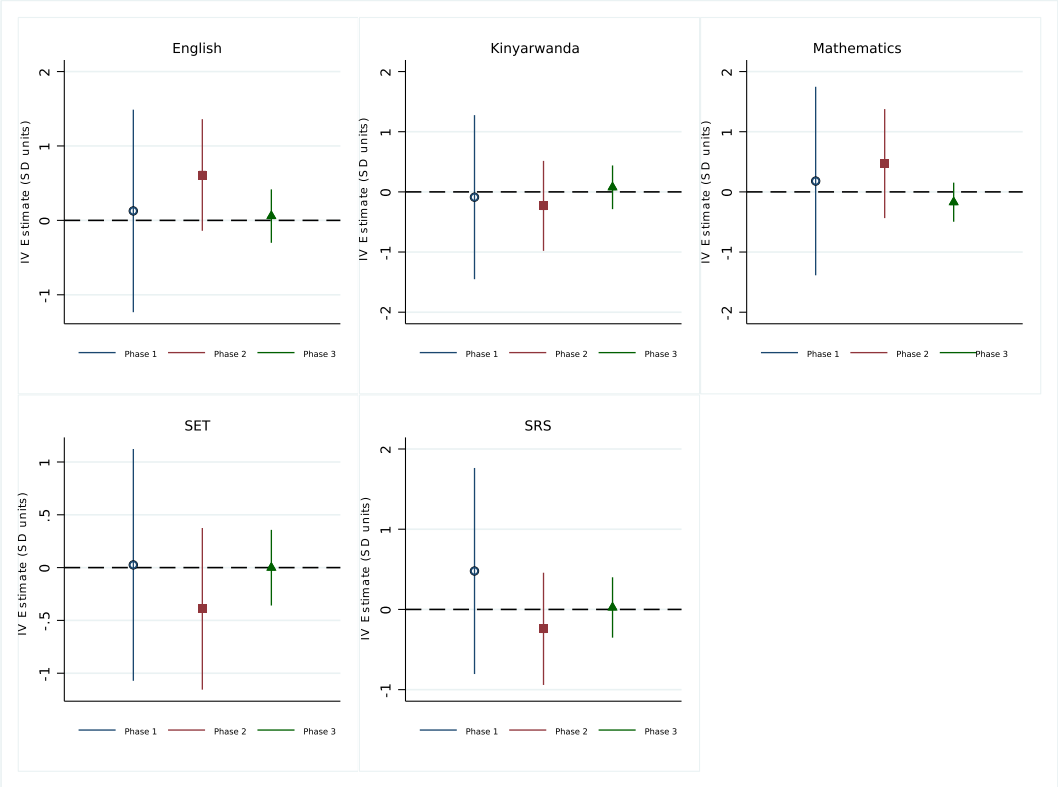
This pattern suggests that the class size reduction effects documented in the main analysis do not meaningfully accumulate with programme exposure within the range observed in our data — approximately a few months to two years. One plausible explanation is that the structured pedagogy model of RwandaEQUIP, which provides teachers with scripted lesson plans and standardized instructional materials, reduces the degree to which class size benefits depend on accumulated teacher experience or programme familiarity. When instructional content is highly standardized, the marginal value of additional programme exposure may be lower than in settings where teachers must develop their own responses to smaller classes. This interpretation is consistent with evidence from structured pedagogy programmes in other low-income contexts, where learning gains have been found to materialize relatively quickly following programme rollout rather than accumulating gradually over multiple years (Piper et al., 2014; Kerwin and Thornton, 2021).

Nevertheless, it is important to exercise caution about conclusions from the phase-specific analysis. This is particularly because Phase 1 and Phase 2 subsamples are substantially smaller than Phase 3, producing wider confidence intervals that limit the statistical power to detect genuine phase differences even if they exist. The pooled specification, which includes programme phase indicators as controls, remains the preferred and most precisely estimated specification.

5.2.4 Differences by Effective Stream Size

In this section, we explore heterogeneity in treatment effects across different stream size thresholds (effective stream size). Specifically, the effective stream size represents the actual number

Figure 10: Effect of class size on T1 learning outcomes by program phase



Notes: **Source**: Author's computation using RwandaEQUIP data Phase 1 schools joined RwandaEQUIP in Feb 2022 (approx. 2 years before data collection). Phase 2 schools joined Sept 2022 (approx. 1 year). Phase 3 schools joined Sept 2024 (few months).

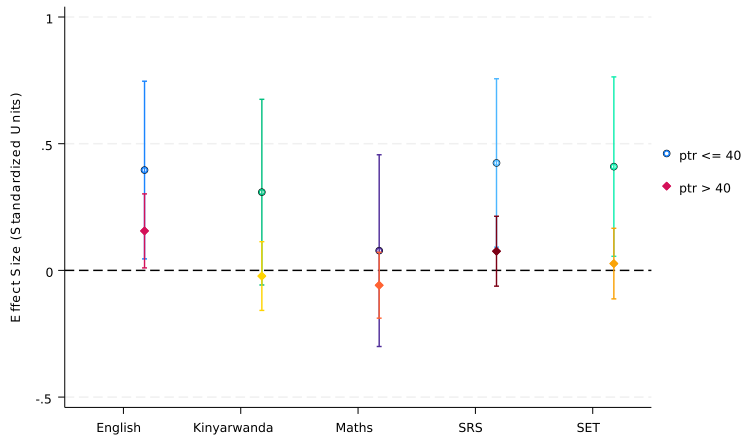
of pupils per stream that the teacher is responsible for during a lesson. This is particularly important to inform the maximum size of each class stream after splitting that will be most beneficial for learning. Overall, the discussion in this subsection allows us to go beyond the question of simply whether class sizes should be reduced, but how small they need to be for reductions to translate into meaningful learning gains**.** This follows from evidence in the literature that the starting size of a class matters for learning (Crawfurd and Le Nestour, 2022).

In Figure 11, we present the result of heterogeneity in the estimated treatment effects by effective stream size of 40 and 45 pupils, in panel A and B, respectively. For additional insights, we present the result with stream size of 35 pupils in appendix 14. In panel a, the results indicate that reductions in stream size generate substantially larger improvements in learning outcomes when streams contain at most 40 - 45 pupils. For English, SRS, and SET, the estimated effects for streams with PTR ≤ 40 range from approximately 0.40 to 0.42 standard deviations, compared with considerably smaller and statistically insignificant estimates for streams with PTR > 40. In panel b, the estimated effects for the same subjects (English, SRS and SET) range from approximately 0.31 to 0.34 standard deviations. In contrast, mathematics outcomes remain largely unaffected across both groups. Furthermore, in appendix 14, we explore further reductions in the stream size from 40 to 35 to test whether learning gains continue to increase when streams

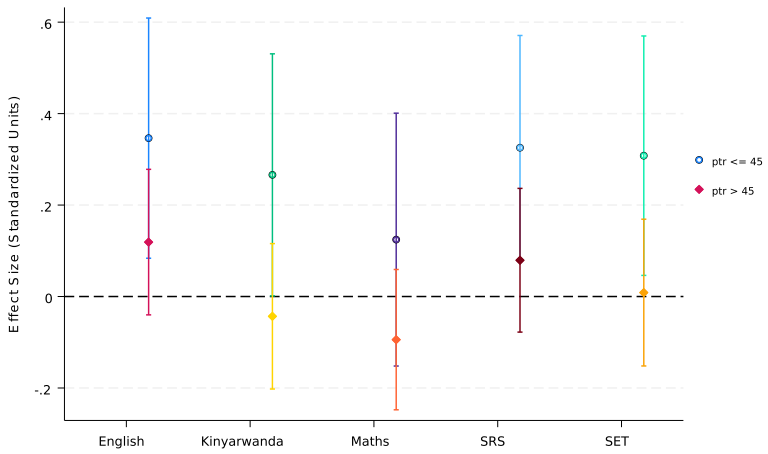
become even smaller than 40, or whether the gains plateau around 40 pupils. Although, the results are broadly consistent with the analysis based on the 40-45 pupil threshold, we find that most of the estimates are positive they are not statistically significant; the magnitudes are not substantially larger than those observed under the PTR ≤ 40 specification. This pattern suggests diminishing marginal returns to further reductions in stream size once streams fall below approximately 40 pupils.

These findings suggest that the benefits of reducing class stream sizes are concentrated among moderately sized streams and provide support for a policy threshold of approximately 40 to 45 pupils per stream, hence serving as a reasonable policy benchmark. This is consistent with evidence in the literature of the existence of nonlinearities in the relationship between class size reductions and learning outcomes; specifically, they argue that beyond a certain threshold, class size reductions may yield null or negative effects on learning outcomes ((Crawfurd and Le Nestour, 2022; Datta and Kingdon, 2023).

Figure 11: Heterogeneity by effective stream size



(a) Stream size threshold of 40 pupils



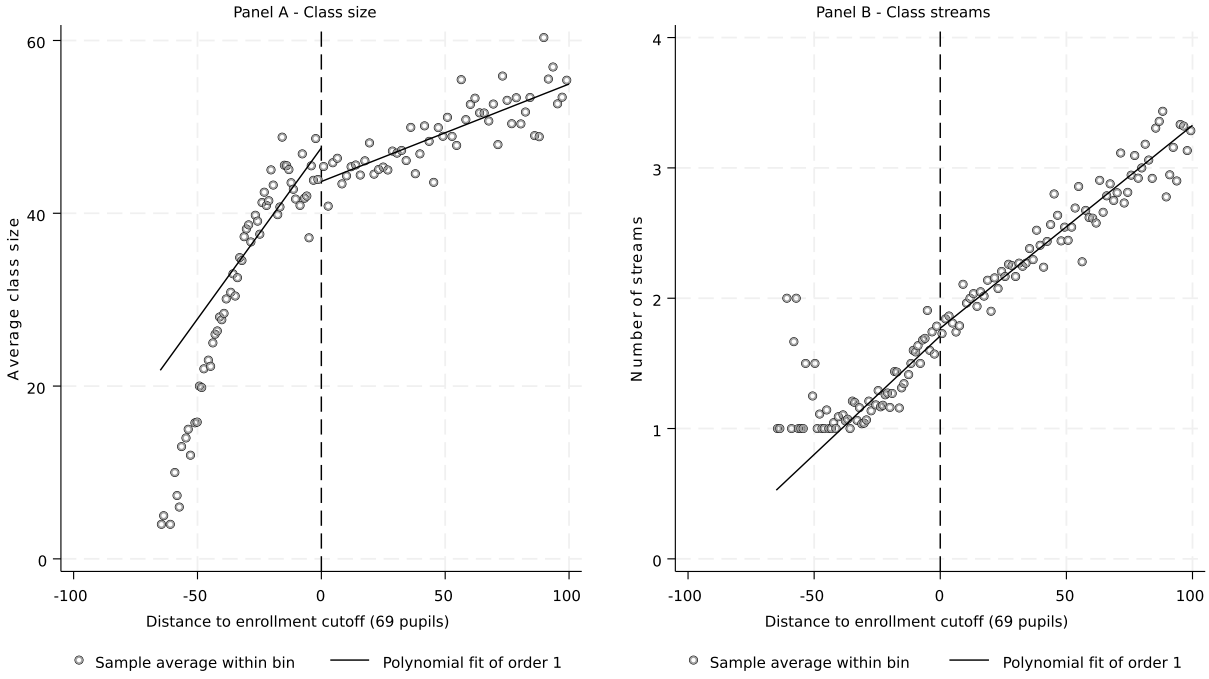
(b) Stream size threshold of 45 pupils

Notes: **Source**: Author's computation using RwandaEQUIP data All outcome variables are standardized to have mean zero and unit standard deviation. We only utilize T1 scores for this analysis.

6 Mechanism

In our discussions so far, we have seen that RwandaEQUIP schools respond to increases in enrollment by splitting classes into additional streams once enrollment reaches the programme threshold of 69 pupils. This mechanically reduces the number of students per classroom and alters the classroom environment in which instruction takes place. In this section, we examine the potential mechanisms through which these enrollment-induced changes in classroom organization affect student-learning outcomes. We focus in particular on individualized attention as a key channel, presenting suggestive evidence that reductions in class size may increase the amount of attention teachers can devote to each student, thereby improving learning outcomes.

Figure 12: Mechanism: Enrolment threshold and class splitting (Streams)



Notes: **Source**: Author's computation using RwandaEQUIP data

To illustrate this mechanism, Figure 12 plots the relationship between distance to the enrollment cutoff and two key measures of classroom organization: average class size and the number of classroom streams. The figure focuses on observations within a local bandwidth around the cutoff and reports binned averages together with local polynomial fits on either side of the threshold. The right panel (panel B) shows that the number of streams increases as enrollment approaches and exceeds the cutoff, indicating that schools respond to enrollment pressure by creating additional classroom streams. The left panel (panel A) shows that while enrollment increases near the threshold, average class size does not increase proportionally because schools split classes into multiple streams. Instead, class sizes increase more slowly as additional students are distributed across multiple streams. Together, these patterns provide sug-

gestive evidence that the enrollment threshold induces changes in classroom organization that reduce effective class size, thereby creating conditions that may facilitate greater individualized attention from teachers. The patterns in Figure 12 are consistent with the interpretation that enrollment-induced class splitting reduces the number of students per classroom, potentially allowing teachers to devote more attention to individual pupils.

7 Robustness Checks

Sample Window Sensitivity Tests

As discussed in Section 3.2.2, the extrapolation-based IV approach adopted for this study relies on functional form assumptions applied across the observed support of the running variable, rather than on local randomization at the cutoff. To assess the stability of our main estimates, we examine results across a range of sample inclusion windows both narrower and wider than our main specification of ±100 pupils. Specifically, we consider windows of ±20, ±25, ±30, and ±40 pupils below the main specification, as well as wider windows of ±60, ±80, and ±120 pupils above it. This range of windows brackets the main specification from both sides, allowing us to assess whether results are sensitive to either excluding observations that carry identification variation or including observations at the extremes of the enrolment distribution. Consistent with the theoretical expectations of the extrapolation-based IV framework, estimates at narrower windows are less precise but directionally consistent with main findings. Estimates at wider windows are similarly consistent in direction and magnitude, suggesting that the main results are not driven by the specific choice of a ±100 window. These results are presented in Appendix 15 and 16, and should be interpreted as evidence of directional stability rather than conventional bandwidth robustness.

Functional Form Sensitivity

We conduct an additional robustness check to strengthen the causal interpretation of our estimates. Specifically, we assess sensitivity to the linear functional form assumption by re-estimating all specifications with a quadratic polynomial in centered enrolment. The quadratic specifications yield sign reversals for several subjects, with statistically significant negative effects for Mathematics and SRS in Term 1 — contrary to the main results. Diagnostic investigation reveals that this reversal is driven by overfitting in the first stage rather than genuine nonlinearity in the enrolment-outcome relationship. As shown in Appendix 18, the quadratic fit produces an implausible U-shaped first-stage relationship on the left side of the cutoff, implying that schools with moderately low enrolments have a lower probability of class splitting than schools with very low enrolments — a pattern inconsistent with the institutional rule and the flat first-stage relationship documented in Figure 2. By contrast, the reduced-form relationship between enrolment and outcomes is virtually identical under linear and quadratic specifications

(Appendix 19), confirming that the sign reversal originates in the distorted first stage rather than in the outcome relationship. Although the quadratic term is statistically significant in the first-stage regression, this reflects the fitting of sparse and noisy observations in the far left tail of the enrolment distribution rather than genuine curvature near the threshold — precisely the overfitting problem identified by Gelman and Imbens (2019) in high-order polynomial RD specifications. The visual and institutional evidence strongly supports the linear specification as the more credible functional form in our setting, and we retain it as our preferred specification.

8 Conclusion

In this paper, we assess the impact of class size reduction on learning outcomes of pupils in public primary schools in Rwanda. We use administrative data from the RwandaEQUIP programme and exploit the discontinuities in class size from the enforcement of a maximum class size rule. We find that in the short-run, class size reduction leads to significant improvements in exam scores across all subjects (0.23 SD in English, 0.13 SD in Kinyarwanda, 0.20 SD in Social and Religious Studies (SRS), and 0.22 SD in Science and Elementary Technology (SET) per SD class size reduction) besides Mathematics, where the effect is negligible. In the medium-run, while the gains in learning outcomes remain, they are smaller in magnitude, indicating that the gains in learning outcomes from class size reduction through class splitting, fade out over time. However, for mathematics, the gains increase over time, most likely arising from the fact that mathematics is a skill-intensive subject, the mastery of which is built cumulatively over time.

We further show that the effect of class size reduction through the introduction of additional class streams on learning outcomes is more beneficial for pupils in early grades (lower primary) than those in later or upper grades, as well as for single shift schools. For the former, the finding possibly arises from the fact that the close interaction that results from class size reduction may be more beneficial for pupils in earlier grade, as they require more guidance during learning compared to those in higher grades. However, going beyond the question of simply whether class sizes should be reduced, we further explore how small the class size needs to be for reductions to translate into meaningful learning gains. Our extrapolation-based IV estimates suggest that the benefits of reducing class stream sizes are concentrated among moderately sized streams, with the largest estimated gains observed when streams contain roughly 40 to 45 pupils. While subject to the usual caveats about functional-form assumptions, this pattern provides empirical support for considering a class-size target in this range as a pragmatic policy benchmark in similar high pupil–teacher ratio contexts.

We provide suggestive evidence on the mechanism of individualized attention arising from the reductions in class size. We show that the class size threshold leads to changes in classroom organization (class splitting into additional streams) that reduce effective class size, thereby cre-

ating conditions that may facilitate greater individualized attention from teachers. In particular, the resulting smaller class sizes from the class reorganization potentially allows teachers to provide more attention to individual pupils.

This paper however is not without limitations. First, a key limitation of the empirical strategy arises from the discrete nature of the running variable. Specifically, because grade-level enrollment takes integer values and exhibits substantial mass points, standard continuity-based regression discontinuity methods cannot be implemented in their ideal form. As a result, the analysis relies on parametric extrapolation around the enrollment threshold, which introduces some degree of model dependence. Second, the analysis provides only suggestive evidence regarding the mechanisms underlying the observed effects. While the results are consistent with the hypothesis that reductions in class size increase opportunities for individualized attention, the available data do not contain direct measures of teacher–student interactions or instructional practices. As a result, the mechanism interpretation should be viewed as indicative rather than definitive. Third, the analysis does not incorporate detailed teacher-level information such as experience, subject specialization, or instructional quality, which could mediate the relationship between class size and learning outcomes. Fourth, given that our estimates are obtained from RwandaEQUIP schools, where teachers receive structured pedagogical support and digital tools, the external validity of our findings to non-EQUIP schools may be limited. In environments with weaker instructional support, the marginal effect of class size reductions could differ, potentially being larger if teachers struggle to manage very large classes or smaller if other constraints (e.g., teacher absence, infrastructure) bind more tightly.

References

Altinok, N., & Kingdon, G. (2012). New evidence on class size effects: A pupil fixed effects approach. *Oxford Bulletin of Economics and Statistics*, 74(2), 203–234.

Angrist, J.D. & Lavy, V. (1999). Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement, *The Quarterly Journal of Economics*, 114(2), 533-575, https://doi.org/10.1162/003355399556061

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Banerjee, A., Cole, S., Duflo, E., & Linden, L. (2007). Remedying education: Evidence from two randomized experiments in India. *Quarterly Journal of Economics*, 122(3), 1235–1264.

Bruhn, M., & McKenzie, D. (2009). In pursuit of balance: Randomization in practice in development field experiments. *American Economic Journal: Applied Economics, 1*(4), 200–232.

Cattaneo, M. D., Idrobo, N., & Titiunik, R. (2020). *A practical introduction to regression discontinuity designs: Foundations*. Cambridge University Press.

Cattaneo, M. D., Idrobo, N., & Titiunik, R. (2024). *A Practical Introduction to Regression Discontinuity Designs: Extensions*. Cambridge: Cambridge University Press.

Crawfurd, L., & Le Nestour, A. (2022). How much should governments spend on teachers? In J. Sandefur (Ed.), *Schooling for all: Feasible strategies to achieve universal education* (pp. 80–97). Center for Global Development.

Datta, S., & Kingdon, G. G. (2023). Class size and learning: Has India spent too much on reducing class size? *The World Bank Economic Review*, 37(1), 24–48.

Duflo, E., Dupas, P., & Kremer, M. (2015). School governance, teacher incentives, and pupil–teacher ratios: Experimental evidence from Kenyan primary schools. *Journal of Public Economics*, 123, 92–110.

Gelman, A., & Imbens, G. (2019). Why high-order polynomials should not be used in regression discontinuity designs. *Journal of Business & Economic Statistics*, 37(3), 447–456.

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

Imbens, G., & Lemieux, T. (2008). *Regression discontinuity designs: A guide to practice*. Journal of Econometrics, 142(2), 615–635.

Jepsen, C., & Rivkin, S. (2009). Class size reduction and student achievement: The potential tradeoff between teacher quality and class size. *Journal of Human Resources*, 44(1), 223–250.

Kerwin, J. T., & Thornton, R. L. (2021). Making the grade: The sensitivity of education program effectiveness to input choices and outcome measures. *Review of Economics and Statistics*, 103(2), 251–264.

Köhler, T. (2022). Class size and learner outcomes in South African schools: The role of school socioeconomic status. *Development Southern Africa*, 39(2), 126–150.

Krueger, A. B. (2003). *Economic considerations and class size*. *Economic Journal*, 113(485), F34–F63. https://doi.org/10.1111/1468-0297.00098

Lindo, J.M., Sanders, N.J. & Oreopoulos, P. (2010). Ability, Gender, and Performance Standards: Evidence from Academic Probation. *American Economic Journal: Applied Economics* 2 (2): 95–117.

Lindo, J.M., Sanders, N.J. & Oreopoulos, P. (2010). *Replication data for Ability, Gender, and Performance Standards: Evidence from Academic Probation*. Nashville, TN: American Economic Association. https://doi.org/103886E113751V1

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

McKenzie, D. (2012). Beyond baseline and follow-up: The case for more T in experiments. *Journal of Development Economics, 99*(2), 210–221.

McKenzie, D. (2012, June 24). *Experimental design and Ancova when everyone has the same baseline outcome value*. World Bank Development Impact Blog. https://blogs.worldbank.org/en/impactevaluations/experimental-design-and-ancova-when-everyone-has-same-baseline-outcome-value

Ministry of Education, Republic of Rwanda. (March 2025). *2023/2024 Education Statistical Yearbook*: School Year ended in July 2024. Kigali: Ministry of Education. https://www.mineduc.gov.rw/index.php?eID=dumpFile&t=f&f=123716&token=d67b88abed7ff24796f67eccd727ef58e2ebb6e0

Moshoeshoe, R. (2015). Average and heterogeneous effects of class size on educational achievement in Lesotho. *Economic Research Southern Africa Working Paper*, (496).

Piper, B., & Korda, M. (2010). Early Grade Reading Assessment (EGRA) Plus: Liberia. Program evaluation report. RTI International. Prepared for USAID/Liberia. https://www.eddataglobal.org/documents/index.cfm?fuseaction=pubDetail&ID=283

Piper, B., Zuilkowski, S.S., & Mugenda, A. (2014). Improving reading outcomes in Kenya: First-year effects of the PRIMR Initiative. International Journal of Educational Development, 37, 11-21. https://doi.org/10.1016/j.ijedudev.2014.02.006

Rodriguez-Segura, D., Rugwizangoga, P., & Lu, P. (2023). *Can data-informed management and structured pedagogy improve learning? Evidence from public schools in Rwanda*. RwandaEQUIP year 1 impact evaluation Report, April 2023.

Trines, S. (2019, October 15). *Education in Rwanda*. World Education News & Reviews. https://wenr.wes.org/2019/10/education-in-rwanda.

UNESCO (2022). *Spotlight on Basic Education Completion and Foundational Learning in Rwanda 2022*. UNESCO. https://knowledgehub.adeanet.org/files/documents/ADEA2022D007E.pdf

Urquiola, M., & Verhoogen, E. (2009). Class-size caps, sorting, and the regression-discontinuity design. *American Economic Review*, 99(1), 179–215.

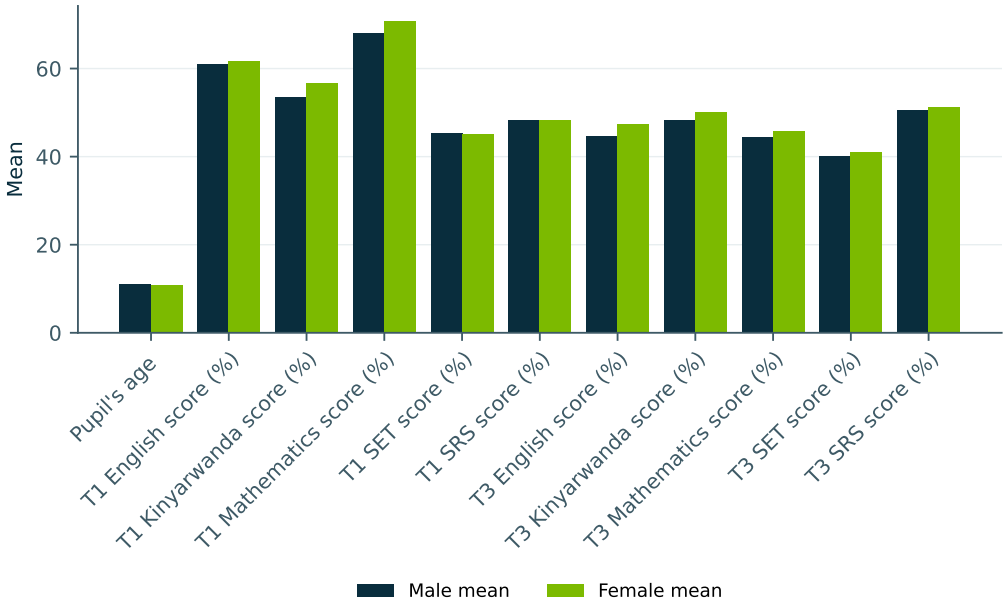
Appendix A: Additional tables and figures

Appendix 1: Balance Test (T-test)

Variable	Mean below cutoff (0)	Mean above cutoff (1)	Difference	Std. Error	T-test p-value
(A) Demographics					
Female (=1)	0.492	0.493	-0.001	0.002	0.768
Age (years)	10.068	10.035	0.032	0.030	0.272
(B) Baseline Test Scores					
Literacy score (P4)	1.185	1.164	0.020	0.022	0.351
Literacy score (P5)	1.766	1.774	-0.008	0.026	0.762
Literacy score (P6)	2.314	2.317	-0.003	0.027	0.915
Numeracy score (P4)	1.087	1.074	0.013	0.016	0.423
Numeracy score (P5)	1.616	1.633	-0.017	0.036	0.634
Numeracy score (P6)	2.243	2.292	-0.050	0.037	0.173
(C) School-level Characteristics					
Province/Region	3.703	3.538	0.165	0.078	0.033

Notes: **Source**: Author's computation using RwandaEQUIP data *p<0.10; **p<0.05; ***p<0.01 Each cell reports the coefficient from a separate regression of the covariate on a treatment dummy (equal to 1 if above the cutoff) and a linear function of the running variable. Standard errors are robust to heteroskedasticity.

Appendix 2: Descriptive statistics by gender



Notes: **Source**: Author's computation using RwandaEQUIP (Endline) data

Appendix 4a: Observations at the closest mass points

Running Var.	Treatment Status	Below Cutoff	Above Cutoff
-3	Control	31	
-2	Control	28	
-1	Control	28	
0	Control		22
1	Treatment		37
2	Treatment		19
3	Treatment		22

Notes: **Source**: Author's computation using RwandaEQUIP data

Appendix 6: Effect of class size reduction on term 1 (T1) learning outcomes

	(1)	(2)	(3)	(4)	(5)
Outcome Variable	T1 English	T1 Kin-yarwanda	T1 Mathe-matics	T1 SET	T1 SRS
Treatment	0.229***	0.125	0.00380	0.217***	0.192**
	(0.0831)	(0.0840)	(0.0864)	(0.0830)	(0.0867)
Missing Dummies	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Constant	0.139	-1.217***	-0.451	-0.370	-1.156***
	(0.330)	(0.350)	(0.345)	(0.353)	(0.333)
N	3164	3159	3165	3157	3163

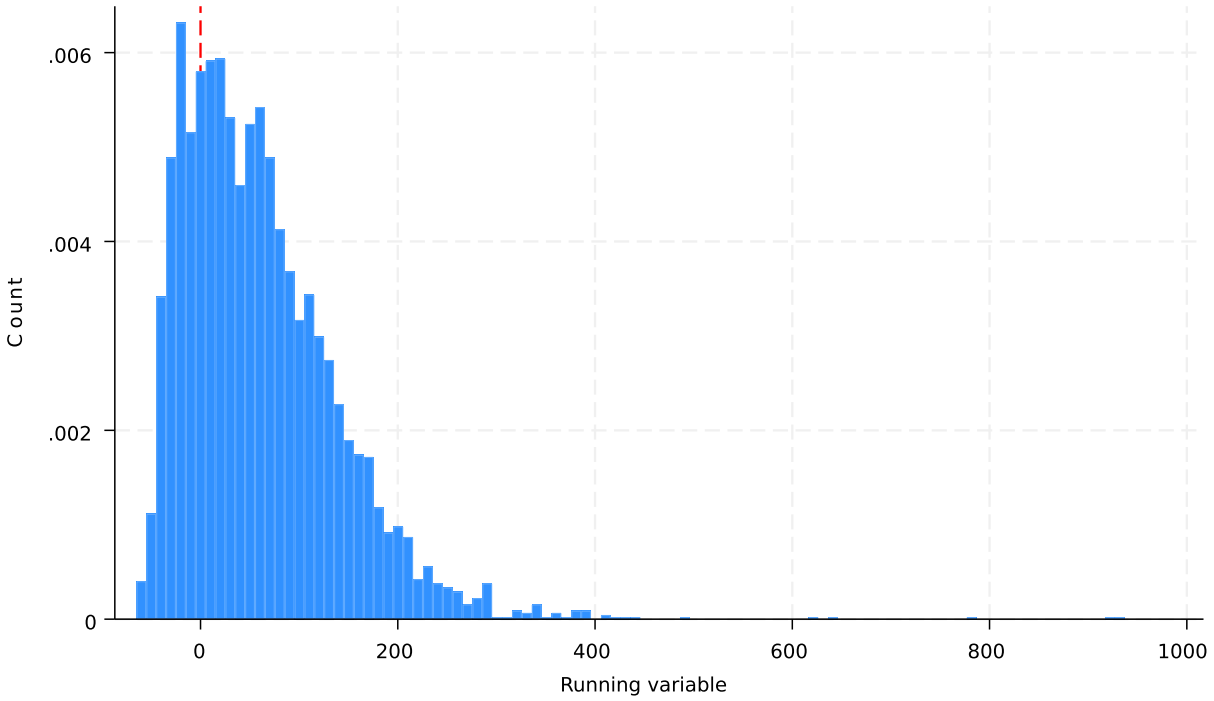
Notes: **Source**: Author's computation using RwandaEQUIP data; Standard error in parenthesis; *p<0.10; **p<0.05; ***p<0.01 All outcome variables are standardized to have mean zero and unit standard deviation.

Appendix 7: Effect of class size reduction on term 3 (T3) learning outcomes

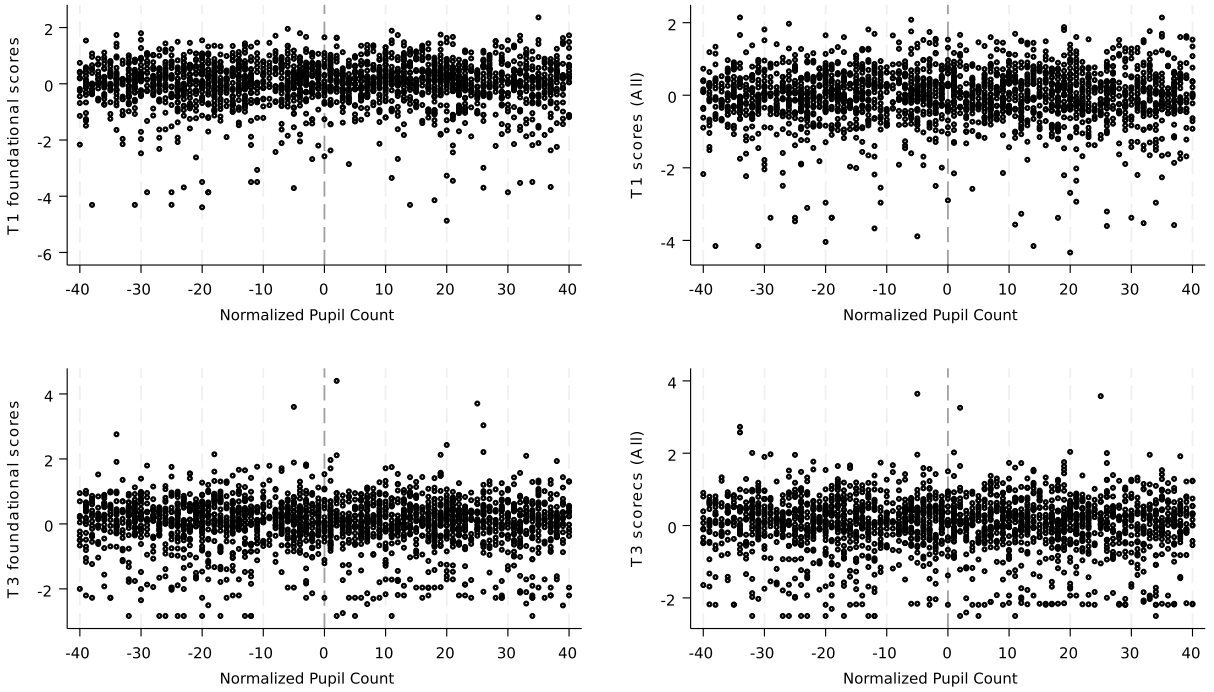
	(1)	(2)	(3)	(4)	(5)
Outcome Variable	T3 English	T3 Kin-yarwanda	T3 Mathe-matics	T3 SET	T3 SRS
Treatment	0.0445	-0.0707	0.0585	0.0416	0.0152
	(0.0770)	(0.0662)	(0.0745)	(0.0803)	(0.0782)
Missing Dummies	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Constant	0.931***	-0.110	0.203	0.634*	0.166
	(0.336)	(0.309)	(0.320)	(0.346)	(0.339)
N	3111	3030	3111	3108	3101

Notes: **Source**: Author's computation using RwandaEQUIP data; Standard error in parenthesis; *p<0.10; **p<0.05; ***p<0.01 All outcome variables are standardized to have mean zero and unit standard deviation.

Appendix 3: Distribution of the running variable.



Appendix 4b: RD Effects on Predetermined covariates based on local randomization (+-2)



Appendix 8: Grouped Term 1 learning outcomes

	(1) T1 Foundational Subjects	(2) T1 - All Subjects	(3) T1 Foundational Subjects	(4) T1 - All Subjects
Treatment	0.0703	0.113*	0.129*	0.161**
	(0.0637)	(0.0605)	(0.0739)	(0.0701)
Missing dummies	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes
Constant	-0.0583	-0.110**	-0.512*	-0.614**
	(0.0545)	(0.0523)	(0.293)	(0.280)
N	3303	3305	3168	3170

Notes: **Source**: Author's computation using RwandaEQUIP data; *p<0.10; **p<0.05; ***p<0.01 All outcome variables are standardized to have mean zero and unit standard deviation. *Foundational* composite scores are constructed as weighted averages of English, Kinyarwanda, and Mathematics exam scores. *All-subject* composite scores additionally include Social and Religious Studies (SRS) and Science and Elementary Technology (SET). Term one (T1) outcomes are based on assessments administered in the first school term, while Term 3 (T3) outcomes correspond to examinations administered at the end of the school session/year in the same subjects.

Appendix 9: Grouped Term 3 learning outcomes

	(1) T3 Foundational Subjects	(2) T3 - All Subjects	(3) T3 Foundational Subjects	(4) T3 - All Subjects
Treatment	-0.0394	-0.0417	0.0145	0.0171
	(0.0542)	(0.0552)	(0.0608)	(0.0624)
Missing dummies	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes
Constant	0.0121	0.0158	0.348	0.358
	(0.0461)	(0.0474)	(0.277)	(0.280)
N	3257	3260	3122	3125

Notes: **Source**: Author's computation using RwandaEQUIP data; *p<0.10; **p<0.05; ***p<0.01 All outcome variables are standardized to have mean zero and unit standard deviation. *Foundational* composite scores are constructed as weighted averages of English, Kinyarwanda, and Mathematics exam scores. *All-subject* composite scores additionally include Social and Religious Studies (SRS) and Science and Elementary Technology (SET). Term one (T1) outcomes are based on assessments administered in the first school term, while Term 3 (T3) outcomes correspond to examinations administered at the end of the school session/year in the same subjects.

Appendix 10: Effect of class size on T1 learning Outcomes by grade level

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Lower Primary					**Upper Primary**				
Outcome Variable	T1 English	T1 Kin-yarwanda	T1 Mathe-matics	T1 SET	T1 SRS	T1 English	T1 Kin-yarwanda	T1 Mathe-matics	T1 SET	T1 SRS
Treatment	0.353**	0.371**	0.0372	0.598***	0.499***	0.194**	0.0377	-0.0303	0.0904	0.0816
	(0.170)	(0.168)	(0.167)	(0.174)	(0.183)	(0.0865)	(0.0900)	(0.0946)	(0.0841)	(0.0871)
Missing Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	-1.088**	-1.360**	-1.439***	-0.0775	-0.891*	0.763*	-1.492***	0.379	-0.678	-1.507***
	(0.534)	(0.528)	(0.514)	(0.519)	(0.541)	(0.437)	(0.521)	(0.507)	(0.507)	(0.488)
N	1321	1319	1323	1320	1323	1843	1840	1842	1837	1840

Notes: **Source**: Author's computation using RwandaEQUIP data; *p<0.10; **p<0.05; ***p<0.01 All outcome variables are standardized to have mean zero and unit standard deviation. Term one (T1) outcomes are based on assessments administered in the first school term, while Term 3 (T3) outcomes correspond to examinations administered at the end of the school session/year in the same subjects. Lower Primary corresponds to grade levels one to three, while upper primary corresponds to grades four to six.

Appendix 11: Effect of class size on T3 learning Outcomes by grade level

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Lower Primary					**Upper Primary**				
Outcome Variable	T3 English	T3 Kin-yarwanda	T3 Mathe-matics	T3 SET	T3 SRS	T3 English	T3 Kin-yarwanda	T3 Mathe-matics	T3 SET	T3 SRS
Treatment	0.344**	0.0836	0.395***	0.232	0.382**	-0.115	-0.159*	-0.0943	-0.0624	-0.169**
	(0.152)	(0.0945)	(0.148)	(0.156)	(0.151)	(0.0837)	(0.0855)	(0.0817)	(0.0848)	(0.0854)
Missing Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	0.760	-0.219	0.371	1.362***	0.471	1.401***	0.279	0.268	0.252	0.182
	(0.488)	(0.385)	(0.464)	(0.515)	(0.489)	(0.487)	(0.487)	(0.460)	(0.492)	(0.467)
N	1316	1237	1317	1315	1316	1795	1793	1794	1793	1785

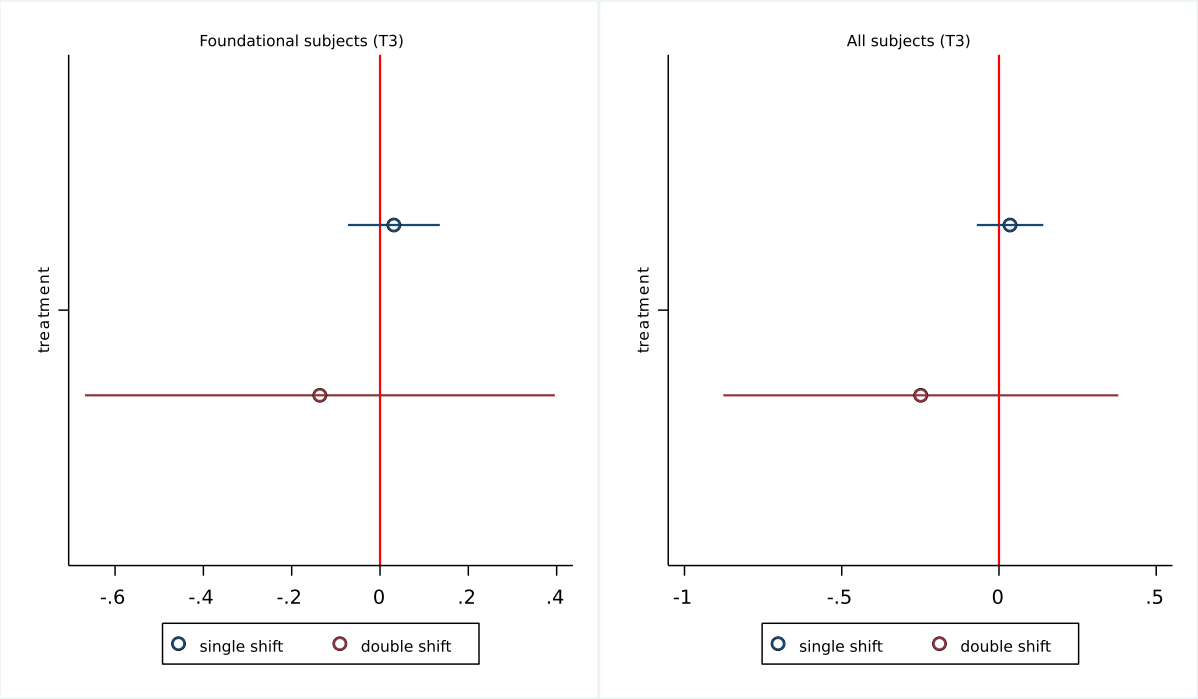
Notes: **Source**: Author's computation using RwandaEQUIP data; *p<0.10; **p<0.05; ***p<0.01 All outcome variables are standardized to have mean zero and unit standard deviation. Term one (T1) outcomes are based on assessments administered in the first school term, while Term 3 (T3) outcomes correspond to examinations administered at the end of the school session/year in the same subjects. Lower Primary corresponds to grade levels one to three, while upper primary corresponds to grades four to six.

Appendix 12: Effect of class size on T1 learning Outcomes by shifting model

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Double Shift					**Single Shift**				
Outcome Variable	T1 English	T1 Kinyarwanda	T1 Mathematics	T1 SET	T1 SRS	T1 English	T1 Kinyarwanda	T1 Mathematics	T1 SET	T1 SRS
Treatment	0.134	0.0113	-0.0545	0.0472	0.0907	0.183**	0.114	-0.0543	0.249***	0.203**
	(0.0950)	(0.0974)	(0.102)	(0.0906)	(0.0948)	(0.0813)	(0.0843)	(0.0857)	(0.0825)	(0.0866)
Missing Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	1.342***	-1.562***	0.697	-0.874	-1.651***	0.561*	-1.080***	-0.0575	-0.0774	-0.668**
	(0.510)	(0.605)	(0.545)	(0.560)	(0.557)	(0.300)	(0.313)	(0.336)	(0.329)	(0.315)
N	347	338	348	347	347	3839	3760	3839	3836	3829

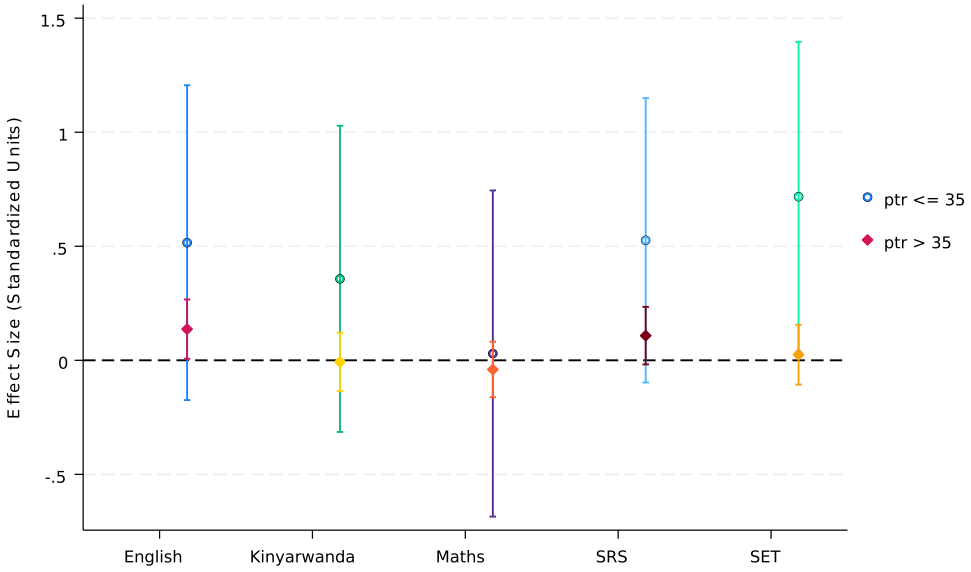
Notes: **Source**: Author's computation using RwandaEQUIP data; *p<0.10; **p<0.05; ***p<0.01 All outcome variables are standardized to have mean zero and unit standard deviation. Term one (T1) outcomes are based on assessments administered in the first school term, while Term 3 (T3) outcomes correspond to examinations administered at the end of the school session/year in the same subjects. Lower Primary corresponds to grade levels one to three, while upper primary corresponds to grades four to six.

Appendix 13: Effect of class size on T3 learning outcomes by shifting model



Notes: **Source**: Author's computation using RwandaEQUIP data All outcome variables are standardized to have mean zero and unit standard deviation. *Foundational* composite scores are constructed as weighted averages of English, Kinyarwanda, and Mathematics exam scores. *All-subject* composite scores additionally include Social and Religious Studies (SRS) and Science and Elementary Technology (SET). Term one (T1) outcomes are based on assessments administered in the first school term, while Term 3 (T3) outcomes correspond to examinations administered at the end of the school session/year in the same subjects.

Appendix 14: Effects of class size reduction by stream size



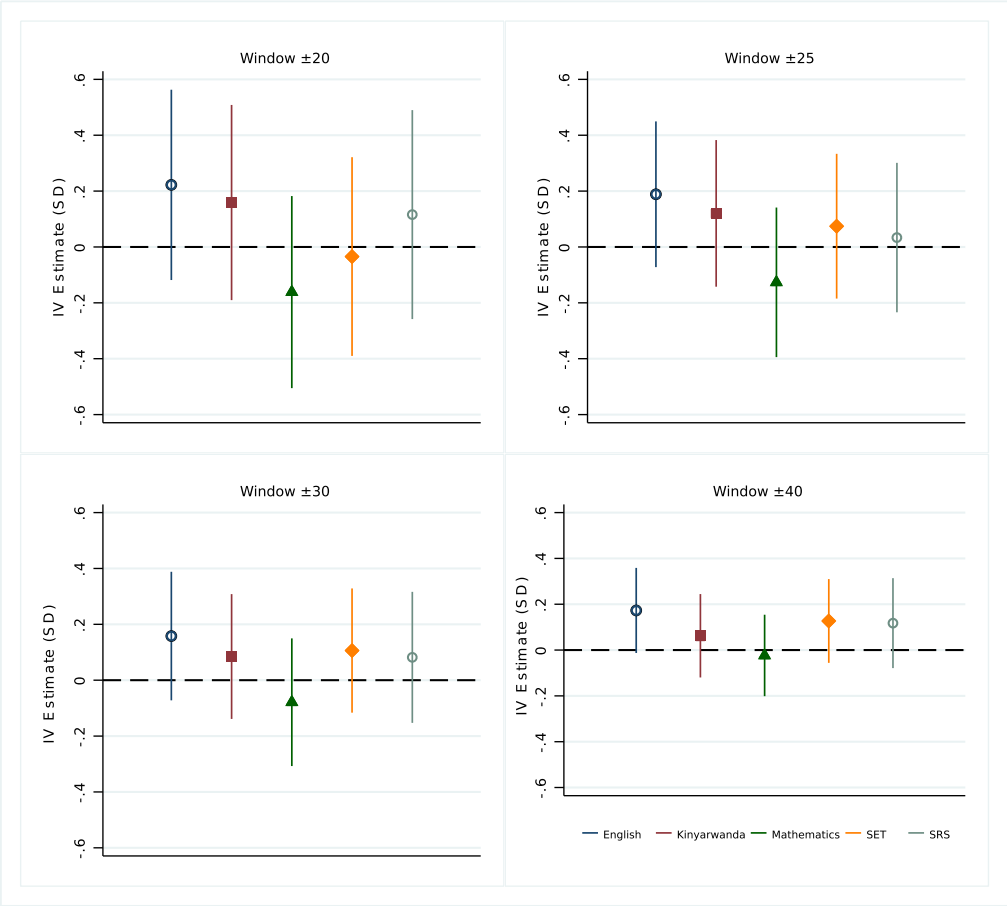
Notes: **Source**: Author's computation using RwandaEQUIP data All outcome variables are standardized to have mean zero and unit standard deviation.

Appendix 20: Robustness: Quadratic Polynomial Specification

	T1					T3				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Outcome Variable	T1 English	T1 Kinyarwanda	T1 Mathematics	T1 SET	T1 SRS	T3 English	T3 Kinyarwanda	T3 Mathematics	T3 SET	T3 SRS
Treatment	-0.162	-0.329	-1.022**	-0.512	-0.644*	-0.123	-0.115	-0.479	-0.298	-0.554
	(0.428)	(0.435)	(0.438)	(0.375)	(0.391)	(0.377)	(0.336)	(0.370)	(0.375)	(0.343)
Missing Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Running var.	0.00461	0.00568	0.0136**	0.00896*	0.0107**	0.00191	0.000647	0.00668	0.00413	0.00743*
	(0.00553)	(0.00567)	(0.00560)	(0.00462)	(0.00478)	(0.00441)	(0.00389)	(0.00429)	(0.00444)	(0.00398)
Squared centred enrolment	-0.0000262	-0.0000417	-0.000119**	-0.0000616	-0.0000855**	-0.00000978	-0.00000744	-0.0000492	-0.0000284	-0.0000647**
	(0.0000468)	(0.0000479)	(0.0000472)	(0.0000384)	(0.0000397)	(0.0000361)	(0.0000317)	(0.0000350)	(0.0000367)	(0.0000323)
Constant	0.397	-0.907**	0.269	0.124	-0.579	1.042**	-0.0780	0.575	0.867**	0.564
	(0.437)	(0.451)	(0.473)	(0.439)	(0.442)	(0.424)	(0.379)	(0.421)	(0.427)	(0.415)
N	3164	3159	3165	3157	3163	3111	3030	3111	3108	3101

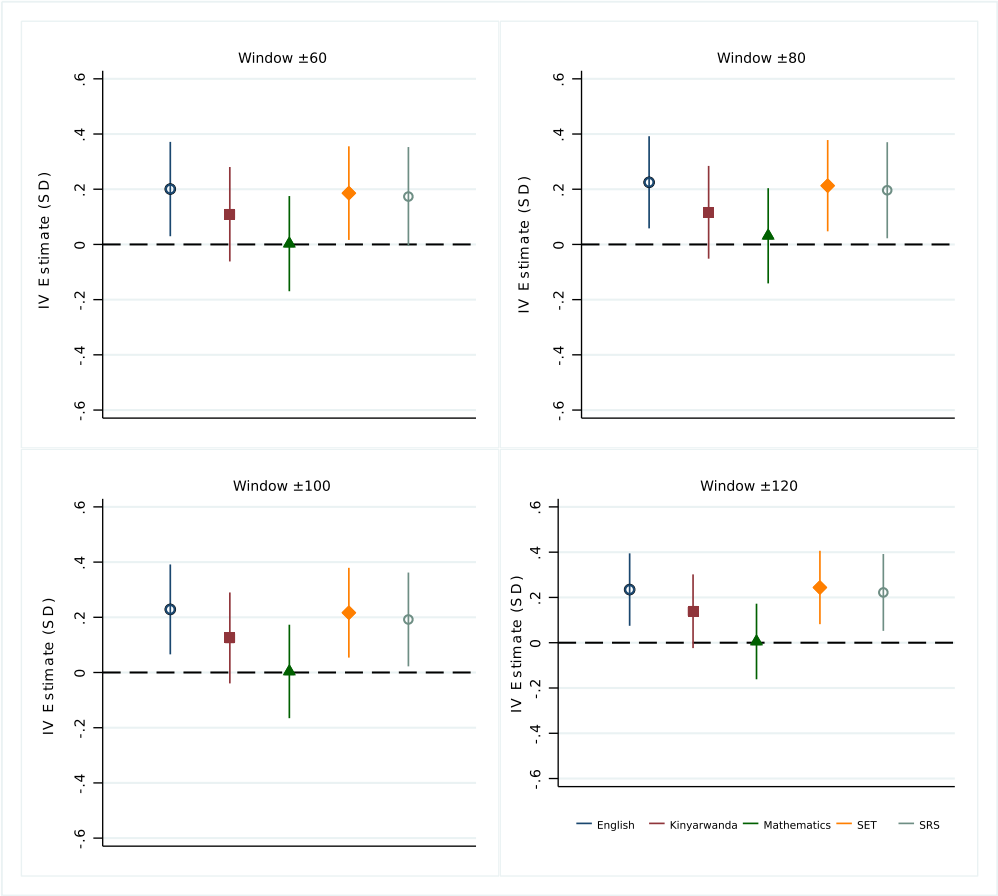
Notes: Source: Author's computation using RwandaEQUIP data

Appendix 15: T1 IV Estimates: Sensitivity to Sample narrow Window (20, 25, 30 and 40)



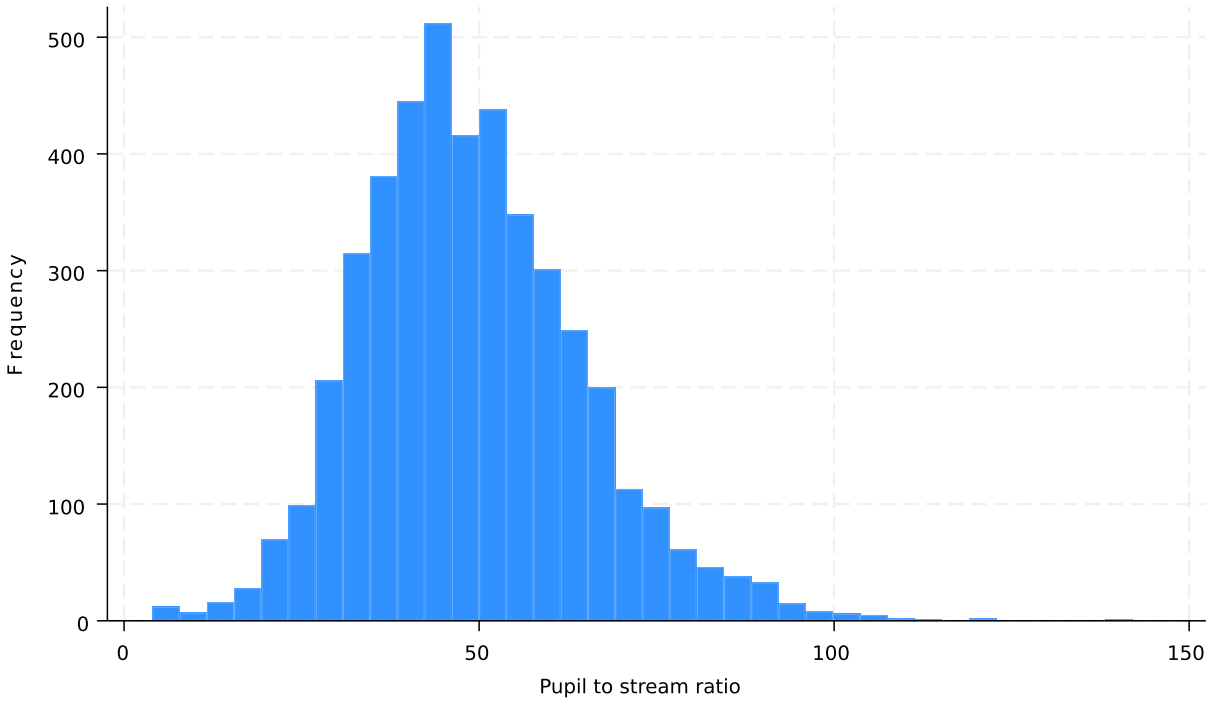
Notes: **Source**: Author's computation using RwandaEQUIP data Extrapolation-based IV following Cattaneo et al. (2024). Point estimates with 95% CIs; clustered standard errors. Declining precision at narrower windows is mechanically expected under this framework. Main specification uses Window ±100.

Appendix 16: T1 IV Estimates: Sensitivity to Sample Window (60, 80, 100 and 120)



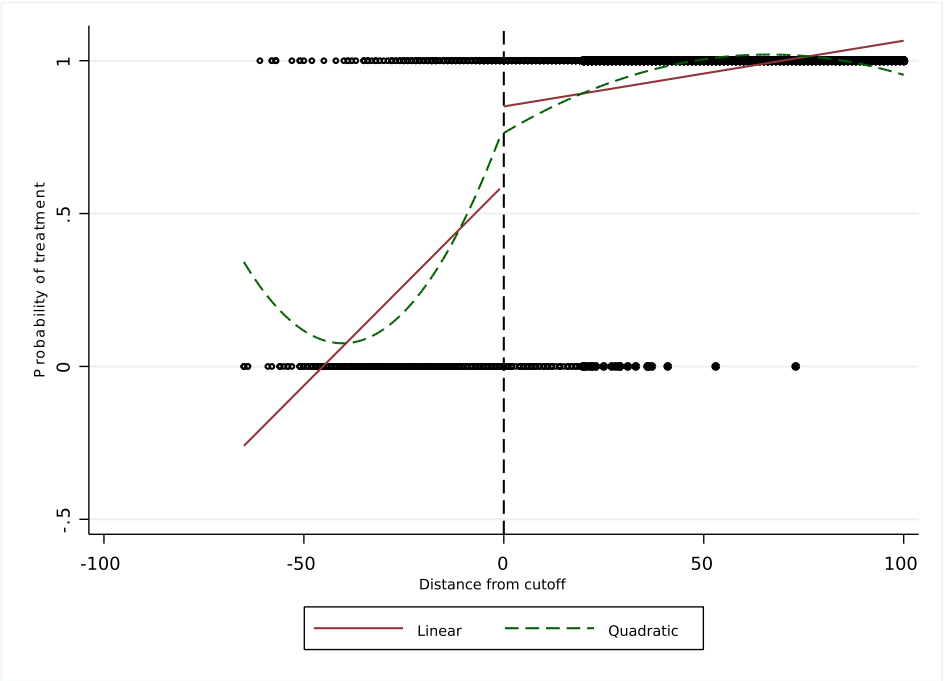
Notes: **Source**: Author's computation using RwandaEQUIP data Extrapolation-based IV following Cattaneo et al. (2024). Point estimates with 95% CIs; clustered standard errors. Consistency across these wider windows supports the conclusion that the ±100 main specification is not driving results through extreme observations.

Appendix 17: Distribution of pupil-stream ratio.



Notes: **Source**: Author's computation using RwandaEQUIP data

Appendix 18: First-stage: Linear Vs Quadratic Fit.



Notes: **Source**: Author's computation using RwandaEQUIP data

Appendix 19: Reduced form – Linear Vs Quadratic



(a) Mathematics T1 (b) SRS T1

Notes: **Source**: Author's computation using RwandaEQUIP data