



HAL
open science

Learning Effects of Conditional Cash Transfers: The Role of Class Size and Composition

Jules Gazeaud, Claire Ricard

► **To cite this version:**

Jules Gazeaud, Claire Ricard. Learning Effects of Conditional Cash Transfers: The Role of Class Size and Composition. *Journal of Development Economics*, 2023, 166, pp.103194. 10.1016/j.jdeveco.2023.103194 . hal-04255219

HAL Id: hal-04255219

<https://uca.hal.science/hal-04255219v1>

Submitted on 23 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Learning Effects of Conditional Cash Transfers: The Role of Class Size and Composition*

Jules Gazeaud Claire Ricard

September 2023

Abstract

A large literature has documented the positive effects of conditional cash transfer programs on school enrollment. However, evidence on their learning impacts is mixed and provides little insights into the mechanisms at play. Using a regression discontinuity analysis of a program implemented at scale in Morocco, we identify a substantial decrease in dropout rates in treated municipalities, which increased both class size and heterogeneity in class composition, and ultimately had negative impacts on children’s test scores. These patterns indicate that, when there is a lack of coordinated supply-side responses, conditional cash transfer programs can constrain learning by putting additional pressure on existing resources. They also highlight the importance of class size and composition for children learning.

JEL Classification: I21, I38, J24, O12, O15.

Keywords: Conditional cash transfers, Learning outcomes, Class size, Program scale-up, Morocco.

* Jules Gazeaud: J-PAL MENA, The American University in Cairo, Egypt & Université Clermont Auvergne, CNRS, IRD, CERDI, France. E-mail: jgazeaud@povertyactionlab.org; Claire Ricard: Université Clermont Auvergne, CNRS, CERDI, France. & FSJES-Aïn Chock, University Hassan II, Morocco. E-mail: claire.ricard@uca.fr. We thank the editor, Jeremy Magruder, and two anonymous referees for their insightful comments. We also thank Abdelilah Abbaia, Martine Audibert, Cátia Batista, Lhcen Belhcen, Simone Bertoli, Isabelle Chort, Bruno Crépon, Zack Devlin-Foltz, Hamidou Diallo, Lydie Drouard, Eric Edmonds, Abdelhaq Elhayani, João Firmino, Catalina Franco, Seema Jayachandran, Abdelhak Kamal, Jason Kerwin, Elhabib Kinani, Abdellatif Komat, Francesca Marchetta, Sandrine Mesplé-Somps, Teresa Molina-Millán, Vasco Molini, Fouzi Mourji, Louise-Pierrette Mvono, Adam Osman, Berk Özler, Olivier Santoni, Olivier Sterck, Holger Strulik, Thomas Thivillon, conference participants at AFES, AFSE, ESPE, GLAD, ICDE, NCDE, JMA, and seminar and workshop participants at ONDH Rabat, Journées Doctorales du Développement, AFD, Moroccan Ministry of Education, and Nova SBE for their helpful comments. We are very grateful to the Moroccan Ministry of Education (in particular Elhabib Kinani) for providing us access to the MASSAR dataset, and to Nourddine Bouhemaid, Mohamed Elarrassi, Abdeljaouad Ezzrari, and Younes Nail for helping us extract and understand the data. Part of this research is based on work conducted while Gazeaud was affiliated with NOVAFRICA, Nova SBE, Universidade NOVA de Lisboa. The authors gratefully acknowledge the support received from the *Agence Nationale de la Recherche* of the French government through the program “*Investissements d’avenir*” (ANR-10-LABX-14-01). This paper was previously circulated under different titles: “*Conditional Cash Transfers and the Learning Crisis: Evidence from Tayssir Scale-up in Morocco*” and “*Conditional Cash Transfers at Scale: A Regression Discontinuity Analysis in Morocco*”. The usual disclaimers apply.

1 Introduction

Empirical studies of conditional cash transfer (CCT) programs have repeatedly found positive impacts on enrollment but mixed effects on learning (Baird et al., 2014; Molina Millán et al., 2019; Garcia and Saavedra, 2022). This evidence is somewhat puzzling as positive impacts on enrollment point to the potential for clear learning gains, and CCT programs have generally been motivated by the ultimate objective of enhancing human capital and breaking the inter-generational transmission of poverty (Fiszbein and Schady, 2009). Existing explanations for the lack of clear learning gains relate to the short-term nature of most evaluations, which may not provide enough time for the effects to materialize. Indeed, a majority of CCT studies rely on phase-in designs in which the intervention expands to the control group after a relatively short evaluation period (e.g., 18 months in the flagship PROGRESA evaluation). While such designs are useful for studying static impacts on enrollment, they make it harder to estimate effects for outcomes such as learning that typically require cumulative investments over an extended period of time (Cahyadi et al., 2020).

In this paper, we propose and evaluate an alternative explanation for the apparent puzzle of why we see large impacts of CCT programs on enrollment but not on learning. Motivated by reports that “schooling is not learning” (Pritchett, 2013; Bold et al., 2017),¹ we focus on supply-side factors and suggest that when no accompanying measures are taken to manage the gains in enrollment, the introduction of a CCT program can deteriorate school quality and thus constrain learning for children who enroll in school regardless of the existence of the program.

We explore this mechanism empirically using a regression discontinuity analysis of a CCT program implemented at scale in Morocco. Morocco’s CCT program, known as *Tayssir*, began operating in 2008 and quickly became the flagship education policy of a government strongly committed to reducing dropout rates. Benhassine et al. (2015) provide experimental evidence that the pilot version of *Tayssir* had substantial positive effects

¹Although most children in low- and middle-income countries are now enrolled in school, less than half are able to read and understand a simple text at the age of 10 (World Bank, 2019). This situation has been characterized as a “learning crisis” by the international community.

on enrollment but null effects on learning. Following this evaluation, Tayssir was scaled-up and targeted all municipalities with a poverty rate above 30% (and all households with children aged 6-15 within these municipalities).² The parents of up to 800,000 children in 434 municipalities received regular cash transfers conditional on their children's school enrollment and regular attendance. Annual transfers were equivalent to between 6% (for children in grades 1 and 2) and 10% (for children in grades 5 and 6) of the average annual spending per capita. Because transfer allocation remained remarkably stable between 2010 and 2018, the scaled-up version of Tayssir offers an ideal setup to study how CCT programs affect learning with enough time for the effects to manifest.³

We exploit the fact that program eligibility was determined on the basis of the poverty rate of each municipality to conduct a fuzzy regression discontinuity analysis. Municipalities were eligible to receive Tayssir if they had a poverty rate above 30% according to the poverty map of 2004. We find that this cut-off is highly predictive of Tayssir allocation. We use novel administrative data from the information system of Morocco's Ministry of Education to study the impacts of the program. This information system provides unique identification for all students in Morocco and is cited by [Abdul-Hamid \(2017\)](#) as an example of a successful education management information system – although to our knowledge it has never been used by academics. To implement this research, we rely on anonymized information on the universe of primary school students in municipalities with poverty rates between 20% and 40% in 2004. This data contains information on over 8,700 schools and 900,000 students annually.

We first assess Tayssir effects on dropout rates and check for possible differences with [Benhassine et al. \(2015\)](#) estimates on the pilot. This exercise is important because impacts may differ as programs are taken to scale ([Banerjee et al., 2017](#); [Muralidharan and Niehaus, 2017](#)). We confirm that the program reduced dropout in all primary school grades. The grade-specific dropout rate decreased by 1.3 p.p. on average, equivalent to 41% of the sample mean. We find a greater decrease for girls: 1.8 p.p. or 50% of the sample mean.

²The pilot evaluation was conducted in municipalities with a poverty rate above 30% and so control schools in the pilot received the intervention in the scale-up.

³In our empirical set-up, treated municipalities have benefited from Tayssir for 5-6 years (or more if they were part of the pilot).

These estimates are in line with those of [Benhassine et al. \(2015\)](#) on the pilot. Overall, the decrease in grade-specific dropout rates translated into an increase in enrollment of about 9 p.p. by the end of grade 6.

Next, we study how this boost in enrollment affected the quality of school supply as proxied by class size and class composition. Without coordinated supply-side responses, the reduction in dropout rates induced by Tayssir may have affected both class size and class composition by retaining lower-ability students, potentially leading to negative effects on learning outcomes through peer effects and less effective teaching practices. Our estimates show that class size in targeted areas increased by 3.6 students by the end of primary school, equivalent to 12% of the sample mean. Additionally, heterogeneity in class composition increased by 0.13 SD by the end of primary school. Looking at effects by grade, we find greater increases in higher grades, suggesting that the reductions in dropout rates accumulated over time and progressively overburdened school resources. We also show that test scores – especially boys’ test scores – are negatively associated with class size and class heterogeneity in this context. Overall, we argue that this evidence is highly suggestive of a negative downstream effect of Tayssir on learning through changes in class size and composition.

To further test whether effects on class size and composition constrained learning, we assess Tayssir impacts on test scores at the primary school graduation exam. Naturally, Tayssir may have impacted test scores through other channels (e.g., increased investments in complementary inputs such as food, school supplies, and parental time), complicating the interpretation of the results. However, we argue that negative reduced-form estimates on learning would still provide empirical support for the importance of the class size and composition channel, as effects from other channels likely run in the opposite direction. In addition, an important empirical challenge for this analysis is that we do not observe the scores of children who drop out before the exam. As a result, estimates using the sample of test-takers are prone to a downward bias.⁴ We rely on two imputation strategies to correct for this bias. First, we use information on the GPA of students prior to dropout

⁴This is because dropouts are negatively selected and because the program caused a reduction in dropout rates (more children with low test scores take the exam in beneficiary municipalities).

and impute test scores corresponding to the likely degree of selection into dropout. Second, because dropouts are negatively selected, we conjecture that they would have obtained test scores on the lower end of the distribution and consider two extreme scenarios: (i) we assume dropouts would have obtained a score of zero; (ii) we assume dropouts would have obtained the median score. Estimates using these imputed values provide lower and upper bounds to Tayssir true effect on test scores.

Consistent with the class size and composition channel, results indicate that Tayssir had negative effects on test scores, especially for boys. Using the first imputation strategy, we find that the program reduced test scores by 0.12 SD for the full sample (p-value<0.05) and by 0.18 SD for the sample of boys (p-value<0.01). Using the second imputation strategy, we estimate effects between -0.02 SD and -0.15 SD for the full sample (significant on the lower-end of the interval) and between -0.09 SD and -0.20 SD for the sample of boys (p-value<0.05 for imputed scores above the 7th percentile). In both cases, effects for girls are smaller and not statistically significant at conventional levels. We also look at impacts on test scores at various points in the distribution, and find suggestive evidence that negative effects are concentrated on the higher end of the distribution. Overall, these patterns indicate that Tayssir constrained learning for children who enroll in school regardless of the existence of the program.

We conclude our analysis by assessing impacts on educational attainment. Despite the negative effects on graduation exam scores, we find that the program increased the probability of student enrollment in secondary school by 4.5 p.p. (equivalent to a 7% increase relative to the sample mean of 63.8%), with stronger effects for girls (7 p.p. or 11% relative to the sample mean of 64.4%).

These results contribute to three strands of the literature. First, they provide novel insights into the learning effects of CCT programs. In a recent review, [Molina Millán et al. \(2019\)](#) document positive effects on grade completion and school attainment,⁵ but mixed evidence on learning. While [Baird et al. \(2011\)](#) and [Duque et al. \(2019\)](#) report evidence

⁵For example, [Cahyadi et al. \(2020\)](#) find a positive effect of a CCT program on high school completion rates in Indonesia – found also in Mexico ([Parker and Vogl, 2018](#); [Araujo and Macours, 2021](#)), Colombia ([Barrera-Osorio et al., 2019](#)), and Honduras for non-indigenous populations ([Molina Millán et al., 2020](#)).

of learning gains, most studies in the literature found null effects (Behrman et al., 2005; Baez and Camacho, 2011; Filmer and Schady, 2014; Benhassine et al., 2015; Baird et al., 2019; Dustan, 2020; Barham et al., 2023).⁶ Our paper adds to this literature by providing estimates from a new context, and, more importantly, by showing that CCT programs can constrain learning by putting additional pressure on existing resources. This mechanism had not been documented in the literature but is expected to be particularly relevant in settings where programs are implemented with no particular measures to manage the enrollment gains. Such ambient effects on learning are unlikely to be captured in studies where transfers are targeted at the individual level and where identification relies on within-school comparisons between treated and untreated children.⁷

Our paper also contributes to the literature on the effects of class size and composition on children learning. Pritchett and Sandefur (2014) argue that the evidence from hundreds of studies on class size effects using non-experimental designs is mixed and highly context specific. Focusing on the most methodologically robust studies in higher-income countries, Bouguen et al. (2017) propose that the learning effects of class size reductions are large – especially in comparison to other educational policies conducted at a large scale and assessed rigorously. In lower-income settings, experimental reductions in class size have typically been implemented in tandem with contract teacher interventions (Muralidharan and Sundararaman, 2011; Duflo et al., 2015; Bold et al., 2018), complicating efforts to isolate their effects (McEwan, 2015). Nonetheless, Duflo et al. (2015) indicate that class size reductions by themselves explained little or none of the learning gains in their context.⁸ In contrast, Duflo et al. (2011) show that reducing heterogeneity in class composition by tracking students into separate classes based on prior achievement benefited children at all levels of the distribution. We add to this literature by documenting the

⁶Table A1 in the Online Appendix summarizes the literature on the learning effects of CCT programs. Unlike previous studies, which evaluated relatively limited exposure to cash transfers or interventions bundling both education and health transfers, in this paper we focus on effects corresponding to a sustained exposure to education transfers.

⁷In the presence of the class size and composition channel documented in this paper, estimates of learning effects using such comparisons are prone to an upward bias.

⁸Regression discontinuity analyses using rule-induced class size reductions found mixed results (Pritchett and Sandefur, 2014), and multiple studies document endogenous sorting or manipulation around the eligibility cutoffs (Urquiola and Verhoogen, 2009; Otsu et al., 2013; Angrist et al., 2019), raising concerns over the internal validity of the estimates.

negative effects on children’s test scores of larger and more heterogeneous classes induced by a CCT program in Morocco – with evidence that negative effects were concentrated in the higher end of the distribution and within the sample of boys.⁹

Finally, we contribute to the emerging literature on the scale-up of development programs by analyzing whether a flagship CCT program in Morocco, whose pilot has been proven particularly successful at increasing enrollment (Benhassine et al., 2015), achieves similar effects once implemented at a national scale. While recent studies suggest that effects may be smaller (Banerjee et al., 2017; Muralidharan and Niehaus, 2017; Bold et al., 2018; Ganimian, 2020; Vivalt, 2020; Araujo et al., 2021; Kerwin and Thornton, 2021), we find that in this setting the effects on enrollment remained very stable despite both the ten-fold increase in the number of beneficiaries and the expansion to numerous new locations. We believe that two aspects of Tayssir help explain this result. First, the pilot and the scale-up were implemented by the same government agency, guaranteeing a high degree of organizational and political buy-in. Second, capacity constraints may be less binding in the case of cash transfer programs, which scale with less requirements in human capital than other interventions such as pedagogical programs. The negative learning effects we estimate contrast with the null effects in Benhassine et al. (2015). However, we do not attribute this difference simply to the one-time scale-up of the program but rather think it emerged because increases in enrollment accumulated over time and progressively overburdened school resources.¹⁰

The rest of the paper is structured as follows: Section 2 presents the background of our study, Section 3 describes the data and the main variables of interest, Section 4 presents the empirical strategy, Section 5 shows the results, and Section 6 concludes.

⁹See Section 5.3 for a discussion of why negative effects may be concentrated within the sample of boys.

¹⁰Randomization in Benhassine et al. (2015) was done at the sector level (with 5 schools per sector on average), therefore, we do not expect their estimates to be prone to the upward bias highlighted above.

2 Background

2.1 Education in Morocco

Public education in Morocco is governed by the Ministry of Education (MENFPESRS or *Ministère de l'Éducation Nationale, du Préscolaire et des Sports*). Private education remains low, despite a sizable expansion during the past decade (especially in urban areas).¹¹ In rural areas, schools are typically organized into clusters consisting of a central, relatively well-resourced school and several smaller satellites. The latter are often one-room schools with one teacher. According to [Soumaya et al. \(2018\)](#), the quality of education in these schools is particularly low. Primary schools include grades 1 through 6 (generally attended by children aged 6 to 12). Lower secondary schools include grades 7 through 9 (generally attended by children aged 13 to 15). Each cycle ends with a final graduation exam. Our article focuses on the exam at the end of primary school, which covers Arabic, Islamic education, French and math. To access secondary education, students are required to get at least 5/10 on this graduation exam. Half of the final grade relies on continuous assessment, 25% on an exam at the end of the first semester, and 25% on a provincial exam at the end of the year. 84% of the students who took this graduation exam in 2014 passed it ([Soumaya et al., 2018](#)).

The average performance of Moroccan students is low compared to international standards. Morocco ranked 73rd out of 78 countries in the latest PISA assessment. According to [Soumaya et al. \(2018\)](#), several factors explain this low performance. First, teachers often lack formal education and pedagogical skills: 40% of fourth-grade teachers have no formal post-secondary education (international average: 3%),¹² and 66% have not participated in any training during the past two years (international average: 16%) ([CSEFRS, 2019](#)). Second, learning is still largely memory-based even though a new curriculum prior-

¹¹The share of students enrolled in private schools doubled between 2007 and 2017, from 8.4% to 16.7% for primary education, and from 4% to 9.3% for lower secondary education (Source: [Atlas territorial de l'enseignement privé 2018](#)). In 2018, 61.4% of the students enrolled in private primary education were living in the most urbanized regions of Casablanca-Settat, Rabat-Salé-Kénitra and Fès-Meknès ([Ministry of Education, 2018](#)).

¹²Until 2007, only a higher secondary school level was required to be a teacher. Since 2007, at least three years of post-secondary education in a Regional Center for Education and Training Professions is required.

itizing skills-based learning was adopted in 2002. Third, instruction is provided in unified modern Arabic whereas the mother tongue of most students is *Darija* (Moroccan Arabic) or *Amazigh* (Moroccan Berber).¹³ Finally, enrollment in early childhood education remains relatively low. During school year 2015-2016, only 43% of children aged 4 to 5 were enrolled in school, of whom 39.5% were enrolled in religious schools.¹⁴

2.2 The Tayssir conditional cash transfer program

Tayssir began operating in 2008 and quickly became the flagship education program of a government strongly committed to reducing school dropout rates. Tayssir provides parents of children aged 6-15 bi-monthly cash transfers conditional on their children's enrollment in and regular attendance of school. The monthly transfer per child is based on their grade: parents receive US\$8 for children in grades 1 and 2, US\$10 for children in grades 3 and 4, US\$13 for children in grades 5 and 6, and US\$18 for children in grades 7 to 9.¹⁵ Transfers are received for all 10 months in which an eligible child attends school regularly (at most 4 absences per month in primary school and 6 absences per month in lower secondary school). Annual transfers for children complying with this condition correspond to between 6% (in grades 1 and 2) and 13% (in grades 7 to 9) of the average annual spending per capita in rural areas.¹⁶ Transfers are restricted to a maximum of three children per household. Parents can withdraw the cash transfers at the local post office. In remote areas, they can also withdraw transfers upon the visit of mobile cashiers. During school years 2015/16 and 2016/17, the program was plagued by liquidity issues, causing significant payment delays (*L'Economiste*, 2017). Parents were supposed to be informed they would receive the transfers for all the months in which eligible children had complied with the conditionality (at most 4 absences per month), irrespective of whether

¹³People typically communicate using Darija (90.9%) or Amazigh (26.8%) but not modern Arabic (*Ministry of Education*, 2018). Amazigh is expected to become a language of instruction by 2030 (*Soumaya et al.*, 2018).

¹⁴Source: *Ministry of Education, 2015-2016 statistical yearbook*.

¹⁵We use the 2008 exchange rate of 7.75 Moroccan dirhams for one US dollar (but this rate has remained relatively stable over time).

¹⁶Authors' calculations using data from the *Enquête Nationale sur la Consommation et les Dépenses des Ménages 2013/14*. Because the program was targeted towards the poorest municipalities, these estimates should be interpreted as lower bounds.

or not the children were still in school at the time of actual payment. This means that payment delays should not drastically affect the incentives of parents to keep children in schools (relative to a scenario with no payment delays).¹⁷

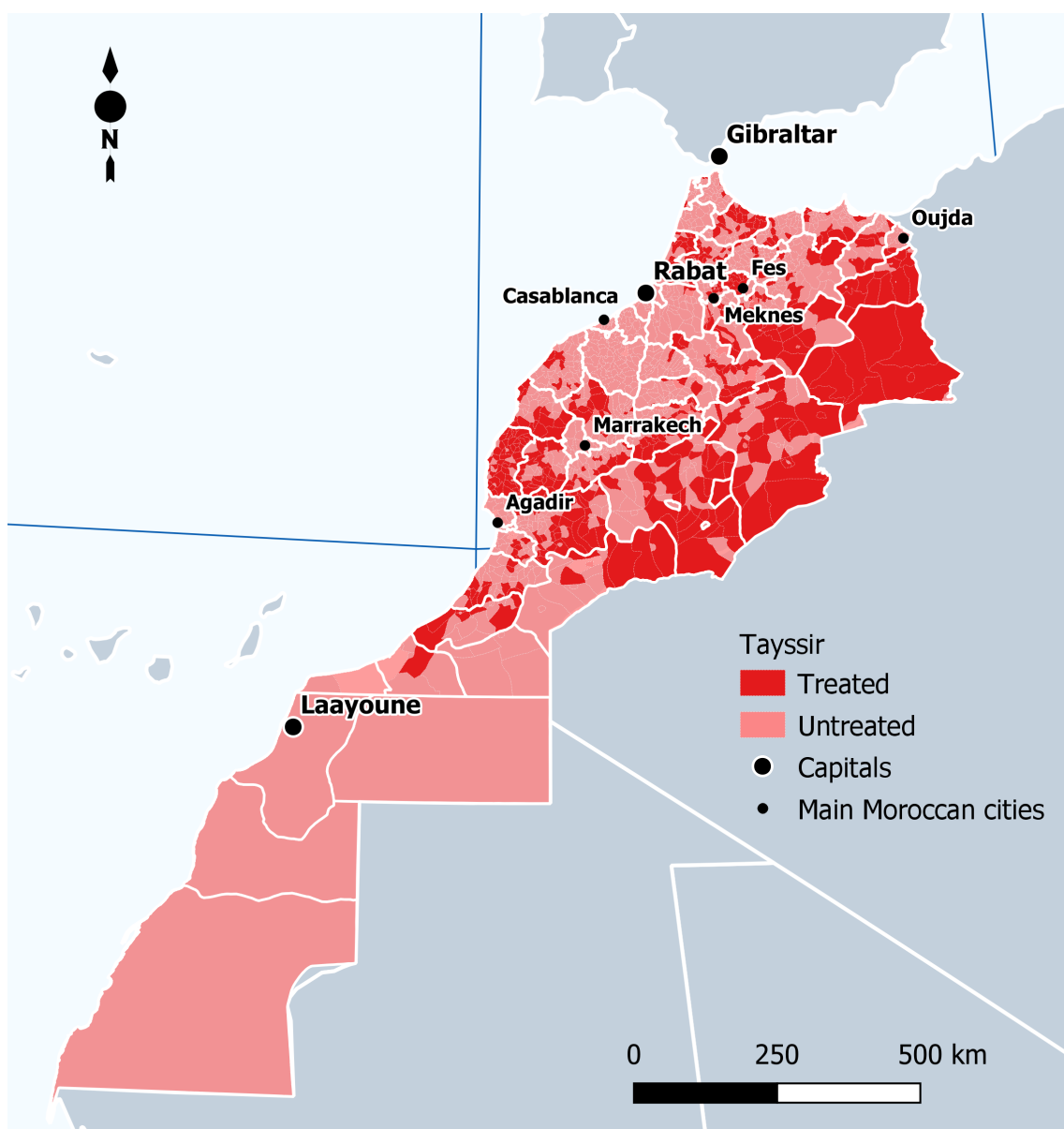
Started as a pilot program in 2008, Tayssir was subsequently rolled-out to 434 poor, rural municipalities in 2010. By the end of 2010, it had already expanded to cover 609,000 children. Municipalities were eligible to receive Tayssir if they had a poverty rate above 30% according to the poverty map of Morocco’s *Haut Commissariat au Plan*.¹⁸ The treatment status of Morocco’s 1,687 municipalities is shown in Figure 1. Initially, all primary school children within treated municipalities were eligible for the transfers. In subsequent years, in order to follow the cohorts of beneficiaries, the program was gradually expanded to lower secondary school (i.e., grades 7 to 9), ultimately reaching around 800,000 children annually. Most recently, during the 2018/19 school year, an important reform enacted a fourfold increase in the budget and the expansion to beneficiaries in new municipalities using proxy means testing (Médias24, 2018). The present paper focuses on the period prior to this reform.

Benhassine et al. (2015) evaluate the impacts of the pilot with a special focus on the importance of conditions and recipient gender. The authors compare households assigned to four variants of the pilot program and to a control group: (i) “labeled” cash transfers to mothers, (ii) “labeled” cash transfers to fathers, (iii) conditional cash transfers to mothers, and (iv) conditional cash transfers to fathers. “Labeled” cash transfers (LCT) were not conditional on school attendance but explicitly labeled as assistance for education costs. The version that was ultimately rolled-out provides CCT to one parent but does not distinguish between father and mother. In what follows, we discuss the effects of the CCT arms but note that results are similar for the LCT arms. Outcomes were measured

¹⁷The amount received stayed the same nominally and inflation was limited over the period (less than 2% annually). Payment delays might still affect our results on dropout if parents had a strong preference for the present or if they believed that in order to ultimately get the money, children had to be kept in school during the whole period with payment delays. In addition, if parents were credit constrained or if they did not behave according to the predictions of the permanent income hypothesis, the payment delays might have muted some of the positive effects of the transfers on learning (e.g., investments in complementary inputs such as food and school supplies).

¹⁸Eligible municipalities were also supposed to have school dropout rates above 8%, but in practice this condition was never enforced (according to our discussions with program officials).

Figure 1: Municipalities receiving Tayssir CCT



Notes: This map depicts the treatment status of the different municipalities. Plain white lines correspond to provincial administrative borders. Dotted white lines correspond to municipal administrative borders. Source: Authors' elaboration using treatment data from the *Haut-Commissariat au Plan*.

two years after the start of the program. The authors find that over the 2-year period dropout rates decreased by 62% in the CCT to fathers arm and by 72% in the CCT to mothers arm. Reductions in dropout were larger for girls (8.3 p.p. in the CCT to fathers arm and 9.5 p.p. in the CCT to mothers – from a base of 12%) than for boys (4.8 p.p. in the CCT to fathers arm and 5.7 p.p. in the CCT to mothers arm – from a base of 8.3%). The authors also administered an ASER arithmetic test to a random subset of

children at endline (i.e., after two years of transfers).¹⁹ They find small, non-significant effects on standardized test scores (+0.04 SD in the CCT to fathers arm; +0.00 SD in the CCT to mothers arm), and no evidence of heterogeneous effects by gender. Our paper assesses whether positive effects on enrollment persisted at scale and whether learning gains emerged for children exposed to 5-6 years of transfers (i.e., during most of their primary education).

3 Data

This study collates data from three sources. First, we rely on the 2004 poverty map built by Morocco’s *Haut-Commissariat au Plan* which was used to determine eligibility to the program. Second, we rely on data from the 2004 and 2014 censuses. We use data from the 2004 census to conduct balance checks and as control variables in the robustness checks. We use data from the 2014 census to test the effect of the program on population growth and on outcomes unlikely to be affected by Tayssir. Third, we use data from the information system of the Ministry of Education, MASSAR, to construct the outcomes of interest. The rest of this section gives more details on MASSAR and on the outcomes of interest.

3.1 The MASSAR database

MASSAR was officially launched in 2013 with the aim of providing unique identification for all students in Morocco. It is cited by [Abdul-Hamid \(2017\)](#) as an example of a successful education management information system. Teachers and school directors enter information on students (age, gender, GPA) into MASSAR using a dedicated website or mobile application, and every student receives a unique ID number, which tracks them throughout the duration of their education in Morocco. MASSAR also includes information at the school level, such as the school municipality, the number of classes per level,

¹⁹ASER is a mathematics test measuring the ability of children to perform basic arithmetic such as recognizing one-digit or two-digit number and performing addition, subtraction, and division. Learning in ASER tests is less cumulative over the years and can show learning gains in the short-term.

the number of teachers per class, and the number of rooms per school.

To conduct our research, Morocco’s Ministry of Education granted us access to a subset of MASSAR database containing anonymized information on primary school students in municipalities with 2004 poverty rates in the [20%, 40%] range.²⁰ ID numbers were used to follow students even if they had migrated or moved to schools outside the 20-40% interval (this is crucial to avoid considering these students as dropouts). Map A1 shows that the set of municipalities with poverty rates inside the 20-40% interval is relatively evenly distributed across Morocco. The only provinces with no sampled municipalities are those in Southern Regions and the greater Casablanca-Rabat area.

According to [Abdul-Hamid \(2017\)](#) and to our own discussions with Moroccan officials, MASSAR covers the universe of students. Table A2 in the Online Appendix further shows that the number of students and schools in our sample has remained remarkably stable over time. This is consistent with the fact that MASSAR was immediately operational at scale. Overall, our sample contains about 8,700 schools and 900,000 students for each year of the 2013-2018 period. Looking at graduation exam scores, we see many missing values in the first two years of MASSAR (40.6% in 2013/14; 31.0% in 2014/15). More reassuringly, the share of missing test scores in the following years was much lower, 0.8% in 2015/16, and 1.0% in 2016/17.²¹ There are several reasons why scores are missing, including student absence on the day of the exam or teachers/directors not entering information in MASSAR.²² To limit the issue of missing scores, we restrict the sample to the years 2015/16 and 2016/17.

²⁰We took into account both practical and methodological considerations while negotiating access to MASSAR data. On the practical side, it would have been challenging for our partners at the Ministry of Education to extract additional data, or data that would be representative of all students. On the methodological side, given our regression discontinuity design, it seemed preferable to sample within a relatively narrow range around the targeting cut-off in order to maximize statistical power ([Cattaneo et al., 2019b](#)).

²¹The prevalence of missing scores is slightly higher for boys than for girls (1.1% vs. 0.8%), and similar in treated and control municipalities (0.9%).

²²The higher prevalence of missing scores in 2013/14 and 2014/15 may reflect potential issues in the first years of MASSAR, including teachers and school directors struggling to get acquainted with the new system, or initial protests from students and teachers ([Médias24, 2014](#)). Qualitative research by [Ennaji \(2018\)](#) highlights that many students protested against MASSAR because they believed it would change the grading system and might therefore increase school failure.

3.2 Outcome variables

Using data from MASSAR, we construct outcome variables related to four main domains: dropout, learning, attainment, and class size and heterogeneity. This subsection describes in details the construction of these variables. Table A3 and Maps A2a to A2d in the Online Appendix provide descriptive statistics.

Dropout We consider a student as a dropout in the school year t if he or she is enrolled in year t but not in year $t+1$. This means that a student who stopped attending school during a given year counts as a dropout only if he or she is not re-enrolled in the following year. Table A3 in the Online Appendix reports the dropout rates by school year and grade. Considering all years and grades, we estimate an average annual dropout rate of 3.3%.²³ The dropout rate remained very stable over the period, consistent with the fact that Morocco did not experience major shocks during any of these years. Looking at grade-specific dropout rates, we see that higher grades are associated with higher dropout rates and that this relationship holds for all years. Taken together, these patterns provide some reassurance about the consistency of MASSAR over time. If anything, data from 2013/14 produce higher dropout rates, especially in the initial grades. Because dropouts in 2013/14 are derived using enrollment data from MASSAR's first year of implementation, this could be a symptom of an imperfect roll-out. As mentioned in Section 3.1, in order to limit data quality concerns and missing scores, we do not use data from 2013/14 and 2014/15 to investigate the effects of Tayssir. Our analyses are restricted to the years 2015/16 and 2016/17, for which data produce patterns of dropout rates that are hardly distinguishable from one another. Overall, we are confident that this data capture an accurate account of actual outcomes. One limitation of the dataset, however, is that we do not observe students in grade 6 during their whole primary education and therefore do not know the exact number of students who dropped out prior to grade 6.²⁴ In addition,

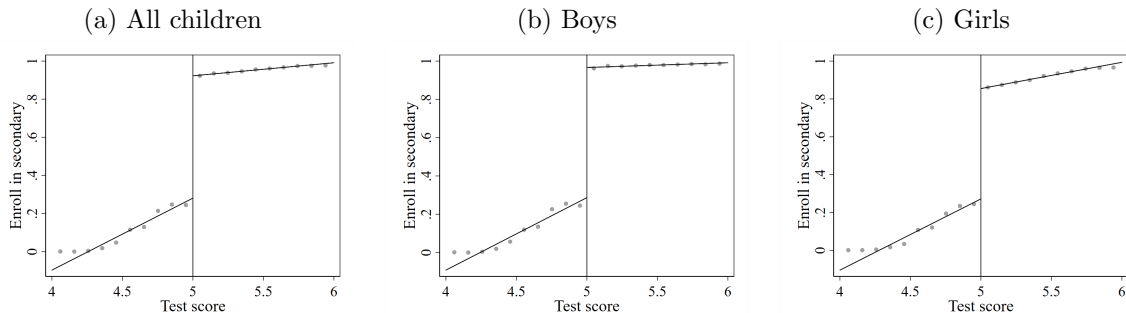
²³This corresponds to the annual dropout rate in primary school. In order to approximate the overall dropout rate in primary school, one should multiply this rate by the average number of years spent in primary school before graduation.

²⁴This presents a challenge for the imputation strategy we use in Section 5.3 to address selection issues. See footnote 50 for more information on how we overcome this issue. In short, we proxy the total number of dropouts prior to grade 6 using information on cross-sectional dropouts throughout grades 1 to 6.

we do not observe children who never enrolled however, according to UNICEF (2015), in 2012, among the 2% of all children aged 6-11 who were not enrolled in school, only 14.4% had never been enrolled in school, suggesting that only 0.28% of children have never been enrolled in school.

Learning We measure learning using the score obtained at a high-stakes exam, called *Certificat d'études primaires*, which is administered to all children in grade 6. This exam assesses student knowledge in four areas: Arabic, Islamic education, French, and mathematics. We focus on the part of the exam administered at the end of the year because it is the only part that is common at the provincial level and therefore allows for comparisons between students from different schools and municipalities.²⁵ In each province, questions to assess student learning are determined by a local committee composed of school inspectors and teachers designated by the regional authorities. Tests are graded anonymously and centrally by provincial teachers.

Figure 2: Selectivity of grade 6 examination



Notes: Each figure represents the probability of a student enrolling in secondary school in year $t+1$ depending on the score obtained at the grade 6 final exam in year t . Authors' elaboration using data from MASSAR. Sample includes grade 6 students in municipalities with poverty rates between 20% and 40%.

For the analysis, because the content of the tests varies across provinces, we divide the test score of each student by the mean test score of the control group of his or her province.

To ease interpretation, we then standardize test scores using the mean and the SD of the

²⁵Note, however, that the bulk of the exam (75%) is composed of the grades obtained during the school year through continuous assessment. Unfortunately, this part of the exam is poorly suited to estimate the effects of Tayssir because of the absence of standardization. Different teachers are likely to use different tests and to grade differently. Moreover, there is no variation in treatment conditions within the group of students assigned to a teacher.

full sample.²⁶ In theory, students should get an overall score of at least 5 (out of 10) to access secondary school. To explore the extent to which this rule is enforced in practice, we relate the overall score to enrollment in secondary school the following year (Figures 2a to 2c). While the figures indicate a clear discontinuity in access to secondary schools for scores just above or below 5, it appears that a significant share of students with scores below 5 do access secondary school (possibly due to rounding of 4.5 to a pass). In addition, despite scores above 5, some students (girls in particular) do not access secondary school – a likely consequence of the high grade 6 dropout rates observed for girls (Table A4 in the Online Appendix).

Educational attainment We measure educational attainment using two dummies indicating (i) graduation from primary school, and (ii) enrollment in secondary school. MASSAR includes no explicit data on graduation; however, as shown in Figure 2, the score obtained at the grade 6 exam provides a relatively good proxy. We define graduation in a theoretical sense, that is, for students with test scores above 5. Enrollment in secondary school is derived from the changes in MASSAR over time, by observing whether students in grade 6 are enrolled in secondary school the following year.

Class size and heterogeneity We derive class size using data at the school level on the number of students in each grade. We divide the number of students in each grade by the number of classes. To derive a measure of class heterogeneity, we also rely on GPA data and derive the standard deviation of GPAs within each class. We do not have information on class size for the 15,243 students (or less than 0.83% of the sample) who moved to municipalities outside the sample. In addition, we do not have information on class heterogeneity for the 31,368 students (or less than 1.71%) whose classes have missing data on student GPAs.

²⁶Results are similar if we standardize relative to the mean and the SD of the control group.

4 Empirical strategy

4.1 Regression discontinuity design

As mentioned in Section 2.2, only children in municipalities with poverty rates above 30% in 2004 were eligible for Tayssir. We use a regression discontinuity (RD) design to compare students in municipalities above and below this eligibility threshold.

Figure 3a illustrates the discontinuity in treatment allocation in the neighborhood of the cut-off using treatment data and poverty rates at the municipality level. The 30% cut-off is highly predictive of Tayssir allocation, although the discontinuity is not perfect: some municipalities with scores below the cut-off receive Tayssir and some municipalities with scores above the cut-off do not receive Tayssir. This means that our RD design is fuzzy. Map A3 distinguishes "complying" municipalities from the "never takers" (i.e., eligible municipalities not receiving Tayssir) and the "always takers" (i.e., non-eligible municipalities receiving Tayssir). Compliance is somewhat lower in the north of the country (in the Rif region) where the share of always takers is particularly high.

We use the following simple linear model to estimate treatment effects:

$$Y_{it} = \lambda_0 + \lambda_1 \text{Tayssir}_i + \lambda_2 Z_i + \lambda_3 \text{Tayssir}_i \times Z_i + \varepsilon_{it} \quad (1)$$

where Y_{it} is the outcome of child i in school year t , Tayssir_i is an indicator variable for whether student i is in a treated municipality, Z_i is the continuous poverty rate of the student's municipality, and ε_{it} is the error term. Because the cut-off is fuzzy, we instrument Tayssir_i with Above_i , which is an indicator for being above the 30% eligibility cut-off. The parameter of interest, λ_1 , corresponds to Tayssir's local average treatment effect (LATE). It can be interpreted as the effect of receiving Tayssir in complying municipalities near the cut-off.

Our baseline strategy for estimating Equation (1) makes use of the full sample of students in primary school described in Section 3.1. We do not rely on Imbens and Kalyanaraman (2012)'s or Calonico et al. (2014)'s data-driven bandwidth selection methods because our sample is already restricted to students in municipalities close to the

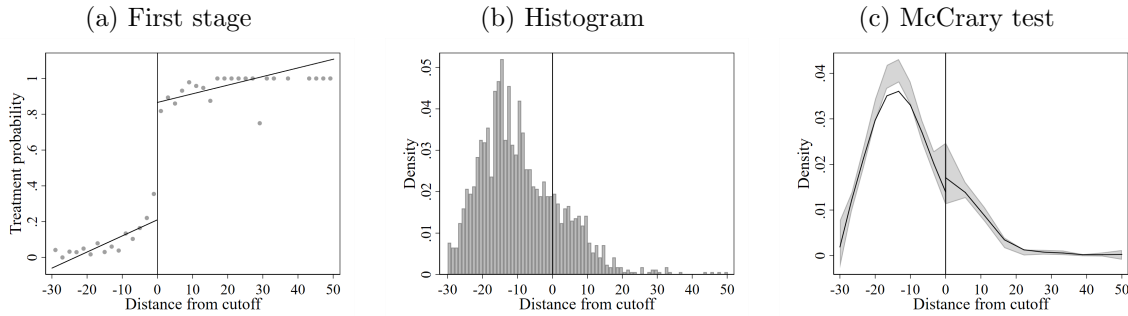
cut-off, and because such methods have poor behaviors when the underlying bias is close to zero (Cattaneo et al., 2019a).²⁷ We use a simple linear model because higher order polynomials can lead to erratic behavior of the estimator at the cut-off (Cattaneo and Vazquez-Bare, 2016; Gelman and Imbens, 2019). Finally, in line with Cattaneo et al. (2019a), we weight observations using a triangular kernel function and account for within-cluster data dependence by clustering standard errors at the municipality year level using the nearest-neighbor estimator. We show that results are largely robust to (i) weighting observations using a uniform kernel function, (ii) adding municipality controls accounting for baseline characteristics (illiteracy rate, net school enrollment rate, share of households with access to electricity, share of households with running water, share of households with waste-water, total population), (iii) using a quadratic polynomial approximation, (iv) narrowing the bandwidth to 7.5 p.p. on either side of the cut-off, and (v) narrowing the bandwidth to 5 p.p. on either side of the cut-off (see Tables A10, A11, A12, A13, and A14 in the Online Appendix). Results are sensitive to using a quadratic polynomial approximation or a smaller bandwidth, however, this is not too concerning because these specifications are noisier and can lead to erratic behaviors of the estimator at the cut-off.

4.2 Identifying assumptions

A key assumption for RD designs to provide unbiased causal estimates is that of no sorting around the eligibility threshold. In our setting, precise manipulation of municipalities around the threshold is unlikely because poverty rates were determined prior to Tayssir

²⁷McKenzie (2016) shows that the optimal bandwidth selected by Calonico et al. (2014)’s method depends on whether the sample is restricted to observations near the cut-off, “*even if [one is] just throwing away units that were outside the initial optimal bandwidth*”. McKenzie (2016) uses Cattaneo et al. (2015) data on U.S. Senate elections to estimate the optimal bandwidths for restricted and unrestricted samples. The research question is whether the incumbent status of previously elected legislators translated into an electoral advantage. McKenzie (2016) finds that using the default settings and the full range of the data, Calonico et al. (2014)’s method yields an optimal bandwidth of 12.45 vote share on either side of the cut-off. In contrast, using the restricted [-20, 20] range, Calonico et al. (2014)’s method yields an optimal bandwidth of only 3.63 vote share on either side of the cut-off. McKenzie (2016) concludes that “[*he does*] not think [*one*] necessarily wants to choose a narrow range, and then choose an optimal bandwidth within that range”. As mentioned in Section 3.1, for practical reasons, our sample is already restricted to students in municipalities with poverty rates in the [-10, 10] range. For reference, we report results using Calonico et al. (2014)’s MSE-optimal bandwidth in the robustness checks (Tables A10, A11, A12, A13, and A14 in the Online Appendix). However, these estimates must be interpreted with caution due to the shortcoming of using an optimal bandwidth when samples are already restricted around the cut-off. In our setting, this yields highly imprecise estimates because small bandwidths are selected.

Figure 3: First stage and distribution of the running variable



Notes: Figure 3a shows a linear prediction of Tayssir CCT receipt on municipality poverty scores, run separately on each side of the normalized eligibility cut-off. Municipalities are sorted into bins of 2 percentage points of width. Figure 3b displays the density of the running variable across municipalities. Figure 3c features a graphical representation of the [McCrary \(2008\)](#) test of no discontinuity in the running variable at the cut-off (p-value=0.33). The vertical lines indicate the normalized eligibility cut-off. Source: Authors' calculation using poverty data from Morocco's *Haut-Commissariat au Plan* and treatment data from MASSAR.

implementation (and prior to the announcement of the eligibility threshold). To confirm this, we report the density of municipalities across poverty rates (Figure 3b) and test for the presence of a discontinuity at the cut-off using [McCrary \(2008\)](#)'s test (Figure 3c). Overall, we see no jump in the density at the cut-off and the McCrary test does not reject the null of no discontinuity (p-value = 0.33). Using population data from the 2004 and 2014 censuses, we also check for the presence of individual sorting. Individual sorting could correspond to households migrating to eligible municipalities to benefit from the transfers.²⁸ We find no increase in population growth in beneficiary areas between 2004 and 2014 (Table A6 in the Online Appendix).

Another crucial identifying assumption is that students in municipalities that just qualify for Tayssir (i.e., municipalities whose poverty rate is just above 30%) are sufficiently similar to those in municipalities that just miss out on Tayssir (i.e., municipalities whose poverty rate is just below 30%). In Table 1, we test for the presence of discontinuities around the 30% poverty rate threshold using data from the 2004 census and information on predetermined variables at the municipality level.²⁹ Although the estimates are quite noisy, reassuringly, the coefficients are small and non-significant at conventional levels.

²⁸Relatedly, an emerging literature documents the effects of cash transfers on emigration (see e.g., [Hidrobo et al., 2022](#); [Gazeaud et al., 2023](#)), although the magnitude of transfers is typically greater than in Tayssir.

²⁹We lack pre-program data at the student level because MASSAR was only launched after the program in 2013.

We also use data from the 2014 census to explore effects on outcomes that are unlikely to be affected by Tayssir but could have been affected by other programs placed on the basis of the poverty rate. Results are reported in Table A5 and show no evidence of significant imbalances.

Table 1: Pre-intervention balance

	(1)	(2)	(3)	(4)	(5)	(6)
	Illiteracy rate	Net school enrollment rate (7-12 y.o.)	Share of households with access to electricity	Share of households with access to running water	Share of households with access to waste-water	Total population (in thousands)
LATE	-1.078 (0.948)	-1.653 (1.577)	-3.967 (3.294)	2.832 (3.235)	0.086 (0.898)	0.419 (0.752)
Sample mean	63.271	68.940	36.356	18.413	2.199	9.693
SD	9.958	14.552	28.033	25.564	9.280	6.677
N	645	645	645	645	645	645

Notes: The table reports the results of local linear regressions using a bandwidth of 10 percentage points on each side of the cut-off. In each regression, the other variables are used as controls (e.g. in column 1 the variables in columns 2 to 6 are used as controls). Observations are weighted using a triangular kernel function. Robust standard errors in parenthesis. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Source: Authors' calculation using data from the 2004 Moroccan census and Morocco's *Haut-Commissariat au Plan*.

Finally, to interpret λ_1 in Equation 1 as the causal effect of Tayssir on educational outcomes, one needs to assume that no other programs determined eligibility using the 2004 poverty rate and had effects on educational outcomes. The only program we are aware of using the 2004 municipal poverty rate is the *Initiative Nationale pour le Développement Humain*.³⁰ This program was implemented between 2006 and 2010 and had four main components: (i) a rural component, (ii) an urban component, (iii) a vulnerability component, and (iv) a “cross cutting” component. Only the rural component of the program used the 2004 municipal poverty rate to determine eligibility. This component targeted 348 municipalities with poverty rates over 30% and 55 municipalities with poverty rates between 22% and 30% – which were nominated by provincial committees.³¹ It was designed as a community-driven development (CDD) intervention, whereby municipalities received a block grant and projects were identified and prioritized through a participatory planning

³⁰Morocco's nationwide medical assistance program, RAMed, has been targeted at the household level using proxy means testing (Cottin, 2019). Other demand side education programs do not depend on the municipal poverty rate. In particular, the initiative *Un million de cartables* – which provides school bags and other school supplies – is universal. Boarding schools and school feeding programs depend on student characteristics (remoteness, number of siblings, academic performance, etc.).

³¹The rural component was extended after 2010 to cover 299 new municipalities with a poverty rate above 14% (in addition to the 403 municipalities initially covered).

process. Each municipality received 1 million Moroccan dirhams (or US\$129,000) per year (World Bank, 2006).³² Funded projects could include: investing to improve access to basic infrastructures and services,³³ supporting local NGOs and civil society groups (local women or youth community organizations), developing local capacity (training of municipal employees), and supporting local businesses (through business training). This CDD intervention could threaten our identification strategy if it had long-run effects on educational outcomes. While our data do not allow us to fully rule out this concern, we believe that it remains limited for a number of reasons. First, evidence on the impact of CDD programs is mixed at best (Casey, 2018), especially in the long-run (Mvukiyehe and Van der Windt, 2020).³⁴ Second, even if the Moroccan CDD program had impacts in the long-run, given the activities that were undertaken, it seems unlikely to us that the first-order effects would be on educational outcomes. Finally, even if the program had long-run effects on educational outcomes, it would likely attenuate the negative effect on learning as well as the positive effect on class size we estimate. To the best of our knowledge, and according to the experts we consulted, no other interventions have used the same poverty rate or cutoff as Tayssir.

5 Results

We start by presenting program effects on school dropout rates. We then assess effects on class size and composition. Next, we present program effects on test scores, discussing potential issues arising from differential dropout rates and our proposed solution. Finally, we report effects on educational attainment. We are interested in the effect of receiving

³²For reference, municipalities targeted by Tayssir received US\$187,000 per year on average over the period we study.

³³Under this item, the only activity directly related to schooling is the rehabilitation of schools: “Projects [...] can be related to the rehabilitation, extension or improvement of road and street infrastructure, electricity or water supply, sanitation or solid waste services. They can also contribute to rehabilitating schools, local health centers or community centers. Finally, they can include limited construction activities for public gardens, s or other public facilities.” (See Annex 4A II p56 of World Bank, 2006).

³⁴Evidence on the short-run effects of the Moroccan CDD program is also mixed. El-Kogali et al. (2016) found transitory impacts on some economic outcomes but no impacts on early childhood development. Nguyen and Rieger (2017) found positive impacts on people’s contributions to a public goods game but no effects on contributions to both a dictator game and a trust game.

Tayssir payments and therefore focus the discussion on LATE estimates (λ_1 in Eq. 1).

5.1 Impacts on school dropout

Table 2 shows the effects of Tayssir on grade-specific dropout rates. Columns 1 to 6 report the results for each school grade and column 7 displays the results for all grades. Panel A shows estimates for all children, while panels B and C show estimates for boys and girls, respectively. Figures 4a to 4c show reduced-form estimates.

Table 2: Fuzzy RD analysis of dropout effects around Tayssir eligibility threshold

	Dep Var: Drop-out						
	(1) Grade 1	(2) Grade 2	(3) Grade 3	(4) Grade 4	(5) Grade 5	(6) Grade 6	(7) All Grades
Panel A. All students							
LATE	0.000 (0.002)	-0.006*** (0.002)	-0.010*** (0.003)	-0.011*** (0.004)	-0.018*** (0.005)	-0.031* (0.017)	-0.013*** (0.004)
Mean drop-out	0.005	0.011	0.013	0.022	0.040	0.101	0.032
Observations	333,225	305,730	304,358	293,988	286,277	310,821	1,834,399
Panel B. Boys							
LATE	0.001 (0.002)	-0.006** (0.003)	-0.008** (0.003)	-0.011*** (0.004)	-0.017*** (0.005)	-0.014 (0.015)	-0.010*** (0.004)
Mean drop-out	0.006	0.011	0.013	0.021	0.036	0.084	0.028
Observations	174,878	160,229	160,590	155,123	151,247	164,642	966,709
Panel C. Girls							
LATE	-0.001 (0.002)	-0.007** (0.003)	-0.011*** (0.004)	-0.011** (0.005)	-0.020*** (0.007)	-0.049** (0.021)	-0.018*** (0.004)
Mean drop-out	0.005	0.011	0.014	0.024	0.043	0.120	0.036
Observations	158,347	145,501	143,768	138,865	135,030	146,179	867,690

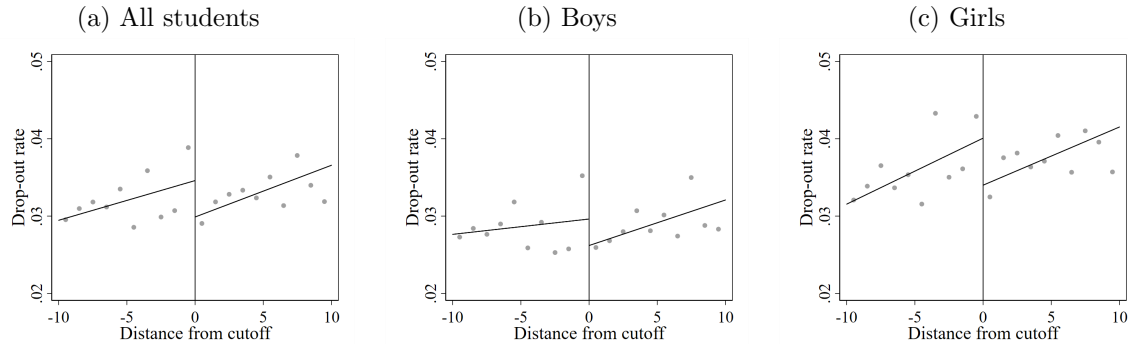
Notes: This table reports local average treatment effects of Tayssir cash transfers (Eq. 1 in the text). The dependent variable is the grade-specific dropout rate, that is a dummy equal to one if a student drops out of school in Grade x (with $x = [1, 6]$). Columns 1 to 6 report results for grades 1 to 6, respectively. Column 7 reports results for all grades. The unit of observation is a student-year for the school years 2015/16 and 2016/17. Sample restricted to students in primary schools in municipalities with 2004 poverty rates in the [20%, 40%] range. Observations are weighted using a triangular kernel function. Robust standard errors in parenthesis are clustered at the municipality year level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Source: Authors' calculation using data from MASSAR and Morocco's *Haut-Commissariat au Plan*.

The estimates reveal that the program led to a significant reduction in dropout rates. Point estimates imply that the average grade-specific dropout rate in primary school decreased by 1.3 p.p. (a 41% reduction relative to the average dropout rate in the sample).³⁵ As expected, we see a positive correlation between school dropout and the poverty rates on each side of the eligibility cut-off (Figure 4a). Looking at effects by grade, we find small and non-significant effects in grade 1 but negative and statistically significant effects in

³⁵To approximate the cumulative effect on dropout rates throughout primary school one needs to multiply -1.3 p.p. by the average number of years spent in primary school (about 7 years on average given that there are 6 grades and that the annual repetition rate is 15%).

higher grades. Higher grades are consistently associated with larger absolute reductions in dropout rates: from -0.6 p.p. in grade 2, to -1 p.p. in grades 3 and 4, to -1.8 p.p. in grade 5, and to -3.1 p.p. in grade 6. Regressions by gender show that negative effects on dropout are stronger for girls. Girls benefiting from Tayssir are on average -1.8 p.p. less likely to drop out of school (a 50% reduction compared to the sample mean). Boys benefiting from Tayssir are on average -1 p.p (or 36%) less likely to drop out of school. The largest absolute reductions are found for boys and girls in the highest grades.

Figure 4: Reduced form effect on school dropout rates



Notes: Figures show a linear prediction plot of dropout rates on municipality poverty rates, run separately on each side of the normalized eligibility cut-off. The unit of observation is a student year for the school years 2015/16 and 2016/17. Sample includes students in primary school in municipalities with 2004 poverty rates in the [20%, 40%] range. Observations are sorted into bins of one percentage point width and weighted using a triangular kernel function. The vertical lines indicate the normalized eligibility cut-off. Source: Authors' calculation using poverty rate from Morocco's *Haut-Commissariat au Plan* and dropout data from MASSAR.

These results appear very consistent with those of [Benhassine et al. \(2015\)](#) on the pilot. Two years after the start of the pilot, they found a reduction in dropout rates between 62% and 72% depending on the parent receiving the transfers (equivalent to an annual reduction between 31% and 36%).³⁶ We should be cautious in comparing their results with ours because of the different empirical strategies. [Benhassine et al. \(2015\)](#) use a randomized controlled trial while in this paper we rely on a regression discontinuity design. This means that estimates in [Benhassine et al. \(2015\)](#) correspond to average treatment effects whereas ours correspond to local average treatment effects (which apply to children in municipalities near the eligibility cut-off but not necessarily in municipalities further away). That said, the 41% reduction we estimate is of a similar sign and magnitude,

³⁶Some children may have dropped out of school temporarily, so these should be interpreted as a lower bound estimates.

and the heterogeneous patterns with respect to gender are fully consistent. These similar results are remarkable given the expansion of the program to numerous new locations and the ten-fold increase in the number of beneficiaries. If anything, they provide some reassurance regarding the ability of pilot evaluations to produce insightful estimates of at-scale impacts – at least for interventions which are not too demanding in terms of human capital.

5.2 Impacts on class size and composition

In Table 3, we study whether this boost in enrollment affected the quality of school supply as proxied by class size and heterogeneity in class composition. Figures A6a and A6b present reduced-form estimates. The results show that Tayssir had a positive and significant impact on class size and class heterogeneity. Class size increased by as much as 3.6 students in grade 6 (panel A, column 6), which corresponds to 12% of the average class size in the sample (30.7 students). This effect is remarkably similar in magnitude to the 9 p.p. increase in enrollment by the end of grade 6 induced by the reduction in dropout rates.³⁷ Looking at effects by grade, we find that higher grades are generally associated with greater increases, with the exception of grade 1. Large effects in grade 1 likely reflect the fact that children in targeted municipalities started school earlier (possibly to benefit from the transfers) and repeated grade 1 more often.³⁸ We also find positive effects of Tayssir on class heterogeneity (Panel B), as measured by the standard deviation of GPAs within classes (+0.13 SD). Effects are greater in higher grades (+0.21 SD and +0.30 SD in grades 5 and 6, respectively).

These results suggest a lack of accompanying measures to manage the increased number of students, leading to larger and more heterogeneous classes in treated municipalities. Although the effect of larger classes on learning is mixed and highly context specific (Pritchett and Sandefur, 2014), evidence from higher-income settings points to negative

³⁷To derive the total increase in enrollment by the end of grade 6, we multiply the grade-specific dropout rate by the number of years spent in primary school (i.e., the number of primary school grades plus the average number of grades repeated).

³⁸Indeed, there is some evidence in Table A9 that Tayssir increased the probability of students repeating grade 1.

Table 3: Effect of Tayssir on class size and heterogeneity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6	All Grades
Panel A. Effects on class size							
LATE	2.805*** (0.846)	1.056 (0.706)	1.679** (0.713)	1.744** (0.703)	1.803** (0.857)	3.599*** (0.887)	2.132*** (0.526)
Sample mean	30.076	29.136	29.689	29.440	29.319	30.682	29.738
Sample SD	9.196	9.201	8.759	9.021	9.169	9.365	9.138
Observations	330,981	303,193	301,795	291,416	283,863	307,908	1,819,156
Panel B. Effects on class heterogeneity							
LATE	0.055*** (0.020)	-0.029 (0.025)	-0.017 (0.020)	0.053*** (0.020)	0.076*** (0.028)	0.111** (0.045)	0.044*** (0.015)
Sample mean	1.165	1.053	1.037	1.002	1.011	1.121	1.068
Sample SD	0.323	0.322	0.316	0.330	0.357	0.369	0.342
Observations	330,122	300,491	298,634	288,394	280,757	304,633	1,803,031

Notes: This table reports local average treatment effects of Tayssir cash transfers (Eq. 1 in the text). Each panel in this table represents a separate outcome variable. *Class size*: number of students per class. *Class heterogeneity*: standard deviation of the GPAs within a class. The unit of observation is a student-year for the school years 2015/16 and 2016/17. Sample restricted to students in primary schools in municipalities with 2004 poverty rates in the [20%, 40%] range. Observations are weighted using a triangular kernel function. Robust standard errors in parenthesis are clustered at the municipality year level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Source: Authors' calculation using data from MASSAR and *Haut-Commissariat au Plan*.

effects (see [Bouguen et al. 2017](#) for a review). In addition, greater heterogeneity in class composition can lead to instruction that is less tailored to the needs of students, resulting in reduced learning outcomes ([Duflo et al., 2011, 2020](#); [Angrist and Meager, 2023](#)).

In Table 4, we further examine how primary school graduation exam scores are associated with class size and heterogeneity. These results are not intended for causal interpretations but provide useful information on the Moroccan context. In addition, we note that class size is more plausibly exogenous in our setting as the typical school has at most one class per grade and cannot sort higher- or lower-ability students into smaller classes.³⁹ Overall, we see that test scores are lower in larger classes and in classes which are more heterogeneous. Looking at results by gender, we see stronger negative associations for boys than for girls. A ten-student increase in class size is associated with a decrease in test scores by 0.07 SD for boys and 0.00 SD for girls. A one standard deviation increase in heterogeneity in class composition is associated with a 0.24 SD decrease in the test scores of boys and a 0.13 SD decrease in those of girls. We obtain similar results controlling for

³⁹In schools with more than one class per grade, we use the average class size by school and grade instead of the individual class size to overcome the endogenous selection of stronger or weaker students into small classes at the school level.

year and municipality fixed effects.

Table 4: Association between class size, class heterogeneity and test scores

	Dep. Var.: Standardized graduation score					
	All children		Boys		Girls	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Association with class size						
Class size	-0.004*** (0.001)	-0.004*** (0.001)	-0.007*** (0.001)	-0.005*** (0.001)	0.000 (0.001)	-0.002** (0.001)
Dep. Var. SD	1.000	1.000	0.935	0.935	1.066	1.066
Observations	277,274	277,137	149,640	149,464	127,634	127,478
Panel B. Association with heterogeneity in class composition						
Class heterogeneity	-0.183*** (0.018)	-0.165*** (0.019)	-0.221*** (0.019)	-0.198*** (0.021)	-0.139*** (0.024)	-0.128*** (0.020)
Dep. Var. SD	1.000	1.000	0.935	0.935	1.066	1.066
Observations	274,985	274,958	148,453	148,370	126,532	126,461
Year and municipality FE	No	Yes	No	Yes	No	Yes

Notes: Panel A reports the association between class size and end-of-primary school test scores. Panel B reports the association class heterogeneity and end-of-primary school test scores. Columns 1, 3 and 5 report the simple correlations for all children, for boys, and for girls, respectively. Columns 2, 4 and 6 add year and municipality fixed effects. The unit of observation is a student-year for the school years 2015/16 and 2016/17. Sample includes grade 6 students in municipalities with 2004 poverty rates in the [20%, 40%] range. Robust standard errors in parenthesis are clustered at the municipality year level. *** p<0.01, ** p<0.05, * p<0.1. Source: Authors' calculation using data from MASSAR and Morocco's *Haut-Commissariat au Plan*.

5.3 Impacts on test scores

In this section, we further test whether effects on class size and composition constrained learning by directly assessing Tayssir impacts on primary school graduation exam scores. Of course, Tayssir may have impacted exam scores through other channels, complicating the interpretation of the results. However, we argue that negative reduced-form estimates would still provide empirical support for the importance of the class size and composition channel, as effects from other channels are likely to run in the opposite direction. For example, transfers could improve learning through increased investments in complementary inputs such as food, school supplies, and parental time.⁴⁰ The conditionality on attendance could also lead to more learning for children who do not attend school regularly.

⁴⁰These effects could emerge either directly using the transfers (Fiszbein and Schady, 2009) or indirectly through an information effect signaling the value of education (Benhassine et al., 2015).

The selection problem. Estimating the impact of Tayssir on test scores is prone to selection issues because the scores of children who drop out before the exam are not observed, and because, as documented above, the program caused a reduction in dropout rates. If dropouts are negatively selected, estimates will be biased downward (more children with low test scores will take the exam in beneficiary municipalities). In Table 5, we compare the characteristics of children who drop out relative to the characteristics of children who do not drop out.⁴¹ Overall, the two groups are very different, and dropouts appear to be negatively selected. Dropouts have, on average, a lower GPA (standardized differences of -0.51 and -0.88 in t and $t-1$, respectively),⁴² especially in grades where the effect of Tayssir on dropout is the largest. Student dropouts are also more likely to have repeated grades (+45 p.p. in $t-1$). This pattern of negative selection into dropout indicates that an analysis of the effects of Tayssir on exam scores using the sample of test-takers is likely to introduce a downward bias in the estimates.⁴³

Table 5: Descriptive statistics: Selection into dropout

	All grades				Std. diff. by grade					
	(1) Drop-out = 1	(2) Drop-out = 0	(3) Std. diff	(4) N	(5) Grade 1	(6) Grade 2	(7) Grade 3	(8) Grade 4	(9) Grade 5	(10) Grade 6
Age	12.301 (2.485)	8.609 (2.291)	1.545	1,834,398	1.326	0.665	1.058	1.286	1.366	1.197
Boy	0.469 (0.499)	0.529 (0.499)	-0.119	1,834,399	0.053	-0.016	-0.002	-0.068	-0.087	-0.199
GPA (t)	4.943 (2.154)	5.835 (1.205)	-0.511	1,771,175	-0.921	-0.149	-0.152	-0.388	-0.766	-0.667
GPA ($t-1$)	4.520 (1.554)	5.730 (1.187)	-0.875	1,538,313	-0.800	-0.527	-0.683	-0.850	-0.910	-0.969
Repeat grade ($t-1$)	0.587 (0.492)	0.141 (0.348)	1.046	1,818,776	3.098	0.458	0.757	0.931	0.963	1.146
Repeat grade ($t-2$)	0.275 (0.446)	0.127 (0.333)	0.376	1,527,464	0.430	0.156	0.463	0.485	0.450	0.447

Notes: Columns 3 and 5–10 report the standardized difference between dropouts and non-dropouts (i.e., the difference in means in units of standard deviation). The unit of observation is a student-year for the school years 2015/16 and 2016/17. Sample includes students in primary schools in municipalities with 2004 poverty rates in the range [20%, 40%]. Standard deviations are in parenthesis. Source: Authors' calculation using data from MASSAR.

⁴¹Differences are qualitatively similar when we consider boys and girls separately (see Table A7 in the Online Appendix).

⁴²Estimates in $t-1$ are cleaner than in t because of the lower prevalence of missing GPAs for dropouts (2% in $t-1$ against 95% in t).

⁴³In Table A8, we also explore whether Tayssir affected the characteristics of dropouts using equation (1) and a set of observable characteristics as outcome variables. We find suggestive evidence that the pattern of negative self-selection into dropout is more severe in targeted municipalities (dropouts have lower GPAs and are more likely to have repeated a grade prior to dropout). In what follows, we focus on correcting the overall pattern of selection instead of the specific patterns by treatment status. We note, however, that this decision must be seen as conservative: imputing lower test scores to treated dropouts would strengthen the negative effect we find below.

We rely on two imputation strategies to correct for this bias. First, using information on the GPA of students prior to drop out, we impute test scores corresponding to the likely degree of selection into dropout. For each student who drops out, we extract their location in the distribution of their municipality’s GPAs in $t-1$,⁴⁴ and impute test scores corresponding to p^* , the average percentile of dropouts in the GPA distribution.⁴⁵ For students who drop out, we start by extracting their location in the distribution of their municipality’s GPAs in $t-1$. Then, we derive p^* as the average percentile of dropouts in the GPA distribution. We do this separately for our three panels: all children, boys, and girls. We estimate $p^* = 32$ overall, $p^* = 32$ for boys, and $p^* = 33$ for girls. The key assumption for this imputation strategy to be valid is that the location of dropouts in the GPA distribution is a good proxy for their (unobserved) location in the test score distribution.

Second, we impute test scores which correspond to different degrees of selection into dropout and provide lower and upper bounds to Tayssir true effects. Because the overall pattern of selection documented in Table 5 is negative, we conjecture that children who drop out from school would have obtained a score on the lower end of the distribution (i.e., anywhere below the median score), and consider two extreme scenarios: (i) we assume dropouts would have obtained a score of zero; (ii) we assume dropouts would have obtained the median score.⁴⁶ Since dropouts are negatively selected, their (unobserved) test scores should stand somewhere between these two values, and, as a consequence, estimates using these imputed values should provide lower and upper bounds to the true effect of Tayssir

⁴⁴We refrain from using only information on the school-specific distribution of GPAs because dropouts are likely to be concentrated in some schools and this would tend to overestimate their average location in the distribution, overstating the negative effect on test scores we document in the next sub-section.

⁴⁵We acknowledge that GPAs obtained in different schools are not necessarily comparable. To make them more comparable, we use information on standardized test scores and derive a conversion factor corresponding to the average test score of students in a given school relative to the best performing school in their municipality. This conversion factor deflates the GPA of students from schools that perform worse in the exam. If worse-performing schools already incorporate this information in their GPAs, and if dropouts are more likely to be concentrated in worse-performing schools, the use of this conversion factor will attenuate both p^* and the negative effect on test scores we document below.

⁴⁶We compute median scores separately for boys and girls in each year because girls tend to have higher scores (Table A4 in the Online Appendix).

on test scores.^{47,48} Imputing zero gives the more conservative bound, while imputing the median score gives the less conservative bound. To illustrate effects associated with more intermediate patterns of selection, we also report estimates imputing scores corresponding to the 10th and 25th percentiles of the test score distribution.

Results. The results on test scores with and without corrections for selective dropout rates are shown in Table 6. We focus here on the score obtained at the graduation exam administered to all students at the end of grade 6. This exam is managed at the provincial level and therefore allows us to compare children from different schools and municipalities within each province (see Section 3.2 for more details). Reduced-form estimates are presented in Figures A4 and A5.

Estimates with no correction for differential dropout rates suggest that receiving cash transfers is associated with a decrease in exam scores of 0.16 SD (column 1, panel A). However, because of the downward bias documented above, it is not clear whether this estimate is due to a selection bias or to a genuine effect of cash transfers. Looking at impacts by gender, we find that negative effects are concentrated within the sample of boys (-0.22 SD). This result is interesting because boys are *less* affected by the reduction in the dropout rate induced by Tayssir (see Section 5.1) and should therefore be less affected by the selection bias highlighted above.⁴⁹ If anything, this suggests that negative estimates in column 1 are not entirely driven by selection issues and must in part reflect a genuine negative effect of cash transfers.

Estimates using the two imputation methods described above are presented in columns 2 to 6.⁵⁰ Column 2 reports the results from the first method, using information on the

⁴⁷More specifically, because Tayssir decreased the number of dropouts and because dropouts are negatively selected, imputing the median score bounds the effect of the program from below, while imputing the score of zero bounds the effect of the program from above.

⁴⁸Although we believe positive selection into dropout is unlikely (based on the literature and on the results from Table 5), we note that our results are qualitatively unchanged (and in fact reinforced) for scores imputed from the higher end of the distribution.

⁴⁹This result is not due to different patterns of selection across genders. Table A7 in the Online Appendix shows that negative selection into dropout is similar for both gender. In particular, the standardized differences between dropouts and non-dropouts are similar for boys and girls (column 3).

⁵⁰Due to data availability, we do not observe students in grade 6 during their whole primary education and therefore do not know for each cohort the exact number of students who drop out prior to grade 6. We proxy this quantity by relying on cross-sectional dropouts, that is, the total number of dropouts observed in a given year in grades 1 to 6. Because grade-specific dropout rates have been very stable over

Table 6: Effect of Tayssir on test scores

	Dep. Var.: Standardized graduation score					
	(1)	(2)	(3)	(4)	(5)	(6)
	correction diff. drop-out: imputed score:	No	Correction 1	Correction 2		
		p*	0	p10	p25	p50
Panel A. All children						
LATE	-0.157**	-0.110**	-0.025	-0.057	-0.102*	-0.136**
	(0.063)	(0.054)	(0.073)	(0.059)	(0.054)	(0.053)
Sample mean	-0.000	-0.064	-0.353	-0.236	-0.096	0.015
Sample SD	1.000	0.925	1.209	1.058	0.939	0.915
Observations	279,880	335,419	335,419	335,419	335,419	335,419
Panel B. Boys						
LATE	-0.206***	-0.160***	-0.100	-0.129**	-0.152***	-0.176***
	(0.064)	(0.055)	(0.073)	(0.058)	(0.056)	(0.055)
Sample mean	-0.069	-0.126	-0.368	-0.247	-0.155	-0.063
Sample SD	0.935	0.875	1.130	0.965	0.889	0.864
Observations	151,024	176,734	176,734	176,734	176,734	176,734
Panel C. Girls						
LATE	-0.089	-0.050	0.066	0.029	-0.034	-0.079
	(0.072)	(0.060)	(0.079)	(0.068)	(0.061)	(0.061)
Sample mean	0.081	0.019	-0.335	-0.224	-0.029	0.102
Sample SD	1.066	0.969	1.291	1.152	0.988	0.961
Observations	128,856	158,685	158,685	158,685	158,685	158,685

Notes: This table reports local average treatment effects of Tayssir cash transfers on test scores using Equation 1. The dependent variable is the standardized primary school graduation score. Column 1: Results with no correction for differential dropout rates. Columns 2 to 6 report the results using two different correction methods for selection into dropout. Column 2: We use information on the GPA of students prior to dropout and impute test scores corresponding to the likely degree of selection into dropout p^* ($p^* = 32$ overall, $p^* = 32$ for boys, and $p^* = 33$ for girls). Columns 3 to 6: We conjecture that dropouts would have obtained scores on the lower end of the distribution and impute scores corresponding to different degrees of selection. Specifically, column 3 assumes that dropouts would have obtained a score of 0 (which provides an upper bound to the true effect), whereas column 6 assumes that dropouts would have obtained the median score (which provides a lower bound to the true effect). Columns 4 and 5 report estimates for more intermediate patterns of selection (i.e., imputing scores corresponding to the 10th and 25th percentiles of the test score distribution). The unit of observation is a student-year for the school years 2015/16 and 2016/17. Sample restricted to grade 6 students in municipalities with 2004 poverty rates in the [20%, 40%] range. Sample also includes dropouts in columns 2-6. Observations are weighted using a triangular kernel function. Robust standard errors in parenthesis are clustered at the municipality year level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Source: Authors' calculation using data from MASSAR and Morocco's *Haut-Commissariat au Plan*.

GPA of students prior to dropout to impute test scores corresponding to the likely degree of selection into dropout p^* ($p^* = 32$ overall, $p^* = 32$ for boys, and $p^* = 33$ for girls). We find that the program reduced test scores by 0.12 SD with stronger negative effects for boys (-0.18 SD). We find small and non-significant results for girls. The coefficient is equivalent to a 0.05 SD decrease in test scores. The smallest increase we can reject is 0.07 SD. Such effects are generally considered as small in the literature, and on the lower-end of the distribution of standardized effect sizes provided by [Evans and Yuan \(2022\)](#) for education

time (see Table A3 in the Online Appendix), we are confident that this quantity provides a good proxy of the number of students who dropped out prior to grade 6.

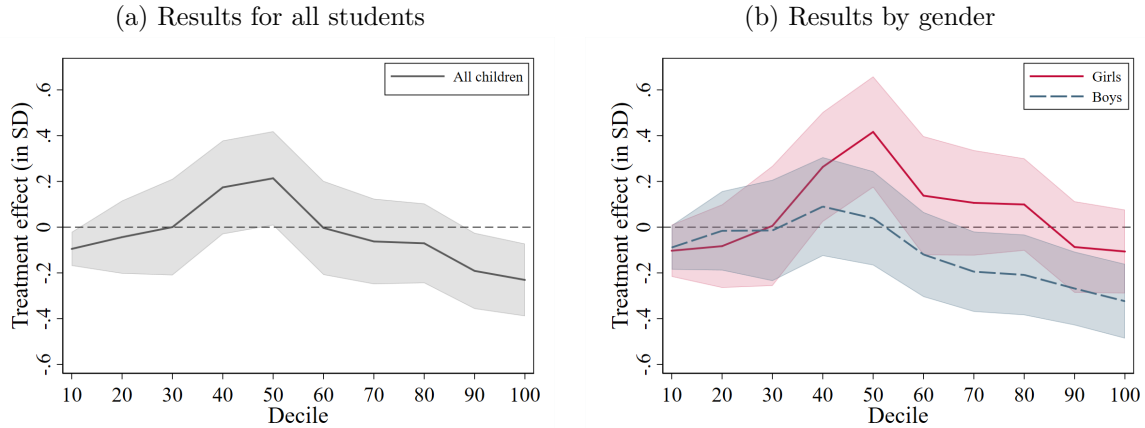
interventions in low- and middle-income countries, suggesting that the lack of significance of the results for girls is not due to a lack of statistical power. Columns 3 to 6 report the results from the second method, where we conjecture that dropouts would have obtained scores on the lower end of the distribution and impute scores corresponding to different degrees of selection. Among the full sample of children, we estimate an effect somewhere between -0.02 SD and -0.15 SD (only significant at the lower end of the interval). The results, by gender, suggest consistent gender dynamics. For boys, we estimate negative effects between 0.09 SD and 0.20 SD (panel B). These effects are generally significant at conventional levels, except for the most severe correction, which imputes the score of zero to all dropouts. We note, however, that imputing zero is very conservative, and that effects are significant when imputing scores corresponding to the 10th percentile (which is also conservative). The effect turns insignificant ($p\text{-value} > 0.05$) when we impute scores below the 7th percentile. For girls, we estimate smaller effects, with bounds well distributed around zero: we find a lower bound of -0.08 SD, an upper bound of $+0.05$ SD, and coefficients that are not statistically significant (panel C).

Results on test scores depend on what happened for two groups of children: (1) the group of children who would have dropped out in the absence of the program; (2) the group of children who enroll in school regardless of the program. We believe that for (1) the treatment effect should be positive, whereas for (2), in light of our results, treatment effects should be negative (at least for boys). While it is not possible to identify children in (1) and (2), looking at impacts on test scores at various points in the distribution can be helpful to assess the magnitude of these effects. Because of the negative selection into dropout, children in (1) are more likely to be located in the lower end of the distribution, while children in (2) are more likely to be located in the upper end of the distribution. To implement this heterogeneity analysis, we reproduce the results from column 3 in Table 6 where we impute the score of 0 to dropouts (doing so is important to limit composition issues and ensure dropouts are located in the lower end of the distribution) by deciles of test scores.⁵¹ Results are reported in Figure 5. Consistent with the theory, negative

⁵¹We stratify by year, municipality, treatment status and gender to construct the deciles.

effects on test scores seem concentrated on the higher end of the distribution, especially for boys.

Figure 5: Impacts of Tayssir on test scores by deciles



Notes: These figures report treatment effects of Tayssir cash transfers on test scores at various points in the distribution of test scores. The dependent variable is the standardized primary school graduation score from column 3 in Table 6 (where we impute the score of 0 to dropouts – doing so is important to limit composition issues and ensure dropouts are located in the lower end of the distribution). We stratify by year, municipality, treatment status and gender to construct the deciles. See notes to Table 6 for more details. Shaded areas represent the 95% confidence intervals. Source: Authors’ calculation using data from MASSAR and Morocco’s *Haut-Commissariat au Plan*.

Overall, the estimates reported in Table 6 and in Figure 5 are consistent with the class size and composition channel, and suggest that Tayssir constrained learning for children who enroll in school regardless of the existence of the program. These findings contrast with those of Lucas and Mbiti (2012), Blimpo et al. (2019) and Valente (2019), who showed that increased enrollment following policies eliminating school fees in sub-Saharan Africa did not generate negative effects on test scores. However, the lack of negative effects in these studies may be explained by the expansion of private education which followed the removal of public school fees and by the increased socio-economic sorting of students into schools (Bold et al., 2015) – a mechanism that is less likely to operate in our setting because private education is generally absent in rural Morocco.

More broadly, our results also highlight the importance of class size and composition for children learning in a lower-income setting. We argue that two factors help to understand why the negative effects of Tayssir on test scores are concentrated within the sample of boys. First, Tayssir increased the proportion of girls in classes, and there is evidence in the literature that this can have detrimental effects on boys’ learning (Black et al.,

2013; Hill, 2015; Briole, 2021), for example, because teachers adapt their practices to the gender composition of the class and tend to favor the learning of girls when they are in higher proportions (Briole, 2021), or because boys can be distracted by a higher proportion of girls in the class (Hill, 2015). Second, Tayssir retained in school students that are negatively selected, and there is evidence suggesting that a higher proportion of low performers can be particularly detrimental to boys' learning. Lavy et al. (2012) show that a higher proportion of repeaters has negative effects on the school performances of other children, especially the worse-performing students (which in our sample are boys – see Table A4 in the Online Appendix). Carrell and Hoekstra (2010) show that boys suffer the most from the presence of “troubled kids” in the class (the authors consider a kid as troubled if he or she is exposed to domestic violence).

5.4 Impacts on educational attainment

In Table 7, we explore program effects on educational attainment by focusing on two dummy variables indicating whether a child (i) graduates from primary school, and (ii) enrolls in secondary school (see Section 3.2 for more details on the definition of these variables). We take into account the issue of selection into dropout by imputing the value of zero to both of these variables for dropouts (who by definition do not graduate from primary school or enroll in secondary school). We present effects with and without corrections for more transparency. Figures A7 and A8 show reduced-form estimates. Without corrections, we find small and non-significant effects on both graduation and enrollment (columns 1 and 3). Correcting for differential dropout, the overall effect on graduation remains non-significant in the full sample (column 2, panel A), but looking at effects by gender we find a 5.7 p.p. increase for girls (significant at 5%). This corresponds to an increase of 9% relative to the sample mean of 65.3%. In contrast, we see no significant improvements for boys, which likely reflects the negative effects on test scores highlighted above. We find that the program also led to a significant increase in enrollment in secondary school once we correct for differential dropout (column 4). We estimate that the program increased enrollment by 4.5 p.p. (panel A), equivalent to a 7% increase relative

Table 7: Effect of Tayssir on educational attainment

dep. var.:	Graduation from primary school		Enrollment in secondary school	
	(1) No correction	(2) Correction	(3) No correction	(4) Correction
diff. drop-out:				
Panel A. All children				
LATE	-0.023 (0.029)	0.020 (0.026)	0.007 (0.021)	0.045** (0.021)
Mean	0.765	0.638	0.765	0.638
Observations	278,739	334,299	278,739	334,299
Panel B. Boys				
LATE	-0.044 (0.035)	-0.009 (0.031)	-0.005 (0.026)	0.024 (0.024)
Mean	0.732	0.625	0.742	0.633
Observations	150,390	176,112	150,390	176,112
Panel C. Girls				
LATE	0.007 (0.025)	0.057** (0.024)	0.024 (0.019)	0.070*** (0.020)
Mean	0.805	0.653	0.793	0.644
Observations	128,349	158,187	128,349	158,187

Notes: The table reports local average treatment effects of Tayssir cash transfers (Eq. 1 in the text). Graduation from primary school: dummy variable coded one if a student in grade 6 graduates from primary school. Enrollment in secondary school: dummy variable coded one if a student in grade 6 is enrolled in secondary school the following year. The unit of observation is a student-year for the school years 2015/16 and 2016/17. Sample restricted to grade 6 students in municipalities with 2004 poverty rates in the [20%, 40%] range. Sample in columns 2 and 4 also include dropouts. Observations are weighted using a triangular kernel function. Robust standard errors in parenthesis are clustered at the municipality year level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Source: Authors' calculation using data from MASSAR and *Haut-Commissariat au Plan*.

to the sample mean of 63.8%. We estimate non-significant effects for boys (panel B) but sizable effects for girls (7 p.p. or 11% relative to the sample mean of 64.4%, panel C).

6 Conclusion

The invention and spread of CCT programs in the past few decades have been described as “*perhaps the most remarkable innovation in welfare programs in developing countries*” (Cahyadi et al., 2020). Despite robust evidence on the positive effects of these programs on enrollment, evidence on their learning impacts has been mixed. In this paper, we point to a new channel to explain this apparent puzzle. Our analyses of a CCT program implemented at scale in Morocco indicate that the program constrained learning by putting additional pressure on existing resources.

Naturally, this finding should in no way be interpreted as evidence that policy makers should not pursue CCT programs. Such programs, including the one studied in this paper, have been particularly effective at increasing access to basic education. Increased education is a necessary first step to affect learning and has many benefits beyond learning, including delayed marriage and childbearing for adolescent girls ([Bergstrom and Özler, 2022](#)). However, our results, together with evidence showing that students in low- and middle-income countries have alarmingly low literacy and numeracy levels ([Bold et al., 2017](#); [World Bank, 2018](#)), suggest that the gains in attendance from CCT programs are unlikely on their own to equip students with the foundational skills they need to thrive. In fact, our results show that CCT programs can have adverse effects on learning when schools are not provided with the necessary resources to absorb the associated extra influx of students.

Recent decades have seen a surge in evaluations focusing on the learning effects of interventions in low- and middle-income countries (for recent reviews see [McEwan, 2015](#); [Ganimian and Murnane, 2016](#); [Glewwe and Muralidharan, 2016](#); [Masino and Niño-Zarazúa, 2016](#); [Conn, 2017](#); [Evans and Mendez Acosta, 2021](#)). Although the literature does not allow one to identify programs that are effective in all contexts ([World Bank, 2018](#)), some lessons have emerged, and [Evans and Popova \(2016\)](#) in particular identified two classes of interventions that have been very consistent at improving learning: (i) pedagogical interventions that tailor teaching to student learning; (ii) individualized and sustained efforts to improve teachers' skills and practices. An interesting question for future research – and a promising avenue for policy makers aiming to address the learning crisis – would be to explore the possible complementarities between these interventions and CCT programs. Recent evidence suggests that combining education interventions can produce effects that are greater than the sum of their individual effects ([Mbiti et al., 2019](#)).

References

- Abdul-Hamid, H. (2017). *Data for Learning: Building a Smart Education Data System*. Washington, D.C.: World Bank Publications.
- Angrist, J. D., Lavy, V., Leder-Luis, J., and Shany, A. (2019). Maimonides' rule redux. *American Economic Review: Insights*, 1(3):309–24.
- Angrist, N. and Meager, R. (2023). Implementation matters: Generalizing treatment effects in education. Available at SSRN 4487496.
- Araujo, M. C. and Macours, K. (2021). Education, income and mobility: Experimental impacts of childhood exposure to *progesa* after 20 years. Working Paper.
- Araujo, M. C., Rubio-Codina, M., and Schady, N. R. (2021). 70 to 700 to 70,000: Lessons from the Jamaica experiment. IDB Working Paper Series.
- Baez, J. E. and Camacho, A. (2011). Assessing the long-term effects of conditional cash transfers on human capital: Evidence from Colombia. World Bank Policy Research Working Paper No. 5681.
- Baird, S., Ferreira, F. H., Özler, B., and Woolcock, M. (2014). Conditional, unconditional and everything in between: a systematic review of the effects of cash transfer programmes on schooling outcomes. *Journal of Development Effectiveness*, 6(1):1–43.
- Baird, S., McIntosh, C., and Özler, B. (2011). Cash or condition? Evidence from a cash transfer experiment. *The Quarterly Journal of Economics*, 126(4):1709–1753.
- Baird, S., McIntosh, C., and Özler, B. (2019). When the money runs out: do cash transfers have sustained effects on human capital accumulation? *Journal of Development Economics*, 140:169–185.
- Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukerji, S., Shotland, M., and Walton, M. (2017). From proof of concept to scalable policies: Challenges and solutions, with an application. *Journal of Economic Perspectives*, 31(4):73–102.
- Barham, T., Macours, K., and Maluccio, J. A. (2023). Experimental evidence from a conditional cash transfer program: Schooling, learning, fertility and labor market outcomes after 10 years. Working Paper.
- Barrera-Osorio, F., Linden, L. L., and Saavedra, J. E. (2019). Medium-and long-term educational consequences of alternative conditional cash transfer designs: Experimental evidence from Colombia. *American Economic Journal: Applied Economics*, 11(3):54–91.

- Behrman, J. R., Parker, S. W., and Todd, P. E. (2005). Long-term impacts of the oportu-
nidades conditional cash transfer program on rural youth in Mexico. Discussion Paper
122. Ibero-America Institute for Economic Research.
- Benhassine, N., Devoto, F., Duflo, E., Dupas, P., and Pouliquen, V. (2015). Turning
a shove into a nudge? A” labeled cash transfer” for education. *American Economic
Journal: Economic Policy*, 7(3):86–125.
- Bergstrom, K. and Özler, B. (2022). Improving the well-being of adolescent girls in
developing countries. *The World Bank Research Observer*, page lkac007.
- Black, S. E., Devereux, P. J., and Salvanes, K. G. (2013). Under pressure? The effect of
peers on outcomes of young adults. *Journal of Labor Economics*, 31(1):119–153.
- Blimpo, M. P., Gajigo, O., and Pugatch, T. (2019). Financial constraints and girls’
secondary education: Evidence from school fee elimination in the Gambia. *The World
Bank Economic Review*, 33(1):185–208.
- Bold, T., Filmer, D., Martin, G., Molina, E., Stacy, B., Rockmore, C., Svensson, J., and
Wane, W. (2017). Enrollment without learning: Teacher effort, knowledge, and skill in
primary schools in Africa. *Journal of Economic Perspectives*, 31(4):185–204.
- Bold, T., Kimenyi, M., Mwabu, G., and Sandefur, J. (2015). Can free provision reduce de-
mand for public services? Evidence from Kenyan education. *The World Bank Economic
Review*, 29(2):293–326.
- Bold, T., Kimenyi, M., Mwabu, G., Sandefur, J., et al. (2018). Experimental evidence on
scaling up education reforms in Kenya. *Journal of Public Economics*, 168:1–20.
- Bouguen, A., Grenet, J., and Gurgand, M. (2017). Does class size influence student
achievement? Notes IPP No. 28.
- Briole, S. (2021). Are girls always good for boys? Short and long term effects of school
peers’ gender. *Economics of Education Review*, 84:102–150.
- Cahyadi, N., Hanna, R., Olken, B. A., Prima, R. A., Satriawan, E., and Syamsulhakim,
E. (2020). Cumulative impacts of conditional cash transfer programs: Experimental
evidence from Indonesia. *American Economic Journal: Economic Policy*, 12(4):88–
110.
- Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014). Robust nonparametric confidence
intervals for regression-discontinuity designs. *Econometrica*, 82(6):2295–2326.

- Carrell, S. E. and Hoekstra, M. L. (2010). Externalities in the classroom: How children exposed to domestic violence affect everyone’s kids. *American Economic Journal: Applied Economics*, 2(1):211–28.
- Casey, K. (2018). Radical decentralization: does community-driven development work? *Annual Review of Economics*, 10:139–163.
- Cattaneo, M. D., Frandsen, B. R., and Titiunik, R. (2015). Randomization inference in the regression discontinuity design: An application to party advantages in the US senate. *Journal of Causal Inference*, 3(1):1–24.
- Cattaneo, M. D., Idrobo, N., and Titiunik, R. (2019a). *A practical introduction to regression discontinuity designs: Foundations*. Cambridge University Press.
- Cattaneo, M. D., Titiunik, R., and Vazquez-Bare, G. (2019b). Power calculations for regression-discontinuity designs. *The Stata Journal*, 19(1):210–245.
- Cattaneo, M. D. and Vazquez-Bare, G. (2016). The choice of neighborhood in regression discontinuity designs. *Observational Studies*, 2(134):A146.
- Conn, K. M. (2017). Identifying effective education interventions in sub-saharan africa: A meta-analysis of impact evaluations. *Review of Educational Research*, 87(5):863–898.
- Cottin, R. (2019). Le ciblage direct des ménages est-il possible pour les politiques de santé? Le cas du RAMed au Maroc. *Mondes en développement*, (3):29–50.
- CSEFRS (2019). *Résultats des élèves marocains dans l’étude internationale sur le progrès en littérature PIRLS 2016*. Rabat, Maroc: Conseil Supérieur de l’Education, de la Formation et de la Recherche Scientifique.
- Duflo, A., Kiessel, J., and Lucas, A. (2020). External validity: Four models of improving student achievement. NBER Working Paper No. 27298.
- Duflo, E., Dupas, P., and Kremer, M. (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya. *American Economic Review*, 101(5):1739–74.
- Duflo, E., Dupas, P., and Kremer, M. (2015). School governance, teacher incentives, and pupil–teacher ratios: Experimental evidence from Kenyan primary schools. *Journal of Public Economics*, 123:92–110.
- Duque, V., Rosales-Rueda, M., and Sanchez, F. (2019). How do early-life shocks interact with subsequent human capital investments? Evidence from administrative data. Working Paper.

- Dustan, A. (2020). Can large, untargeted conditional cash transfers increase urban high school graduation rates? Evidence from Mexico city’s Prepa Sí. *Journal of Development Economics*, 143:102392.
- El-Kogali, S. E. T., Krafft, C., Abdelkhalek, T., Benkassmi, M., Chavez, M., Bassett, L., and Ejjanoui, F. (2016). The impact of a community development and poverty reduction program on early childhood development in Morocco. World Bank Policy Research Working Paper No. 7671.
- Ennaji, M. (2018). Morocco’s experience in gender gap reduction in education. *Gender and Women’s Studies*, 2(1):5.
- Evans, D. K. and Mendez Acosta, A. (2021). Education in africa: What are we learning? *Journal of African Economies*, 30(1):13–54.
- Evans, D. K. and Popova, A. (2016). What really works to improve learning in developing countries? An analysis of divergent findings in systematic reviews. *World Bank Research Observer*, 31(2):242–270.
- Evans, D. K. and Yuan, F. (2022). How big are effect sizes in international education studies? *Educational Evaluation and Policy Analysis*, 44(3):532–540.
- Filmer, D. and Schady, N. (2014). The medium-term effects of scholarships in a low-income country. *Journal of Human Resources*, 49(3):663–694.
- Fiszbein, A. and Schady, N. R. (2009). *Conditional cash transfers: reducing present and future poverty*. The World Bank.
- Ganimian, A. J. (2020). Growth-mindset interventions at scale: Experimental evidence from Argentina. *Educational Evaluation and Policy Analysis*, 42(3):417–438.
- Ganimian, A. J. and Murnane, R. J. (2016). Improving education in developing countries: Lessons from rigorous impact evaluations. *Review of Educational Research*, 86(3):719–755.
- Garcia, S. and Saavedra, J. E. (2022). Conditional cash transfers for education. NBER Working Paper No. 29758.
- Gazeaud, J., Mvukiyehe, E., and Sterck, O. (2023). Cash transfers and migration: Theory and evidence from a randomized controlled trial. *Review of Economics and Statistics*, 105(1):143–157.
- Gelman, A. and Imbens, G. (2019). Why high-order polynomials should not be used in regression discontinuity designs. *Journal of Business & Economic Statistics*, 37(3):447–456.

- Glewwe, P. and Muralidharan, K. (2016). Improving education outcomes in developing countries: Evidence, knowledge gaps, and policy implications. *Handbook of the Economics of Education*, 5:653–743.
- Hidrobo, M., Mueller, V., and Roy, S. (2022). Cash transfers, migration, and gender norms. *American Journal of Agricultural Economics*, 104(2):550–568.
- Hill, A. J. (2015). The girl next door: The effect of opposite gender friends on high school achievement. *American Economic Journal: Applied Economics*, 7(3):147–77.
- Imbens, G. and Kalyanaraman, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of Economic Studies*, 79(3):933–959.
- Kerwin, J. T. and 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.
- Lavy, V., Paserman, M. D., and Schlosser, A. (2012). Inside the black box of ability peer effects: Evidence from variation in the proportion of low achievers in the classroom. *The Economic Journal*, 122(559):208–237.
- L’Economiste (2017). Abandon scolaire: Tayssir, neuf ans après. *Accessed online on 2020-09-23*.
- Lucas, A. M. and Mbiti, I. M. (2012). Access, sorting, and achievement: The short-run effects of free primary education in Kenya. *American Economic Journal: Applied Economics*, 4(4):226–53.
- Masino, S. and Niño-Zarazúa, M. (2016). What works to improve the quality of student learning in developing countries? *International Journal of Educational Development*, 48:53–65.
- Mbiti, I., Muralidharan, K., Romero, M., Schipper, Y., Manda, C., and Rajani, R. (2019). Inputs, incentives, and complementarities in education: Experimental evidence from Tanzania. *The Quarterly Journal of Economics*, 134(3):1627–1673.
- McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, 142(2):698–714.
- McEwan, P. J. (2015). Improving learning in primary schools of developing countries: A meta-analysis of randomized experiments. *Review of Educational Research*, 85(3):353–394.

- McKenzie, D. (2016). Power calculations for regression discontinuity evaluations: Part 2. <https://blogs.worldbank.org/impac evaluations/power-calculations-regression-discontinuity-evaluations-part-2>, Development Impact Blog.
- Ministry of Education (2018). *Atlas territorial de l'enseignement privé*. Rabat, Maroc: Conseil Supérieur de l'Éducation, de la Formation et de la Recherche Scientifique.
- Molina Millán, T., Barham, T., Macours, K., Maluccio, J. A., and Stampini, M. (2019). Long-term impacts of conditional cash transfers: Review of the evidence. *The World Bank Research Observer*, 34(1):119–159.
- Molina Millán, T., Macours, K., Maluccio, J. A., and Tejerina, L. (2020). Experimental long-term effects of early-childhood and school-age exposure to a conditional cash transfer program. *Journal of Development Economics*, 143:102385.
- Muralidharan, K. and Niehaus, P. (2017). Experimentation at scale. *Journal of Economic Perspectives*, 31(4):103–24.
- Muralidharan, K. and Sundararaman, V. (2011). Teacher performance pay: Experimental evidence from India. *Journal of Political Economy*, 119(1):39–77.
- Mvukiyehe, E. and Van der Windt, P. (2020). Assessing the longer term impact of community-driven development programs: Evidence from a field experiment in the Democratic Republic of Congo. World Bank Policy Research Working Paper No. 9140.
- Médias24 (2014). Massar: Rachid Belmokhtar ne fera pas de concession. <https://www.medias24.com/SOCIETE/8833-Massar-Rachid-Belmokhtar-ne-fera-pas-de-concession.html>, Accessed online on 2020-09-23.
- Médias24 (2018). Programme Tayssir: le détail des nouvelles mesures de la rentrée 2018-2019. <https://www.medias24.com/MAROC/EDUCATION/185971-Programme-Tayssir-Le-detail-des-nouvelles-mesures-de-la-rentree-2018-2019.html>, Accessed online on 2020-09-23.
- Nguyen, T. C. and Rieger, M. (2017). Community-driven development and social capital: Evidence from Morocco. *World Development*, 91:28–52.
- Otsu, T., Xu, K.-L., and Matsushita, Y. (2013). Estimation and inference of discontinuity in density. *Journal of Business & Economic Statistics*, 31(4):507–524.
- Parker, S. W. and Vogl, T. (2018). Do conditional cash transfers improve economic outcomes in the next generation? Evidence from Mexico. NBER Working Paper No. 24303.

- Pritchett, L. (2013). *The rebirth of education: Schooling ain't learning*. Washington, DC: Center for Global Development.
- Pritchett, L. and Sandefur, J. (2014). Context matters for size: why external validity claims and development practice do not mix. *Journal of Globalization and Development*, 4(2):161–197.
- Soumaya, M., Julie, B., Marguerite, C., Elizabeth, F., Hannah, K., and Isobel, M. (2018). *Examens de l'OCDE du cadre d'évaluation de l'éducation: Maroc*. Paris, France: OECD.
- UNICEF (2015). Maroc: Rapport national sur les enfants non-scolarisés. *Accessed online on 2022-11-02*.
- Urquiola, M. and Verhoogen, E. (2009). Class-size caps, sorting, and the regression-discontinuity design. *American Economic Review*, 99(1):179–215.
- Valente, C. (2019). Primary education expansion and quality of schooling. *Economics of Education Review*, 73:101913.
- Vivalt, E. (2020). How much can we generalize from impact evaluations? *Journal of the European Economic Association*, 18(6):3045–3089.
- World Bank (2006). Morocco National Initiative for Human Development. Project Appraisal Document. Washington, DC: World Bank.
- World Bank (2018). *World Development Report 2018: Learning to realize education's promise*.
- World Bank (2019). Ending learning poverty: the call of our times. *Accessed online on 2020-11-19*.