

Results-based financing and performance incentives in education: new evidence in the literature

**Author:
Louisee Cruz**

March 2022

Acknowledgements

This report was written by Louisee Cruz and edited by Jessica Lee under the World Bank's Results in Education for All Children (REACH) programme. The author is grateful for Omar Arias, Saamira Halabi, and Jessica Lee for their guidance and support. She would like to thank Tara Beteille, Gabriel Demombynes, and Tazeen Fasih for their insightful feedback and comments.

The findings, interpretations, and conclusions expressed in this paper are entirely those of the author. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments that they represent.

Executive summary

Results-based financing (RBF) has been applied in education with the aim to improve the effectiveness and efficiency of spending. This is particularly relevant in light of the COVID-19 pandemic. Since the onset of COVID-19, the education budget has shrunk in 65% of developing countries. Also, the share of children experiencing learning poverty has only increased, which has led to a call for more efficient use of resources to assure children the right to education and foster human capital (Al-Samarrai et al., 2021; World Bank et al., 2021).

This study considers RBF as an umbrella term referring to any program or intervention that provides rewards to individuals or institutions after agreed-upon results are achieved and verified. The definition of “result” varies per intervention, with some adopting intermediate outcomes (as teacher attendance) or final outcomes (as learning).

Incentivized actors can be students and families, teachers, schools, education administrators, and governments. Incentives for students and families frequently take the form of conditional cash transfers - which look to increase enrollment, attendance, and completion rates - or merit-based rewards to foster student learning. Incentives for teachers try to stimulate teachers’ attendance, to focus on student learning, and, in some instances, to select teachers during recruitment processes. Some countries have also introduced school rewards based on performance indicators to drive quality improvements. In decentralized countries, governments have tried linking funding to subnational levels (as mayors or regional education bureaus) according to their education performance. Finally, development agencies also attach part of disbursements to achieving pre-established performance targets.

This report updates the state of the literature on RBF in education since the first analysis conducted in 2019 (Lee & Medina, 2019). It focuses on incentives for teachers, students and families, schools, and education administrators. The evidence base continues to increase at a fast pace. Still, there is significantly more research on teachers and student-family incentives. Newer evidence on these two actors sheds light especially on whether RBF contributes to student learning and the long-term effects of interventions. Design and implementation aspects continue to have high relevance for the success of initiatives. For example, aligning incentives for different education agents can be more effective than a scheme targeting a single actor, or combining incentives with appropriate support helps engage actors, especially those starting at a lower baseline.

The takeaways for each type of incentive analyzed in this paper are:

- 1. Teachers:** Teacher incentives have been largely implemented as a bonus policy instead of a change in teachers' contracts. Incentives designed to improve teacher attendance can raise attendance, but strong monitoring systems are required, **which can be politically difficult to introduce**. In addition, incentives have been shown to increase test scores and improve learning, but **achieving those results is not straightforward**. Key factors that **seem to** lead to positive results are: i) assuring schools have minimum learning resources, ii) the reward rule clearly puts learning as the main outcome indicator, and iii) teachers set specific targets for their students and articulate them to learning strategies. To be sustainable and fully effective, incentives typically require an accountability mechanism that helps shape teachers' behavior in order to achieve long-term positive effects.
- 2. Students and families:** Incentives can foster schooling and learning, but the design rule generally targets one over the other. Conditional cash transfers can increase enrollment, attendance, and completion rates, especially in primary and lower secondary education. In high school, short-term average impacts on schooling are mixed, though there are positive effects for the most vulnerable students. In the long run, children exposed to CCTs during primary education have increased upper secondary enrollment and completion, probably due to the increased salience of education returns. Evidence on learning effects is mixed, with small positive impacts. Merit-based incentives tied to students' test scores are still growing, with promising results.
- 3. Schools:** Robust evidence on school incentives is currently restricted to two studies on performance-based school grants that use a tournament design. These grants are awarded based on learning outcomes. Although limited, evidence on RBF for schools seems promising, especially when the incentive is combined with capacity-building interventions, such as peer-mentoring to principals of high- and low-performing schools. The lack of robust evidence does not mean that performance grants are rare. In fact, several incentives to schools have been implemented in development projects, most of them adopting intermediary outcomes. This highlights the relevance of the topic and the need to expand the evidence base.
- 4. Subnational governments and education administrators:** Robust impact evaluations on incentives for subnational governments and education administrators (also called meso-level actors) is limited and mostly focus on interventions that combined meso-level incentives with teacher incentives or pedagogical support. Although they have found positive significant effects on student learning, the limited number of analyses prevents broader conclusions. A highlight is the state of Ceará, in Brazil, with a long-lasting incentive policy to local governments that is aligned with other education provisions such as technical and pedagogical support, adequate instruction material and training, and strengthened school management.

1. Introduction

Despite a significant increase in education investment in the last decades, learning outcomes are still low. Improving access to education and learning are key goals of the international community, and important ventures were made on this effort (United Nations, 2015; World Bank, 2018). In the last decades, national governments and development organizations significantly invested in education, such as expanding access, building and improving schools' infrastructure, establishing monitoring systems, providing textbooks, and teacher training. Yet, many challenges remain to leverage learning outcomes, as education systems face several constraints beyond the lack of inputs, as spending inefficiencies and inequalities (Al-Samarrai et al., 2021; Glewwe & Muralidharan, 2016; World Bank, 2018).

The current economic and education scenario arising from the COVID-19 pandemic puts even more pressure on efficiency in education spending. The COVID-19 pandemic impacted public finances dramatically, decreasing governmental revenues and redirecting funds from other areas to health and social security. In low- and lower-middle-income countries, 65% of them have cut their education budget since the offset of the pandemic (Al-Samarrai et al., 2021). Additionally, school closures and deficiencies in remote learning provision generated a massive learning loss, requiring immense efforts to recover (World Bank et al., 2021). These will also entail financial commitment to strengthen education systems' capacity, increasing the value of education's money. Therefore, the current analysis of RBF literature can provide relevant insights to countries that face pressure towards more efficient use of resources to assure children the right to education and foster human capital.

RBF arrangements have gained relevance with the expected potential to impact education service delivery positively. As education provision is fragmented among different actors (such as teachers, textbook providers, and education bureaus that establish curricular and learning assessment standards), it is difficult to make all stakeholders accountable to their role and education's ultimate goal: learning. In this context, RBF is conceived as a tool to strengthen accountability and service delivery by promoting a focus on actions necessary to improve learning, fostering local solutions to common education challenges, and reinforcing the capacity of education systems to measure and track progress. In the context of this report, results-based financing is conceived as an umbrella term referring to any program or intervention that provides rewards to individuals or institutions after agreed-upon results are achieved and verified. The definition of "result" varies per intervention, with some adopting intermediate outcomes (as teacher attendance) or final outcomes (as learning). Rewards can be monetary or non-monetary and are frequently partial (such as a salary bonus or one specific grant, among other revenues received by schools).

This report analyzes how the recent literature contributes to the knowledge base of RBF in education. Research on the theoretical basis of results-based financing have been addressed by other authors (Birdsall & Savedoff, 2011; Clist, 2019; Clist & Verschoor, 2014). The current study builds upon Lee and Medina (2019) literature review of research published between 2000-2018, and adds recent evidence published between 2019-2021.

The search process has focused on the 2019-2021 time window and looked strictly at results-based financing mechanisms implemented in primary and secondary education in developing countries. The papers were retrieved through bibliographic databases, electronic search engines, past reviews and snowballing, and expert recommendations. The search process identified 197 papers, from which 92 were included. Out of those, 16 are mentioned very briefly as most of them address an intervention that is tangent to our study. The analysis is, therefore, narrowed to 76 papers. The inclusion criteria considered studies that convincingly established causality between the intervention and outcomes either through a randomized control trial (RCT), regression discontinuity design, or difference-in-differences. The decision to focus on these methodologies lies in the desire to draw policy recommendations with a high degree of confidence on the intervention's ability to have an impact and the potential impact size on education outcomes. In the sections related to incentives for schools and meso-level actors, operational reports were also mentioned to provide insights on design and implementation, given the limited number of robust evidence. Following the 2019's report, the focus on developing countries aims to clearly understand RBF effectiveness in contexts of constrained resources. Regarding the characteristics of the studies assessed, they include RBF interventions in 24 countries, most of them in Latin America and Asia (figure 1). Most incentives target primary and lower secondary education outcomes (figure 2).

*Figure 1 – Country distribution of studies analyzed in this report.
The countries with the most papers are Brazil (7), Colombia (6), Indonesia (6), and Mexico (6).*

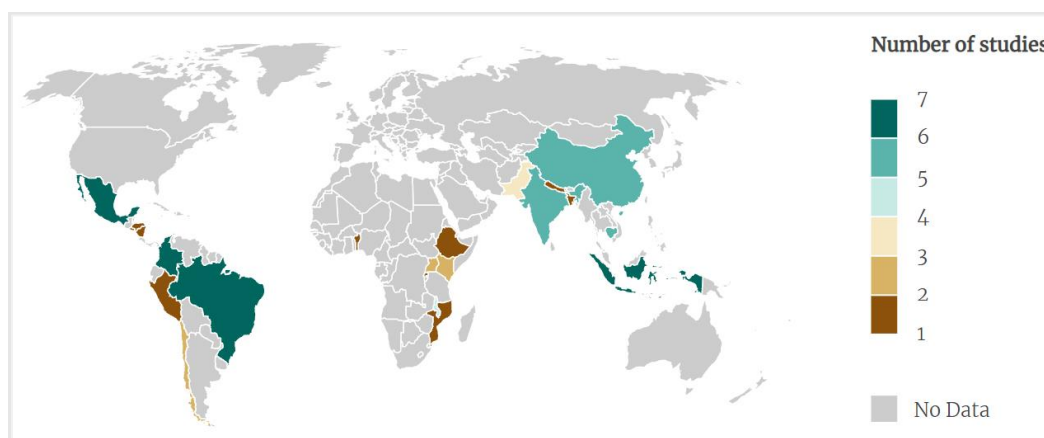
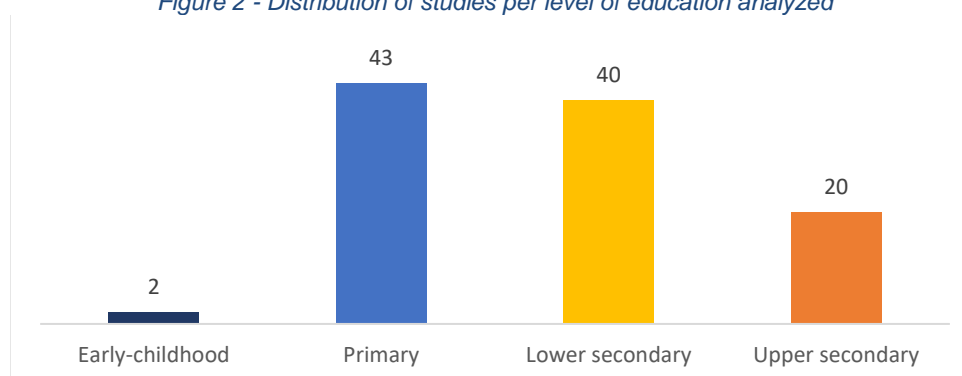


Figure 2 - Distribution of studies per level of education analyzed



Evidence is discussed according to the actor incentivized and the main expected outcome. Following Damon et al. (2016a), this report organizes interventions according to the incentivized actor – teachers, students and families, schools, and meso-level stakeholders – and the primary expected outcome, as incentives’ design greatly vary according to these two characteristics. Incentive mechanisms are grounded on the principal-agent problem, designing a structure that motivates the agent to achieve the optimum output or outcome desired by the principal and aligning priorities between them (Mitnick, 1975). Table 1 presents a framework of education RBFs, identifying in which levels they occur, the main actors involved, and what instruments are put in place. Incentives for students and families frequently take the form of conditional cash transfers - aiming to improve enrollment, attendance, and completion rates - or merit-based rewards to foster student learning. Incentives for teachers try to stimulate teachers’ attendance, to focus on student learning, and less often to select teachers in recruitment processes. Some countries have also introduced school rewards based on performance indicators to drive quality improvements. In decentralized countries, governments experimented linking funding to subnational levels (as mayors or regional education bureaus) according to their education performance. Finally, development agencies started to attach part of disbursements to the achievement of pre-established performance targets.¹

Table 1 - Framework of RBF in education: levels, actors, and instruments

Level	Who is incentivized? (Agent)	Role	Sample constraints	Who incentivizes (Principal)	RBF Instruments
1	National Government	Designers and managers of the system	Monitoring and providing adequate support to service delivery	Development organization	Results-based aid or loan
2	Meso-levels (subnational governments or regional/local education bureaus or supervisors)	Co-responsible for service design and management	Monitoring and providing adequate support to service delivery	National Government (sometimes development organizations)	Performance-based transfers, Output-based disbursement, bonus pay
3	Schools and service providers (public or private) or Investor	Managers of front-line service delivery	Leading school staff, managing inputs	Donor / National government / Local government	Performance-based contracts, school grants, school vouchers ² , impact bonds

¹ RBF for governments has been further explored by a recent World Bank report (Dom et al., 2021) and therefore is not covered in the current review. Other complementary reports are Schipper & Pradham (2022) on the impact of education RBFs on inequality of outcomes, Social Finance (2022) on lessons from RBF in health that can be applied in education, Terway et al (2021) for a qualitative study of incentives for meso-levels and Elsby, Smith, Monk and Ronicle (2022) on impact bonds.

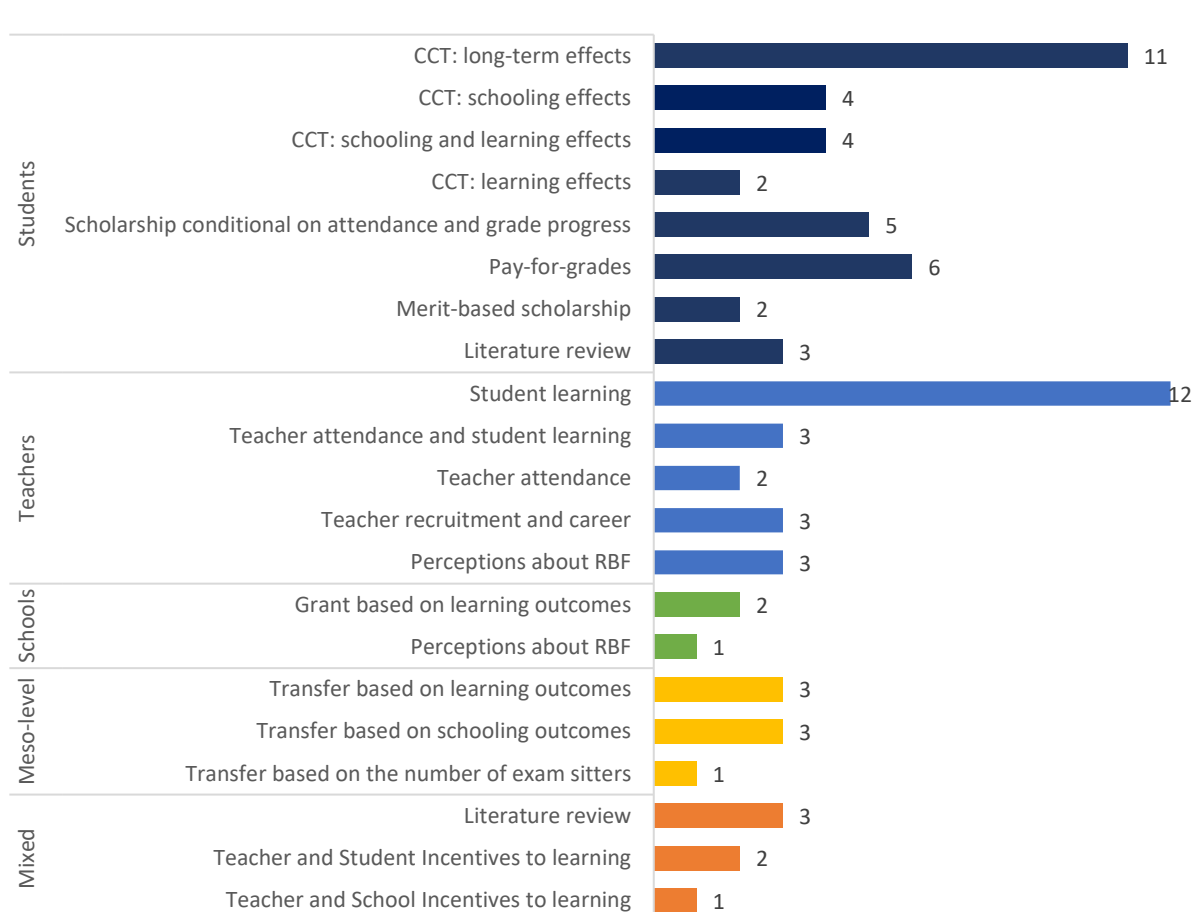
² Following the first report produced by Lee & Medina (2019), the current analysis does not address school vouchers since their primary purpose is expanding education coverage more than improving outcomes. It also does not cover impact bonds, that consists of an agreement between a donor/government agency and the investor, whereas the service provision is under a third-party responsibility (as a school or a local NGO) who may or may not be incentivized - as those have been assessed in a separate piece work Elsby, Smith, Monk and Ronicle (2022).

4	Teachers or school directors	Service delivery agents	Attending school, effective engaging in teaching	Government / School	Performance pay or bonus pay
5	Beneficiaries (students and households)	Service user	Attending school, learning	Government / School	Conditional cash transfers, merit-based incentives

Source: Adapted from Terway et al. (2021) and Lee & Medina (2019).

RBF in education is still largely concentrated on teachers and student-family incentives. Figure 3 presents the distribution of studies by actors incentivized and type of intervention, highlighting the relevance of teacher and student incentives to the evidence base. Newest research on these topics sheds more light on RBF's contribution to student learning and the long-term effects of interventions. Design and implementation aspects continue to have a great relevance for the success of initiatives, something that is almost intrinsic to public policy. Different interventions show that aligning incentives of more than one agent involved in education can be more effective than a scheme targeting a single actor. Likewise, combining incentives with appropriate support helps to engage actors, especially those departing from low baseline outcomes. The following sections dive into each type of incentive, discuss the key factors to policy-making and concludes with a summary of the main findings.

Figure 3 - Distribution of studies by actors incentivized and type of intervention



2. Teachers

2.1. Why incentives for teachers?

Paying teachers for their performance is a strategy to improve education outcomes by encouraging them to change their behavior, whether it is to attend school regularly, or to focus on supporting low-achieving students. Teachers are decisive agents in education, and play a critical role in helping children achieve relevant skills. This is not an easy task and requires a fertile learning environment, which involves having nutritional needs fulfilled, access to appropriate instruction material, good pedagogical instruction and adequate support from teachers (World Bank, 2018). For years, policy-makers have focused on education conditions, especially by improving learning inputs (school infrastructure, textbooks, hiring more teachers, etc). Nevertheless, learning outcomes greatly depend on teacher-student interaction, and there are still many barriers to an effective interplay. For example, high rates of teacher absenteeism reduces children's time exposed to instruction (Abadzi, 2007; Chaudhury et al., 2006; Unesco, 2014). Non-instructional tasks, such as taking attendance, cleaning the blackboard, and distributing papers, frequently reduces learning time (Abadzi, 2007; Bruns & Luque, 2015). Pedagogical practices focused on memorization, teaching only to high-performing students, and shortages in teachers' skills decreases learning efficiency. One way that countries and development partners have sought to improve teacher-student interaction is through results-based financing, which tries to stimulate teachers' effort or performance and raise accountability towards learning. There are, however, multiple ways of motivating individuals, with extrinsic and intrinsic motivators. Pay-for-performance schemes may also direct teachers' attention towards low-achieving students when the formula used for rewards gives more weight to those pupils. From a policy perspective, results-based incentives also represent an approach to motivate and reward better teacher performance without conducting substantial changes in pay-scale, which, due to its rigidity, requires more orchestration (Breeding et al., 2021).

2.2. What results can teacher incentives achieve?

Most of the evidence on incentives for teachers shows positive outcomes. There is evidence that incentives designed to improve teacher attendance can raise attendance but positive effects rely on strong monitoring systems. Likewise, 12 out of 15 interventions that directly incentivized student learning have shown positive and significant effects. A qualitative analysis of programs' characteristics indicates that interventions tend to be successful when students have minimum adequate conditions, the reward rule clearly puts learning as the main outcome indicator, and teachers understand the mechanism well enough to reshape pedagogical strategies and improve its performance. Design and implementation matter, and, in both stages, it is necessary to have the technical requirements, resources, and political will (Breeding et al., 2021). So far, teacher incentives have been largely implemented as a bonus policy instead of a change in teachers' contracts. Pay for performance

contracts can sort teachers that are more responsive to incentives but evidence on selecting more qualified teachers is mixed.

An analysis of the impact of teacher incentives should consider what types of results are being rewarded and the purpose of the intervention, which can be improving teachers' attendance, raising student learning or sorting better teachers. While most teacher incentives are designed as bonus programs, they differ in the purpose of the intervention (i.e.: improving attendance, learning or teachers' selection) and the focus of the reward (which can incentivize teachers' efforts or performance). Since effort and performance are difficult to measure, the incentive's design also depends on what metric is available. The most common metric used to measure teacher effort is attendance, and, for teacher performance, it is student test scores. Breeding et al (2021) constitutes the most recent literature review of teacher incentives. The authors analyzed to what extent a heterogeneous group of interventions (focused on improving teachers' attendance, student learning and teachers' selection) could lead to improved learning. This paper does not seek to duplicate that work, and thus, the focus here will be on categorizing interventions according to their primary purpose. The rationale for this approach is that variations in outcomes may also be driven by variations in design, thus one should consider the theory of change behind each design, and identify whether the incentive (i) achieves its primary purpose, (ii) leads to better learning or improved conditions for learning, and (iii) if there are other sort of impacts (both positive or negative). The following paragraphs present the evidence according to interventions' purpose, grouping papers on three categories: incentives focused on teacher attendance, student learning or selecting better teachers.

2.2.1. Incentives focused on teacher attendance

Evidence from primary schools indicates that teachers respond to incentives to higher attendance, but the verification mechanism and the weight given to attendance in the reward formula seem to be decisive for incentives' effectiveness. Two interventions focused exclusively on raising teachers' attendance and had a positive impact, while five programs attendances as part of the reward formula but had mixed effects. These divergent outcomes seem to be related to which accountability mechanism is established and the relevance of attendance in the bonus formula. When tamper-free timestamps cameras were introduced in rural Indian classrooms,³ it reduced absenteeism rates by 21 percentage points (Duflo et al., 2012). In Uganda, an intervention focused on attendance provided incentives to both teachers and principals, asking the latter to report teacher absenteeism. It successfully increased teachers' presence by eight percentage points (Cilliers et al., 2018). In contrast, multi-purpose interventions only seem to positively impact attendance when there is an explicit focus on it. In Indonesia, attendance was part of a multi-dimensional set of indicators monitored by the community but only had an effective impact on learning in the treatment group that explicitly measured it by cameras (Gaduh et al., 2021). In Rwanda, an intervention funded by REACH that explored how teachers reacted to pay-for-performance contracts linked 25% of the rewarding rule to attendance (measured during surprise school visits) and increased it by 8% (Leaver et al., 2021). Meanwhile, programs less centered on teachers' presence did not impact this indicator. A multi-

³ In the intervention, students were instructed to take pictures of teachers at the beginning and at the end of the school day.

dimensional intervention in Kenya combined teachers' training, learning material, and incentives but did not affect attendance or student learning. The authors suggest that the incentive did not work because headteachers underreported teacher attendance (Chen et al., 2001). A bonus program in Brazil defined that only teachers with a minimum of 65% of attendance rate were eligible to the bonus, but the evaluations did not report attendance outcomes, and principals' perceptions on attendance have not changed over the years (Lépine, 2021). Two other incentives focused on student learning collected data on teachers' presence through unannounced school visits and did not find differences in attendance (Filmer et al., 2020; Muralidharan & Sundararaman, 2011). While strong monitoring systems seem effect to raise attendance, their introduction may face opposition by teachers, making it politically difficult to implement.

Regarding other education outcomes, incentives exclusively focusing on attendance seem to positively impact student learning, enrollment rates, and dropout, but the evidence is still limited.

The intervention in India improved learning outcomes by 0.17 standard deviations (SD), although researchers made an important observation: during their visits to schools, they noticed that present teachers were as likely to be teaching in treatment as in control schools. This suggests that teachers converted the additional time gained from improved attendance into more hours of instruction (Duflo et al., 2012). The program in Indonesia also raised learning outcomes by 0.2 SD (Gaduh et al., 2021). In Uganda, the intervention had substantial and significant impacts on dropout and attainment, suggesting that teachers' presence motivated students to continue their studies. Schools with the incentive had 8% more students enrolled and, two and a half years later, the percentage of cohort students tracked (which can be seen as a proxy of attainment) was 14 percentage points higher than in control schools. As a consequence of these large differentials in retention rates across treatment arms, it was not possible to assess learning outcomes (Cilliers et al., 2018).

2.2.2. Incentives focused on student learning

From a sample of 15 interventions whose primary purpose was improving student learning, 12 showed positive and significant increases in test scores. Effect sizes range from 0.09 SD to 0.57 SD, as presented in Annex 1.⁴ This doesn't mean that incentives always work, but it shows a positive trend. While this finding shows a larger effect size than Breeding et al (2021), the main difference appears to be in the sample composition. Also with a sample of 15 papers, the former review did not include pay-for-percentile schemes (Chang et al., 2020; Gilligan et al., 2019; Loyalka et al., 2019; Mbiti, Romero, et al., 2019), and included interventions whose primary outcome was selecting better teachers (Bo et al., 2013; Cabrera & Webbink, 2016; Leaver et al., 2021), improving teachers' attendance (Duflo et al., 2012), testing the impact of unconditional salary increase (Chewla et al., 2019) or investigating teachers' beliefs (Sabarwal & Abu-Jawdeh, 2018). These differences would probably change median effects in relation to this review since teacher incentives focusing on attendance or selection have

⁴ Positive outcomes are found in (Behrman et al., 2015; Chang et al., 2020; Contreras & Rau, 2012; Filmer et al., 2020; Gaduh et al., 2021; Gilligan et al., 2019; Glewwe et al., 2010; Loyalka et al., 2019; Mbiti, Muralidharan, et al., 2019; Mbiti, Romero, et al., 2019; Muralidharan & Sundararaman, 2011; Oshiro & Scorzafave, 2015). The current review did not harmonize the estimates reported in each paper to get into an average effect size, as they vary in their metrics. Some report as standard deviations, other in percentage increase. Moreover, some report effects only for language or mathematics, others for both subjects. Details in Annex 2.

lower impact estimates for learning, as they are second-order effects. The current review, in contrast, considers only interventions whose primary focus is improving student learning and includes research published in 2021 (Bellés-Obrero & Lombardi, 2021; Gaduh et al., 2021).

Most of the evidence consists of randomized control trials (RCTs) with small sample sizes, designed by researchers and implemented by local NGOs, although four studies analyze large-scale interventions, all in Latin America. These interventions are treated separately, given the specificities of small- and large-scale programs. Moreover, among RCTs, three main groups were identified: (1) initial interventions that tested if simple incentive schemes focusing on student learning gains could impact learning, (2) attempts to combine teacher incentives with other interventions, and (3) comparisons of metrics to identify the most effective interventions.⁵ Findings are discussed in the following paragraphs. Overall, incentives can improve learning but positive outcomes seem to be related to schools having minimum learning resources, having a reward rule that clearly puts learning as the main outcome indicator, and when teachers are able (or supported) to articulate clear and specific targets for their students, developing a strategy to improve learning outcomes.

Two experiments, in Kenya and India, revealed that student learning is more meaningfully assessed through low-stakes tests. A program run in rural Kenya provided group incentives to teachers and head-teachers (in-kind gifts) conditional on student learning improvements.⁶ It raised test scores by 0.14 SD on a government issued exam, but a low-stakes test conducted by the research team showed no impacts on learning. An analysis revealed that two-thirds of the improvement in the high-stakes exam came from increased student participation and one-third from improvements in narrow test-taking skills, indicating that teachers were “teaching to the test” (Glewwe et al., 2010).

In India, an intervention in primary schools had positive effects on student learning as measured by low-stakes tests in Mathematics and language (the two subjects that had incentives attached to student scores) and Sciences and Social Studies (which did not have incentives). At the end of two years, the intervention had a mean treatment effect of 0.22 SD in Math and language (Muralidharan

⁵ As explained in Lee and Medina (2019), incentives for teachers can reward based on individual or group performance: (i) group incentives may favor cooperation between teachers since student learning is beyond teachers’ control, but also induce free-riding which may reduce incentive’s impacts, (ii) individual incentives connect with a person’s effort, but may be challenging to scale-up since it requires calculating the reward for each teacher. On the metrics, rewards can consider levels, gains or rank. (1) Levels provide an objective measure to teachers, which may help them to design strategies to improve student learning. It can involve a single target, as the number of students that passed an exam, or a multiple threshold, as learning scales. The disadvantage is that teachers have an incentive to focus on students that are closer to meeting the target(s), which may jeopardize low-achieving pupils far from the threshold(s). (2) Gains try to favor the opposite, since students with very low levels will have a greater percentage increase than students with already good performance, but once students start improving their performance, the incentive may become less attractive. (3) Rank reward the top performers, usually a defined number, which makes it cost-predictable. (4) Pay-for-percentile is a hybrid method that conciliates gains and rank features. Students are tested at the beginning of the year and gathered according to their levels. At the end of the year, students are tested again and teacher bonuses are calculated based on students’ ranking position.

⁶ The program took place between 1998-2000 and focused on primary and lower secondary schools. The governmental test was multiple-choice while researchers applied fill-in-the-blank questions to cross-check student learning.

& Sundararaman, 2011). After five years, the cohort of students exposed to the program during the whole primary cycle scored 0.54 SD higher in Math, 0.35 in language, 0.52 in sciences, and 0.3 in social studies compared to control schools (Muralidharan, 2012). The experiment divided treatment schools into two groups: one rewarded teachers based on their individual performance and the other based on the average performance of a group of teachers. In the short run, the group and individual incentives had similar impacts, but, at the end of the fifth year, only individual incentives had positive and statistically significant results. The main mechanism driving results seems to be improved teacher effort (i.e., assigning more homework and classwork, supporting low-performing students, and conducting extra classes) among those teachers with regular attendance. The study revealed evidence that the partner NGO in charge of implementing and monitoring the intervention provided regular performance feedback to teachers, through regular visits to schools to collect process indicators and observe classes. This suggests that incentives combined with performance feedback could have a significant role on improving teachers' effort and accountability.

Evidence from Mexico, Tanzania and Indonesia looked at whether combining teacher incentives with complementary interventions could boost learning outcomes. The research found that:

- **Combining incentives for teachers and students seems to have positive results but also to generate dysfunctional responses.** Experiments in Mexico and Tanzania investigated the impacts of providing monetary incentives only for high school students, only for teachers or for both. In Mexico, the intervention focused on incentivizing students to improve their mathematics scores and the combined treatment also included head teachers. It elicited strong cheating in the exams among students that were being incentivized. Using low-stakes tests, it found that incentives only improved test scores among the group in which both students and teachers were incentivized (Behrman et al., 2015).⁷
- In Tanzania, the program consisted of a learning gains tournament in language and mathematics in grade 10 and gave in-kind rewards. After the first year of intervention, positive impacts were found only for the combined scheme (average treatment effect of 0.15 SD), while in the second year, only for the teachers' group (0.13 SD). The reasons for this variation are unclear but the evidence suggests that teacher incentives have the potential to impact learning. Nevertheless, impacts were concentrated in schools and students with higher baseline performance, although the study could not identify the triggers of learning inequality (Filmer et al., 2020).⁸
- **Combining teacher incentives to improve student learning with school inputs and social accountability seems promising:** Another experiment in Tanzania, this time funded targeting primary schools, explored if teacher incentives combined with unconditional grants to schools could increase learning. They found complementarities in providing incentives and grants.

⁷ Using high-stakes test scores, the intervention found substantial effects of the combined scheme (0.57 SD), no evidence for teacher incentives-only (0.04 SD) and positive effects of student incentives-only (0.23SD). Nevertheless, the authors suggest that results may be overestimated due to different student effort in the test. Students belonging to an incentivized group would put more effort on the exam (and have more attempts to cheat) than students in the control and in the teacher's incentive group. Indeed, when comparing results from high- and low-stakes exams for a subsample, only the combined scheme had positive effects.

⁸The reward consisted of in-kind prizes (smartphones, book vouchers, certificates and medals). Prizes were given to the first three collocates, based on a rank of annual average gain in student test scores. Incentives did not have an effect on teacher or student attendance.

Teacher bonuses alone raised test scores by 0.21 SD as measured by high stakes exams, but combining them with grants led to an increase of 0.36 SD. In low-stakes tests, only the combined scheme had a positive impact on learning of 0.23 SD. Findings point to the relevance of education resources on learning outcomes: in the group with only the bonus, test scores increased among schools that had some resources to work with, while in the combined scheme positive outcomes were seen throughout the distribution. When comparing schools with bonus and grants to schools with only grants, the former had higher net and per-student expenditure, deciding to invest the grant on textbooks in the first grades of primary education (the incentivized grades), while the latter opted for reducing tuition fees (Mbiti, Muralidharan, et al., 2019).⁹

- In Indonesia, an intervention tried to improve learning through social accountability and teacher incentives. The community agreed on a set of indicators relevant for student learning, and a committee was responsible for monitoring them. Treatment schools were divided into three groups. All of the groups had community monitoring, one group had monetary incentives attached to teachers' performance (measured by the committee) and another group had incentives attached to teachers' attendance (measured through cameras). After one year, all three groups showed positive learning outcomes, with the highest increase in the group with incentives for attendance (0.2 SD). The community monitoring group improved learning in 0.08 SD and the incentives for teachers' performance in 0.11 SD. The lower effectiveness in the incentives to performance seems to be related to tensions between teachers and the committee, with the former pressuring the latter to give them positive evaluations. One year after the intervention ended, impacts on learning outcomes and parental investments persisted for the group with community monitoring and cameras, although external engagement and supervision did not persist into the second year (Gaduh et al., 2021).
- Taken together, evidence from Tanzania and Indonesia suggest that, depending on the setting and resources available in schools, there are multiple binding constraints to improve learning, thus, incentives to teachers should be complemented with other interventions to boost impact. On sustainability, the Indonesian experience, analyzed in conjunction with the Indian intervention that lasted 5 years (Muralidharan, 2012) suggests that even instruments designed to be sustainable require an accountability routine to maintain effectiveness. In public policy, this can be achieved through a cascade of accountability schemes that keep schools and the community engaged (see the cases of Sobral and Ceará on that (Costa & Carnoy, 2015; Cruz & Loureiro, 2020; Lautharte et al., 2021; Loureiro et al., 2020; Rocha et al., 2018).

In China, Uganda and Tanzania, experiments explored nuances in the metrics of individual teacher incentives, especially pay-for-percentile, but achieved mixed results. Multiple studies have shown that incentive design can shape teachers' behavior, and two key aspects explored in experiments refer to individual versus group incentives and the metric used to calculate the rewards (see Annex 2 for details). Metrics can channel teachers' focus on the group of students they perceive will help them achieve greater rewards, but teachers' response depends not only on the design but also on their understanding of the incentive and how they behave in order to improve student test scores.

⁹ This intervention was partially funded by REACH. The incentive was considered a threshold scheme, providing ca US\$3 for each student who passed an external test in numeracy, English and Kiswahili. Headteachers received ca 0.6 USD for each subject test a student passed.

- **In China, pay-for-percentile has worked with 5th and 6th-grade mathematics teachers.** An experiment tested three schemes of rank-order tournaments (based on levels, gains, and pay-for-percentile).¹⁰ They found that pay-for-percentile incentives increased student achievement by approximately 0.15 SD, with similar size gains for all students across the baseline achievement distribution, including students who are traditionally neglected. Levels and gains schemes had no significant effects (Loyalka et al., 2019). In a subset of the same sample, Chang et al (2020) tested a pay-for-percentile bonus scheme with 5th-grade math teachers. The incentive provided additional rewards (60% more) for improved results of students at the bottom 30% of the distribution in the baseline test. The intervention improved learning by 0.10 SD on average and by 0.15 SD among the bottom 30%. Results seem to be driven by improved curricular coverage and teacher behavior, such as investing more time and energy in raising student outcomes, improving teacher-pupil interactions, reducing teacher absenteeism and giving more math homework.
- **But in Uganda, pay-for-percentile with 6th-grade mathematics teachers did not yield improvements.** This particular experiment had only one treatment, which consisted of a pay-for-percentile scheme that incentivized Uganda teachers to improve student learning and reduce dropout. Overall, the program did not increase learning but improved attendance rates by four percentage points. The analysis indicates that results were driven by schools that provided math books to students. Moreover, students with good performance at the baseline and who had access to textbooks had positive and significant improvements in test scores in the final exam, especially in questions with grades 4-6 content, which were addressed in the textbook (Gilligan et al., 2019). This result may reinforce the findings from the experiment that combined school grants with teacher incentives in Tanzania that other constraints to learning should also be addressed.
- **In Tanzanian primary schools, a threshold scheme was more effective than pay-for-percentile.** An experiment in Tanzania, with REACH funding, focused on primary mathematics and language teachers and found less favorable results for pay-for-percentile. Comparing it to multiple proficiency threshold schemes, rewarding teachers according to student learning levels led to a higher impact (0.19 SD in language and 0.14 SD in mathematics) than pay-for-percentile (0.10 SD in language and 0.09 SD in mathematics). The treatment effect of the levels group was 0.09 SD higher than the pay-for-percentile, a difference that is statistically significant at 10 percent level. Additionally, the levels design led to learning gains more equitably distributed across students, while the pay-for-percentile led teachers to focus on the best students (Mbiti, Romero, et al., 2019).
- **What explains the varying success of pay-for-percentile schemes?** There are some hypotheses for the mixed effects of pay-per-percentile but further research is needed to assess this. One possible explanation is teachers' level of understanding of the mechanism and ability to develop strategies to leverage learning. In Tanzania, the study identified that, although teacher

¹⁰ The levels scheme considered the class average achievement. The gains scheme ranked classes by average learning gains from the start to the end of the school year. The pay-per-percentile considered teacher's percentile performance index (the fraction of contests that students of a given teacher won compared to students taught by other teachers and yet began the school year at similar achievement levels).

comprehension in both schemes was high, teachers in the pay-for-percentile group expected to earn a lower bonus than what they actually received, perhaps because they were not able to foresee how their efforts would be rewarded (for uncertainty or complexity in the scheme), which reduced their responsiveness. In the levels group, in turn, teachers could better articulate clear and specific targets for their students, developing a strategy to improve learning outcomes (Mbiti, Romero, et al., 2019). Another explanation could be in teachers' backgrounds. Although the authors did not report data on teachers' perception of the subject being taught, it could be the case that math teachers better understand the pay-for-percentile scheme, since positive evidence from pay-per-percentile schemes happened to be among math teachers (Chang et al., 2020; Loyalka et al., 2019). In Tanzania, where the intervention focused on both language and mathematics, the threshold scheme outperformed pay-for-percentile. A third hypothesis is the context. In China, the education system is very competitive and there is already a "reward culture" among primary and secondary teachers (Chang et al., 2020, p. 1). Lastly, it would be interesting to assess the feasibility of scaling up a pay-for-percentile scheme since, as indicated in the following section, all programs implemented at scale so far adopt simpler designs.

Programs implemented by governments at large scale usually include multiple outcome indicators and are designed as group incentives. They have mixed effects. Out of the five programs implemented by governments, four are large scale Latin American interventions that adopted group incentives.¹¹ A key advantage of group incentives is that they are easier and cheaper to administer, since they consider data at the school level, while individual incentives require classroom-level management. Another common feature of the interventions implemented by governments is that they all include more than one outcome indicator, which might disperse teachers' efforts. Given that all scaled-up programs cover almost all teachers, quasi-experimental approaches are adopted for the analyses.

A key challenge in large-scale incentive interventions focused on increasing student learning is accounting for heterogeneity in the system. Even successful programs have difficulties finding the best balance.

- **A school tournament in Chile is based on a composed index that balances learning levels and gains, equality of opportunity to students, and school metrics. Still, the bonus incentivizes only a subset of schools with better performance.** A teacher bonus program for public and private-subsidized schools was introduced in Chile in 1996, reaching 90% of enrolled students from primary education to high school. It consists of a tournament that rewards the top 25% of schools in each category. Categories consider schools' geographic and socioeconomic characteristics and the education level provided. A group-based monetary reward is given to the top schools, which are measured by a 6-component national index that accounts for learning levels and gains (65%), school measures (innovation and working conditions), and equality of opportunities (retention rates, passing rates, inclusion of deficient students). The impact evaluation found an average treatment effect on the treated of 0.19 SD on learning outcomes in the first two rounds, using high-stakes test scores from the national standardized exam. Despite the positive impact, the

¹¹ It is interesting to note that the programs in Mexico and Chile were introduced in 1993 and 1996, even before the first experiment in Kenya (1998).

analysis indicated that incentives seem to affect schools above the 60th percentile (Contreras & Rau, 2012). Nevertheless, the program was created with a twin-objective – leverage performance and disclosure performance data to the community – and on its second purpose, it succeeded.

- **In the state of São Paulo, in Brazil, both school staff and regional supervision units are incentivized. There is also a specific rule that only applies to top-performing schools.** The bonus policy was established in 2008 for primary and lower secondary schools, and adopted a piece-rate design in which all schools that achieved the target would receive a reward. It is a group-based incentive to teachers and principals that attaches the reward to improvements in student learning and teacher/school staff attendance.¹² This is an interesting design since it tries to reconcile individual and group incentives. The policy also considered a bonus to staff from regional supervision units proportional to the regional average to strengthen monitoring. There is also a special design for top schools since their results are more difficult to improve (they can receive the bonus if they are listed among the top 10% in the last 2 years). The program was evaluated in the short- and medium-term using low-stakes exams (the national standardized test, while the bonus is based on the state’s standardized test). One year after its implementation, the program had an average effect of 0.28 SD on learning for 5th graders, but this effect faded out in the third year. There were no effects for 9th graders. These results may be explained by difficulties in adjusting schools’ learning targets since the distribution of schools’ results has two modes, 0%, and 120%, for both grades 5 and 9, and with large variation between one year to another. This suggests that the targets are too hard for some schools, and for others, they are too easy to achieve (Oshiro & Scorzafave, 2015). As the bonus policy continued to be implemented, it could affect teachers’ behavior by allowing them to improve over time. Another evaluation conducted seven years into implementation found positive and consistent learning gains in the 5th grade (ranging from 0.06 to 0.29 SD depending on the subject and control group), but for 9th graders, gains are modest and sensitive to the control group adopted. The study also did not find evidence of free-riding effects. The difficulty in improving results for grade 9 may be related to the fact that lower-secondary classes have more teachers, which makes coordination within schools more difficult, and students in grade 9 accumulate learning gaps that are more difficult to close (Lépine, 2021).

Programs that were less effective at improving student learning vary greatly in design. Still, evidence suggests that a common thread is that they did not concentrate efforts on improving student outcomes.

- **The incentive program in Peru includes more than five indicators, which can disperse teachers’ efforts.** A nationwide bonus program implemented in 2015 in lower-secondary schools¹³ rewards

¹² Learning targets are set at the school level based on OECD levels. Only teachers with at least 65% of attendance are eligible for the bonus. Unfortunately, data on teacher attendance was not included in the evaluation, but principals’ perceptions about teacher attendance (measured by survey) does not seem to have improved with the bonus policy.

¹³ There was already a rewarding scheme in place for primary schools which seem to have positive outcomes according to a report commissioned by the Ministry of Education (<http://repositorio.minedu.gob.pe/handle/20.500.12799/6054>). We did not include the paper in this review since it was not peer-reviewed and the identification strategy considers PSM only.

teachers and principals based on a multidimensional index. The index gives 40% of weight to students' test scores in a standardized exam of grade 8, 35% to the school's intra-annual retention rate, and 20% to a school management index, which includes teacher attendance and compliance with class hours, learning environment, and other indicators. Schools compete in specific groups according to their district, school-day length, and urban/rural location. Those ranked among the top 20 percent receive a generous bonus. An analysis of the program's first year found no significant results and did not provide robust evidence to explain the lack of success (Bellés-Obrero & Lombardi, 2021). Despite that, some potential factors are raised. First, the incentive design included several indicators, dispersing schools' efforts. Second, teachers had no experience with the standardized exam, as it was introduced in the same year of the incentive mechanism. Third, the evaluation used schools' internal learning assessments to investigate program's impact, rather than a standardized test applied to all schools.

- **In Pakistan, a pilot had little variations in payments if teachers improved student enrollment, student test-taking, or test results, which likely influenced teachers to focus on indicators that were easier to increase.** In Punjab, Pakistan, a bonus program implemented by the government targeted schools at the bottom of the learning distribution. It was aimed at raising teachers' efforts towards student learning and stimulating schools to raise student enrollment. The bonus is a group incentive with three components: average school scores from 5th graders (gains), student enrollment, and student participation in the test. In the third year of the intervention, the program increased test participation rates and student enrollment in grade 1 but had no effects on learning.¹⁴ Complementary analysis suggests that the test was not reliable, and that inter-cohort variation affected test scores' comparability. Additionally, the bonus formula implied payment for almost all teachers, with little variation even if improvements corresponded to learning, enrollment, or attendance on the exam. Thus, teachers may have directed their efforts towards the components with larger payoffs such as sitting an extra student in the exam (Barrera-Osorio & Raju, 2017).
- **In Mexico, a bonus scheme was introduced as part of a teacher career reform, and student test scores corresponded to only 20% of the index.** Introduced in 1993, the program provided a bonus to teachers based on a 6-dimensional assessment (education degree, seniority, peer feedback on performance, teacher scores on two tests, and student scores). Student incentives corresponded up to 20 out of 100 points and the cutoff was around 70, thus teachers could get promoted even if they did not improve student test scores. Interestingly, the bonus had a permanent status: once promoted to a given level, teachers could not be demoted and continued to receive the bonus throughout their career, which corresponded to roughly 20 to 200 percent of a salary base. Moreover, teachers had to wait 2-4 years to try for another promotion. Given the program's design, it is not surprising that it had no impacts on student learning: test scores had low weight in determining the bonus, cutoff varied by state and by year, reducing predictability, and the permanent status of the reward, added to the fact that teachers could not move to another level

¹⁴ The experiment design considered 3 treatment groups: (1) bonus to headteacher, (2) bonus to teachers, (3) different levels of bonus offered to teachers and headteachers. Effects on exam participation rates are very similar across treatments, with a tiny increase for the group that combined incentives for teachers and headteachers. For instance, in year 3, treatments 1 and 2 increased participation rates in 11-12 percentage points (pp) while group 3 increased in 15pp.

before two years, further diluted the importance of student learning in the bonus incentive (McEwan, Patrick J. & Santibanez, Lucrecia, 2005). This experience suggests that, while it would be nice to include accountability towards learning as part of a teacher's fixed-wage scale, the necessity to negotiate changes in wage scale with teacher unions and other representative institutions may result in cuts in the performance part, which reduces incentives' power to change behavior.

In terms of large-scale, government led interventions, three main lessons emerge:

- **First, given the heterogeneity among schools, it is critical to design a mechanism in which all schools perceive fair competition.** In general, piece-rate mechanisms seem to be more appropriate than tournaments, but other designs could be explored. For example, the state of Ceará has a school-level tournament in which winning schools must partner with low-performing schools and only receive the full prize after the lower-performing one has raised its learning outcomes (Goldemberg et al., 2021).
- **Second, although education systems face multiple challenges, including too many indicators in the reward metric may disperse efforts which results in minimum impacts on learning.** With multiple indicators, teachers may direct their attention towards what is easier to change, thus there should be a clear focus or weight on learning to help teachers direct their efforts.
- **Third, large-scale incentive schemes require regular evaluation and this provides an opportunity to consolidate a culture of learning monitoring and improve teachers' performance in the long-term.** As seen in Brazil and Chile, large-scale exams that are conducted on a regular basis, allow for more comparability among cohorts, and are more likely to be perceived as unbiased. Developing and implementing these assessments is not simple but it greatly contributes to ensure that teachers and communities trust in the system. Additionally, teachers should understand the test and the abilities assessed because this will help them tailor their practice. It is not teaching to the test, but rather being able to measure if students acquired the content being taught. In this process, external support may be required, for example from regional supervisors or implementer agents. The state of São Paulo instructed regional supervisors to organize teacher training about the standardized evaluations and how they related to the curriculum. The state of Ceará, also in Brazil and described in the following section, organized a technical assistance program that aligned teachers' training and pedagogical material with learning assessments.

2.2.3. Incentives focused on selecting teachers

Evidence on the effectiveness of performance-pay contracts to recruit better teachers is mixed but teachers do respond positively to performance-pay salary arrangements. Two experiments tried to identify if offering teachers a pay-for-performance contract would be an effective way to select better professionals who, in turn, would have a higher impact on leveraging student learning. In private schools in Pakistan, teachers from primary and secondary levels were asked to choose between a raise in their salaries following a fixed-rate or a pay-for-performance rate. The study found that teachers who prefer performance schemes generally are more responsive to incentives and have higher abilities, as measured by learning outcomes of their students two years before the intervention. Using

two additional treatments, the study identified that teachers' decisions on performance-pay also depends on switching costs and how well-informed they are about their quality (Brown & Andrabi, 2021). Another experiment run in Rwanda offered primary school teachers in the recruitment stage the possibility of choosing between a performance¹⁵ or fixed-wage contract. It found that teachers who chose performance pay were more responsive to incentives but that they do not have significant differences in ex-ante quality, measured by their Teacher Training College final exam score, or observed skills, measured by a grading task. By the end of the second year, students whose teachers were under performance contracts scored 0.16 SD higher. Learning gains seem to be driven by improved teacher attendance and more effective pedagogy (Leaver et al., 2021).

Box 1 - What do teachers think about incentives?

Previous research has shown that teachers support test-based accountability and perceive they are accountable for student learning. A survey of teachers from 9 countries (Afghanistan, Argentina, Indonesia, Myanmar, Pakistan, Senegal, Tajikistan and Zanzibar) found that more than 59 percent of teachers support the idea that student test scores should be the main factor to assess teacher performance. Apart from Indonesia and Argentina, in all other countries at least 67 percent of teachers support performance bonuses. The authors found variations in preferences among countries and attribute that to local contexts but the high levels of acceptance are well above what is seen in developed countries (Sabarwal & Abu-Jawdeh, 2018).

More recent research also documents teachers support for pay-for-performance schemes. A frequent concern about pay-for-performance, especially among policy-makers, relates to teachers' perceptions and acceptance of the incentive mechanism. To shed light on this topic, some studies have also gathered data on teachers' perceptions. Evidence from different settings indicates that, overall, teachers support pay-for-performance schemes. In Indonesia, 97% of teachers agree with performance-based promotion, 80% are favorable to use student test scores as a criteria to promote teachers. A survey administered to 500 teachers and 100 head-teachers in Indonesia¹⁶ has shown great acceptability towards performance-pay. 72% of teachers are favorable of assessing performance through teachers' performance evaluation or 62% of using students' learning outcomes. Additionally, 65% of headteachers agree with using teachers' evaluation results as part of their performance assessment (Perez-Alvarez et al., 2020). In Tanzania, 96% of teachers support performance pay and at least 61% agree with attaching some part of salary increase to performance. Among headteachers, 80% support performance pay.¹⁷ Even among parents, acceptance is high (55%) (Mbiti & Schipper, 2021).

¹⁵ Performance was measured considering teachers' attendance, quality of lesson plans, quality of observed classes and student test scores (all components having equal weight). The measure of student learning adopted the pay-for-percentile scheme proposed by Barlevy and Neal (2012), rewarding the top 20% teachers within a district.

¹⁶ 100 schools were selected, following several criteria to account for heterogeneity in education level, geography, socioeconomic aspects and student learning outcomes.

¹⁷ Using a nationally representative sample of 350 public primary schools across 10 districts in Tanzania, it surveys education actors (teachers, families, students) to understand what they think about performance pay.

High levels of perceived fairness and transparency within the system are likely to foster teacher trust and acceptance of performance pay. System's reliability seems to greatly influence teachers' motivation and acceptance of performance-linked pay, as noticed in India, especially among teachers with lower pay base (Muralidharan & Sundararaman, 2011).

Teachers' support is higher among high-performing teachers. Teachers' ex-ante support for incentive mechanisms is correlated with ex-post performance, suggesting that teachers are aware of their effectiveness (Muralidharan & Sundararaman, 2011). A similar correlation was found in China, where, despite general support of the incentive, this perception was higher among teachers that could earn big incentive payments (Chang et al., 2021).

However, teachers positive perceptions about results-based programs do not seem to depend on characteristics such as age or gender nor on previous participation in an incentive program. Studies in China and Indonesia investigated if perceptions differed by teacher characteristics (as gender or age) and did not find significant differences (Chang et al., 2021; Perez-Alvarez et al., 2020). Likewise, teachers perceptions about results-based programs do not seem to be related to their previous participation in a performance-pay program (Mbiti & Schipper, 2021).

Pay-for-performance schemes do not seem to damage the school environment, and teachers feel appreciated by the community. Tanzanian teachers participating in the results-based program reported being satisfied with their work environment, indicating that pay-for-performance had not damaged it (Mbiti & Schipper, 2021). In Indonesia, an intervention focused on community monitoring has shown that teachers felt more appreciated by district education officials and their community, alleviating concerns of general dissatisfaction among teachers affected by the program (Gaduh et al., 2021).

2.3. Design and implementation features of teacher incentives

The discussion on the impacts of teacher incentives on distinct outcomes already elicited features that contribute to a program's impact. So far, teacher incentives have been largely implemented as a bonus policy instead of a change in teachers' contracts. On the design, most incentives consist of monetary bonuses, either individual or group-based, with large variation in terms of bonus size. On the metrics, it is not clear if paying-for-percentile is indeed better than levels or gains, and there seems to exist a trade-off between design complexity and ease of use, which involves both teachers' understanding and implementation issues.

The sample participated in a large experimental evaluation that included three interventions: a school grant program, a teacher performance-pay program and a combination of both programs.

Many interventions have tried in their design to prevent an increase in learning inequality, an issue that has been carefully assessed by Schipper & Pradhan (2022). They have found suggestive evidence that students with higher baseline test scores can benefit more from teacher incentives, probably because they have more resources available. They do not find strong evidence that teacher incentives with threshold designs result in more learning inequality.

On implementation, technical capacity, resources and political are decisive pre-conditions for successful incentive mechanisms (Breeding et al, 2021). Moreover, having a strong accountability mechanism - that goes beyond student learning assessments - contributes to strengthening program implementation, which fosters results. Annex 2 points out the studies that addressed specific design and implementation features, adding to the topics already discussed in other analyses.¹⁸

2.4. Conclusion on incentives for teachers

Most of the evidence on incentives for teachers had positive outcomes. Incentives designed to improve teacher attendance can raise attendance when strong monitoring systems are applied. Incentive rules focused on student learning can improve learning (12 out of 15 interventions designed to improve student learning had raised test scores) but key factors that lead to positive results are: i) assuring schools have minimum learning resources, ii) when the reward rule clearly puts learning as the main outcome indicator, and iii) when teachers set specific targets for their students and articulate them to learning strategies.

Design and implementation matter, and, in both stages, it is necessary to have the technical requirements, resources, and political will (Breeding et al., 2021). So far, teacher incentives have been largely implemented as a bonus policy instead of a change in teachers' contracts. Pay for performance contracts can sort teachers that are more responsive to incentives but evidence on selecting more qualified teachers is mixed.

On the design, it is not clear if pay-for-percentile is the best metric, as positive evidence came from Math teachers in China. In fact, most of the positive evidence, both from experiments and scaled-up interventions, use simpler and group-based designs. Programs implemented by governments and on a large scale include multiple outcome indicators but those with a positive impact on learning are the ones that had a substantial part of the incentive attached to student test scores. Additionally, given the heterogeneity among schools, it is not trivial to design a large-scale mechanism in which all schools perceive they can have a fair competition. To be sustainable and fully effective, incentives require an accountability routine that ultimately can shape teachers' behavior and have positive long-term effects.

On gaps, it is still not clear if individual incentives are better than group incentives since there is an unbalance in the evidence: experiments have focused on individual incentives whereas large-scale interventions adopted group-based designs. Additionally, little is known about the long-term effects of incentive policies.

¹⁸ Lee and Medina (2019) have detailed several design issues and Breeding et al (2021) propose a step-by-step guide to identify if performance-pay is an appropriate approach for a given setting, considering preconditions, implementation features and risk analysis.

3. Students and families

3.1. Why incentives for students and families?

Incentives for students and families try to raise the demand for education by rewarding expected behaviors as enrolling children in school, attending school regularly, or improving learning outcomes. The majority of these incentives try to address three main issues: (i) prioritization of education; (ii) reducing financial barriers to access education; and (iii) compensate families for the economic loss associated with sending a child to school. Education incentives for students and families are generally part of a broader goal of poverty alleviation and human capital investment, which means they have first-order results and spillover effects beyond education.

3.2. What results can student and family incentives achieve?

Student and family incentives can effectively increase demand for education and have potential to improve learning. Conditional cash transfers (CCTs) are the most common intervention to foster schooling and have a good track record of increasing enrollment, attendance and completion rates, especially in primary and lower secondary education. In upper secondary school, short-term impacts are mixed, while students who were enrolled in primary school as a result of cash transfers also have increased upper secondary enrollment and completion in the long-run. Evidence on learning effects is also mixed, with small positive impacts from medium or long-term analyses. Merit-based incentives tied to students' test scores are the most common results-based intervention to increase student learning, with promising results. Nevertheless, the evidence base is still limited, which hampers generalizations on which design works better.

While interventions can potentially impact both schooling and learning, the incentive design usually focuses on one over the other. For instance, CCTs generally focus on improving schooling, while merit-based scholarships concentrate on learning. The success of CCTs in increasing enrollment led researchers and policy-makers to test various designs, including information campaigns and learning incentives. The following paragraphs outline the impacts of incentives according to their primary purpose – schooling or learning – and discuss whether incentives (i) achieve their primary purpose, (ii) lead to better learning or improved conditions for learning, and (iii) if there are other sorts of impacts (both positive or negative).

3.2.1. Incentives focused on schooling

Conditional cash transfers are one of the most widespread and effective interventions to raise demand for education and improve schooling, with consistently documented impacts in primary and lower secondary education. Evidence on cash transfers is extensive, accounting for more than

100 studies reporting effects on education (Evans et al., 2021).¹⁹ The traditional design consists of monetary and recurrent payments to parents if children are enrolled and attending school, and conditionality is usually set at an attendance rate between 80-85%. Evidence converges on the positive impacts of CCTs in decreasing school dropout rates and increasing school attendance and completion, and they are considered as the most effective intervention to improve families and students' demand for education (S. Baird et al., 2014; Bastagli et al., 2016; Damon et al., 2016b; Evans & Mendez Acosta, 2021; Evans & Yuan, 2019; Glewwe & Muralidharan, 2016; Lee & Medina, 2019; Molina Millán, Macours, et al., 2019; Snilstveit et al., 2016). Recent research provided more evidence on the impact of CCTs to improve schooling in lower secondary education (De Walque & Valente, 2018; Evans et al., 2021).

Programs targeting high school students have mixed results. While in Mexico and Colombia programs have raised schooling in upper secondary education, evidence for Brazil is debatable and for Mexico City is negative. A possible hampering factor is that, as students get older, education's direct and opportunity costs increase, and many adolescents decide to drop out of school. Since increasing the size of transfers can compromise fiscal space, some alternatives are combining transfers with information campaigns about schooling returns or adopting re-enrollment as a conditionality. For instance, the urban version of Oportunidades in Mexico (that included high school students as beneficiaries) had a slight, positive, and significant effect on education attainment (Whetten et al., 2019). In Colombia, a program targeting high school students increased attendance, especially for subgroups that conditioned the transfer to re-enrollment in the next grade or to high school completion and enrollment in a tertiary program, and among the poorest and low-performing students (Barrera-Osorio et al., 2019). In Brazil, the expansion of Bolsa Família to high school is associated with positive impacts on attendance and completion rates, especially for at-risk students. However, a recent study employing an alternative identification strategy finds null effects, casting doubts on previous results (Draeger, 2021). Also, in Mexico City, a program targeting only high school students did not significantly impact graduation rates or test scores. (Dustan, 2020).

However, children who were recipients of CCTs during primary school have increased upper secondary enrollment and completion. In Colombia, Mexico, Nicaragua, Indonesia, and Pakistan, researchers found positive effects on enrollment and completion of secondary school for children from families who had participated in CCT schemes earlier on (Attanasio et al., 2021; Baez & Camacho, 2011; Barham et al., 2017; Cahyadi et al., 2020; Chhabra et al., 2019; Duque et al., 2019; Filmer & Schady, 2014; Whetten et al., 2019). A possible reason for these effects is that cash transfers help transmitting the message of the importance of education. For example, in Morocco, parents' beliefs about the returns to education have risen after program implementation (Benhassine et al., 2015). In Colombia, parents participating in *Familias en Acción* were 11 percentage points more likely to aspire for higher education for their children, and kids were 20 percentage points more likely to desire higher education. Effects were even higher among low-income households (García et al., 2019). Another intervention in Colombia that tested conditioning transfers to re-enrollment in the next grade of the high school noticed an enrollment increase in tertiary education, suggesting that the program

¹⁹ These refer to 27 unique cash transfer programmes, covering 20 countries in Latin America and the Caribbean, sub-Saharan Africa, East Asia and the Pacific, South Asia, and Middle East and North Africa, ordered by predominance. It involves different levels of conditionalities and different designs.

impacted families' demand for education (Barrera-Osorio et al., 2019). For a summary on long-term effects, see table 2.

Table 2 – Long-term impacts of cash transfers

Schooling	One of the major long-term effects of cash transfers, as already mentioned, refers to increased schooling (Attanasio et al., 2021; S. Baird et al., 2019; Cahyadi et al., 2020; Chhabra et al., 2019; Duque et al., 2019; Molina Millán, Barham, et al., 2019; Whetten et al., 2019).
Learning	Evidence on learning is mixed , with clear learning gains in Colombia (<i>Familias en Acción</i>) and Nicaragua but not significant effects in Mexico or Malawi (Barham et al., 2017; Molina Millán, Barham, et al., 2019). A recent study on Cambodia identified higher cognitive skills among children that received the transfer based on their baseline test scores (Barrera-Osorio et al., 2018).
Other education-related outcomes	Socioemotional skills: benefitting from cash transfers during early childhood led to a positive impact in Mexico and Colombia but null results in Ecuador (Molina Millán, Barham, et al., 2019). Not significant effects were also found in Cambodia for children in primary education (Barrera-Osorio et al., 2018; Molina Millán, Barham, et al., 2019). School readiness: In El Salvador, receiving CCT raised enrollment in pre-school and completion of at least one year by 12 percentage points (Sanchez Chico et al., 2020).
Employment and earnings	Evidence is mixed . In Cambodia, benefits may be concentrated on poorer beneficiaries who are males (Barrera-Osorio et al., 2018). In Nicaragua, former beneficiaries experienced higher labor income and off-farm employment. Effects in Mexico and Colombia also point to positive labor market outcomes. In contrast, other studies found no clear gains in Mexico, Honduras, Ecuador, Cambodia, Pakistan, Malawi, and Indonesia (S. Baird et al., 2019; Cahyadi et al., 2020; Molina Millán, Barham, et al., 2019).
Welfare	While there are many studies reporting effects on welfare, this summary focuses on the latest evidence of CCTs with education conditionalities. There is substantial evidence reporting positive effects on health outcomes . <i>Familias en Acción</i> in Colombia led to a reduction in teenage pregnancy of 2.3pp for women (Attanasio et al., 2021). In Malawi, CCTs caused sustained impacts on incidence of marriage and pregnancy, age at first birth, the total number of births, and desired fertility – but only among the stratum of adolescent females who had already dropped out of school at baseline and were all assigned to CCTs (S. Baird et al., 2019). In Indonesia, stunting fell by 23 percent (Cahyadi et al., 2020). <i>Prospera</i> in Mexico had positive effects, especially on food security and mobility (Aguilar et al., 2019). In Honduras, exposure to the CCT more than doubles the probability of international migration of young men, from 3 to 7 percentage points (Molina Millán et al., 2020). In Colombia, cash transfers reduced arrest rates by 2.7pp for men (Attanasio et al., 2021).

Some CCTs also seem to have positive impacts on learning. Evidence has increased and is prevalent among studies investigating medium-term program effects. In comparison to schooling impacts, there are significantly fewer studies reporting CCTs' impact on learning (17 studies of 12 different programs). Eight out of the 17 studies identified positive, although modest, learning effects. Most of them only appear in medium or long-term analyses published more recently. For instance, between 1997-2014, only two out of 9 studies showed positive effects of cash transfers on learning. All the others did not yield statistically significant results (see table 3). From 2015 onwards, 6 out of 8 studies showed positive effects of cash transfers on learning.

Several factors might explain why learning effects appear in the medium and long run. One could be that there was less data on test scores for the first generation of CCTs. Another factor could be that researchers decided not to report negative or statistically insignificant results. Additionally, CCTs were originally designed to increase the demand for schooling rather than to increase learning. In this sense, learning outcomes are considered second-order effects, which were only further explored after the evidence on the schooling impacts was more established. Another factor that might explain the relatively greater prevalence of positive learning outcomes nowadays is the time elapsed since the beginning of the program. Until 2014, a program's follow-up occurred, on average, three years after the start of the intervention (with a median of 2 years), whereas from 2015 onwards, studies measured impacts after 5.25 years of programs' implementation (with a median of 5 years). It is known from several educational interventions that impacts on learning take longer to appear.

Interventions that did not show meaningful results on student learning suggest that crowding classrooms, changes in student composition, and difficulties in measuring learning hamper positive effects. The literature points to at least three factors related to CCTs' lack of impact on learning, although the prevalence of those factors varies with the context. First, CCTs tend to increase enrollment among children that were not accessing school and, for this reason, have low performance, at least in the short run, decreasing students' average performance in test scores. This case was noticed, for example, in Cambodia (Filmer & Schady, 2009). Second, learning outcomes are often measured through test scores applied at the end of primary or secondary levels, which implies that at-risk children that enrolled in schools due to CCTs but then dropped out for work or other reasons might have benefitted from the program (even more than students that were previously enrolled and that completed education), but the impact evaluation is unable to capture these effects since kids did not take the final test (Akresh & de Walque, 2013; Baez & Camacho, 2011). Third, CCTs have a more significant effect in the most deprived areas, which frequently have low-quality teaching and less infrastructure. As the increase in enrollment tends to increase class size, this could undermine instruction and learning. This situation seems to be the case, for example, in Tayssir's program in Morocco, where the program increased class size by 12% and each additional student in a class led to a reduction in boys' test scores of 0.03 to 0.05 SD in beneficiary municipalities (Gazeaud & Ricard, 2021).

Table 3 - Studies that reported CCTs' impacts on student learning

Year	Authors	Country	Name & Context	Intervention category	Learning outcome	Measure	Follow-up
2009	Filmer et Schady	Cambodia	CESSP Scholarship Program	Scholarship conditional on attendance and grade progress	Not-significant	-	3 years
2011	Baez & Camacho	Colombia	Familias en Acción	CCT: long-term effects	Not-significant overall and in math. Small negative effect on language.	- 0.05 SD	9 years
2011	Baird et al	Malawi	Zomba Cash Transfer Programme	CCT: schooling and learning effects	Positive	+0.13 SD	2 years
2013	Akresh, de Walqye & Kazianga	Burkina Faso	Nahouri Cash Transfers Pilot Project	UCT vs CCT: schooling and learning effects	Not-significant overall and in math. Small significant increase in language (all children average)	+0.2 in z-score	2 years
2013	Mo et al	China	CCT in Northwest China	CCT: schooling and learning effects	Not-significant	-	0.9 year
2013	Baird et al	Malawi	Zomba Cash Transfer Programme	CCT: schooling and learning effects	Positive	+0.15 SD	2 years
2013	Benhassine et al	Morocco	Tayssir	Labelled cash transfer	Not-significant	-	2 years
2014	Filmer et Schady	Cambodia	CESSP Scholarship Program	Scholarship conditional on attendance and grade progress	Not-significant	-	5 years
2014	Evans et al	Tanzania	Tanzanian Social Action Fund Program	CCT: schooling and learning effects	Not-significant	-	2 years
2016	Barrera-Osorio & Filmer	Cambodia	Primary School Scholarship Pilot Program	Scholarship conditional on attendance and grade progress	Positive	+0.16 SD	3 years
2017	Barham et al	Nicaragua	Red de Protección Social	CCT: long-term effects	Positive	+0.2-0.28 SD	10 years
2018	Barrera-Osorio et al	Cambodia	Primary School Scholarship Pilot Program	Scholarship conditional on attendance and grade progress	Positive	+0.11 SD	9 years
2018	De Walque & Valente	Mozambique	CCT in Manica province	CCT: schooling and learning effects	Positive (assessed only math)	+8.5-9.4%	1 year
2019	Duque et al	Colombia	Familias en Acción	CCT: long-term effects	Positive	+0.13 SD	5 years
2019	Behrman et al	Mexico	Prospera	CCT: learning effects	Positive	+0.05 SD	6 years
2020	Dustan	Mexico	Prepara Sí, Ciudad de Mexico	CCT: schooling and learning effects	Not-significant	-	3 years
2021	Gazeaud & Ricard	Morocco	Tayssir	CCT: learning effects	Negative	-0.1-0.18 SD	5 years

* Z-score is defined as the difference between the child's raw test score and the mean test score of the same-aged children, divided by the standard deviation of those same-aged children.

Focusing on marginalized groups and strengthening information components could be possible drivers of interventions that successfully impacted learning. Out of six programs with positive results on student learning, two common factors appear: four focused on marginalized groups, and three had an information component that seemed to have played a significant role.

- **Malawi, Nicaragua, Mexico, and Mozambique identified positive outcomes among marginalized groups of students.** In Malawi, the CCT targeted adolescent girls - whose access to education is still limited in many low-income countries (Evans & Yuan, 2019) - and found positive results on language outcomes in the order of 0.14 SD. The study distinguishes impacts for girls initially enrolled in school from girls who were out of school and returned to studying due to the cash transfer. While both groups had positive and similar results, girls initially out of school benefited a bit more (S. Baird et al., 2011; S. J. Baird et al., 2013). In Mozambique, an intervention targeting girls also had positive results on learning, as described in the next paragraph (De Walque & Valente, 2018). In Nicaragua, a study assessed the impacts of CCT for boys, as, in the region, boys have a higher risk of dropout to begin income-earning activities. Long-term effects for children exposed to CCT for at least three years indicate a substantial and significant impact of 0.2 – 0.28 SD on abilities learned in school, as measured by tests applied in students' homes (Barham et al., 2017). A recent analysis of the Mexican *Prospera* program focused on marginalized students of the poorest areas and indigenous communities. Using data from a standardized national exam and considering individual fixed-effects²⁰, it found positive although small impacts of the program (0.05 SD on language and math) for the most impoverished communities and 0.1 SD on language for indigenous students (Behrman et al., 2019).²¹
- **In Cambodia, Mozambique, and Colombia, information components seemed to play a relevant role in programs with positive effects.** In Colombia, a study assessed the effects of *Familias en Acción's* initial phase (2001-2004) ten years after exposure to the program. It found that children beneficiaries of the CCT who remained in school scored 0.13 SD higher at the secondary graduation test than non-recipient students (Duque et al., 2019). Interestingly, the program included two information features in its initial design. First, families needed to get proof of children's attendance in the school, which allowed them to meet with teachers every two months and receive regular feedback on children's performance. Additionally, families had to attend information meetings and receive booklets about healthcare and education to raise their awareness of human capital investment returns. While the researchers were unable to attribute long-term causality, an assessment of the first year of program implementation showed an increase in parents' and children's aspirations towards

²⁰ The outcome measure for effects on achievement is the difference in test scores, which allow for individual fixed effects as a way to control for selectivity in test-taking, meaning selection bias of those who continued enrolled in school and took the test. Note that this approach is subject to downward bias.

²¹ Other studies have investigated CCTs' impacts, besides learning, on marginalized groups. For example, Evans et al. (2021) found that students with better performance at the baseline were 21pp more likely to complete primary school at the endline and attend secondary school than low-performing students. They also found that the CCT increased girls' schooling, which is in line with previous evidence (Baird et al., 2014; Bastagli et al., 2016; Evans et al., 2021). Lastly, they found that the poorest children benefitted the most, which was seen in Tanzania, Brazil, Colombia, Kenya, Cambodia, and China (Barrera-Osorio et al., 2019; Cardoso & de Souza, 2009; Filmer & Schady, 2008; Mo et al., 2013).

higher education, which may have impacted their efforts during the basic education cycle (García et al., 2019).

- In Cambodia, a program functioning similarly to a CCT tested the effects of providing cash to individuals, labeling it as a scholarship, based on poverty status (treatment 1) or baseline test scores (treatment 2). The amount transferred had free use and was equivalent to 3.3 percent of annual per capita expenditure, an amount similar to other CCT programs. At the same time, unlike CCTs, the program targeted students already enrolled and asked them to keep their passing grades and maintain enrollment and attendance. The “scholarship” improved test scores by 0.16 SD on the group of students that received the money based on baseline academic performance. Families and students in this group also exerted more effort in terms of hours spent studying and invested more in education. In conjunction, this suggests that framing the program as a merit-based transfer may have impacted students' and family's motivation towards schooling, leading to better outcomes (Barrera-Osorio & Filmer, 2016). Nine years after program inception, when individuals were, on average, 21 years old, results show that both scholarship types led to higher long-term educational attainment (about 0.21-0.29 more grades attained). Still, only merit-based scholarships led to improvements in cognitive skills in the order of 0.11 SD, besides greater self-reported well-being (0.18 SD) and employment probability (3.4pp) (Barrera-Osorio et al., 2018).
- An experiment in Mozambique partly funded by REACH compared three approaches to leverage education outcomes among girls and disentangle the effects of information and incentives. In group 1, parents received regular information about the attendance of their daughters in school. In group 2, weekly reports were combined with cash transfers to parents, conditional on school attendance. In group 3, parents received information on girls' attendance, and the daughters with regular attendance received vouchers to exchange for goods valued by pupils, such as school uniforms, shoes, bags, pens, and notebooks. The program successfully raised attendance, with higher impacts in group 3, and had positive effects on learning in groups 1 and 3, in which students scored 8.5% and 9.4% higher than the control group, respectively. These results suggest that students' agency may be an essential factor in raising learning. It is worth noting that the program mainly targeted girls already enrolled in school, which presumably do not change classroom composition (De Walque & Valente, 2018).

3.2.2. Incentives focused on learning outcomes

Merit-based scholarships and pay-for-grades incentives are the most common results-based interventions to foster student learning, with most evidence pointing to positive effects. Merit-based incentives explicitly link the reward to academic performance. These rewards can take the form of scholarships, lump-sum payments, or prizes. The rationale for merit-based incentives for students includes compensating the opportunity costs if the child was not in school and increasing parents' monitoring of student learning. Five out of seven interventions incentivizing students or parents yielded positive learning outcomes, despite focusing on different settings and education levels, including grades 1 to 10 (Berry, 2015; Blimpo, 2014; Hirshleifer, 2017; Kremer et al., 2009; Li et al., 2014). Two interventions found that combining incentives for teachers and students can be more effective than only targeting students (Behrman et al., 2015; Filmer et al., 2020).

Evidence on merit-based incentives is not extensive but encompasses various designs, making it difficult to generalize findings on which design works better. The following paragraphs present some promising design features, but more research is needed on the topic.

- **Grouping students with different learning levels seems more effective than individual rewards.** An intervention in China and another in Benin tested the effects of rewarding children based on their individual or group performance, and both found better results on the latter. In China, the experiment for grades 3 to 6 found that pairing low-performing with high-performing children (and providing cash rewards based on improvements from the low-performing student) led to an increase in test scores by 0.26 SD of low-achieving students without hampering the high-achievers. The intervention also tested rewarding students for their individual results and found no effects (Li et al., 2014). In Benin, an experiment compared three types of intervention: (1) rewarding students based on their individual performance, (2) gathering children in groups of four and paying for the average performance of a group if they achieved specific standards, and (3) a tournament among groups of four students. All types of incentives led to positive and statistically significant impacts. Still, estimates for the group tournament were higher (treatment 1 increased test scores by 0.29 SD, treatment 2 by 0.27 SD, and treatment 3 by 0.34 SD) (Blimpo, 2014).
- Additional evidence on individual incentives is mixed, as seen by an experiment in Kenya, another in Malawi, and a scaled-up program in Chile. In Kenya, girls in 6th grade who scored among the top 15% received a scholarship for tuition and school supplies for the next two years. The intervention increased schooling and learning among girls. Surprisingly, it also raised teachers' attendance, which led to spillover learning effects for boys (Friedman et al., 2016; Kremer et al., 2009). Nevertheless, a very similar intervention in Malawi was unsuccessful. It compared two treatment groups (one equal to the Kenyan program that rewarded the top 15% of 5th to 8th graders, the other in which students were grouped according to their initial score and awarded the top 15% of each group). The first group decreased test scores by 0.27 SD, which is likely driven by an observed reduction in students' motivation to study, especially among the least likely to win. Fortunately, these negative effects did not persist after the incentive was removed. The second group, in turn, had no impact (Berry et al., 2019). In Chile, a scaled-up policy provided a monetary reward to grade

5th to 12th students that scored among the top 30% in their grade and belonged to the bottom 30% of the income distribution. After one year of the program's implementation, no significant results were found. One possible explanation for this is the low understanding of the incentive scheme during its first year of implementation (Crespo, 2019).

- **Cash and in-kind rewards may impact students differently.** In India, children in grades 1-3 had to achieve a learning target (established according to their pre-test scores) to receive an award. Incentives varied in type (cash, a voucher to buy toys or toys) and the recipient (parent or child). While the recipient's identity did not influence results, the type of reward did. Noncash incentives were more effective for initially low-performing children and cash incentives for high-performing ones (Berry, 2015). It would be interesting to see if this study could be replicated in other settings.
- **Quick assessments regularly administered can encourage student learning more than a final long test.** An experiment in India with students in grades 4-6 had two treatment groups. In one of them, children needed to conduct regular quizzes to assess their knowledge. In another group, learning was only evaluated in the final test. For each correct answer, children received a point that was converted into cash. It found that regular quizzes increased learning by 0.57 SD while a final test led to non-significant results (Hirshleifer, 2017).
- **Combining incentives for teachers and students can be more effective than only student incentives.** Experiments in Mexico and Tanzania investigated the impacts of providing monetary incentives only for high school students, teachers or both. In Mexico, incentives only improved test scores among the group in which students and teachers were incentivized (according to a low-stakes test) (Behrman et al., 2015). In Tanzania, incentives for students had no effects. The combined scheme had positive results in the first year, but in the second year, only the teacher-incentive group had a significant increase in learning, although the study could not identify possible explanatory factors (Filmer et al., 2020).

3.3. Design and implementation features of incentives for students and families

Annex 2 presents a summary of the main issues considered in the design of CCTs, but it is worth highlighting the role of informing families about the value of education and their children's performance. Another useful feature is the alignment of transfer payments with important dates or times in the academic year, e.g. for instance, giving transfers right before school fees' deadline or delaying part of the payment and making it conditional on next grade enrollment (Barrera-Osorio et al., 2019). On monitoring conditionality, when neighbors or friends receive a warning about their non-compliance, positive effects are also noticed among their peers (Brollo et al., 2020).

3.4. Conclusion on incentives for students

Incentives to students and families can foster schooling and learning, but interventions tend to target one over the other.

Conditional cash transfers have shown to be the most effective intervention to increase demand for schooling and have a good track record of increasing enrollment, attendance and completion rates, especially in primary and lower secondary education. In high school, short-term average impacts on schooling are mixed, though there are positive effects for the most vulnerable students. In the long run, children exposed to CCTs at the primary level had increased upper secondary enrollment and completion, probably due to the increased exposure to the benefits of education returns. Evidence on learning effects is mixed, with 8 out of the 17 studies pointing to small positive impacts, most of them in medium or long-term analyses published more recently. Possible drivers of learning impacts are focusing interventions on marginalized groups and strengthening information components.

Merit-based incentives tied to students' test scores are the most common results-based intervention to foster student learning, with promising results. Since evidence on these programs is not extensive and encompasses various designs, it is difficult to generalize findings on which design works better, highlighting the need for more research. One aspect that does come out in current research is that grouping students with different learning levels seems more effective than individual rewards.

4. Schools

4.1. Why results-based incentives for schools?

The majority of RBF interventions for schools tries to strengthen schools' autonomy and improve outcomes simultaneously. The most common form of RBF for schools is performance-based school grants intended to improve outcomes and provide schools discretion over where to invest the money.²² The rationale for providing grants to schools, both conditionally and unconditionally, is the belief that local decision-makers, such as principals and the school community, better understand the school's needs and can invest money more effectively. Performance-based school grants may also intend to foster schools' social status by recognizing the high-performing ones, which establishes a sense of pride among educators and the community that ultimately reinforces their efforts to improve education outcomes.

Incentives can also be established for non-government providers or in the form of impact bonds, but these are addressed in other reports. In regions such as South Asia and Sub-Saharan Africa, non-governmental organizations are key education providers, and therefore some RBF interventions have been designed to incentivize those actors. For instance, between 2013 and 2017, the British government funded 15 projects under the Girls' Education Challenge that included performance-based disbursement indicators. NGOs ran the projects, and incentives focused on improving learning outcomes or student attendance (Clist, 2019). Additionally, incentives to "schools" can take place in the form of impact bonds, where private investors provide upfront capital to a service provider (such as an NGO that runs an education center). When the provider achieves pre-agreed results, an outcome funder repays private investors. In developing countries, outcome funders are generally donors and foundations (Gustafsson-Wright & Boggild-Jones, 2019). Although similar to school grants in which the government incentivizes its own network of schools, these types of interventions involve different stakeholders and have been analyzed by other reports (Ecorys, 2022; Terway et al.;(2021).

²² This distinguishes RBF for schools from RBF for principals. The latter are frequently combined with incentives for teachers and it is not possible to isolate the effect for headteachers only. Thus, these interventions are discussed in the Teachers section (Behrman et al., 2015; Bellés-Obrero & Lombardi, 2021; Contreras & Rau, 2012; Glewwe et al., 2010; Lépine, 2021).

4.2. What results can school incentives achieve?

RBF for schools has been applied to incentivize improvements in learning, transition rates, school management, teacher attendance, textbooks usage, and others. Still, there are only two robust impact evaluations, both from incentives to learning. Most of the literature on school grants focus on unconditional transfers (Evans & Mendez Acosta, 2021; Ganimian & Murnane, 2014; Snilstveit et al., 2016), and, at the time of this writing, there were only two publicly available impact evaluations that studied performance-based grants to schools.²³ However, this lack of robust evidence does not mean that performance grants are rare. In fact, there are several incentives to schools introduced in development projects (see Box 2). Some of these initiatives are very recent; others were established for all schools, limiting the capacity to assess their impact. But, in general, what differs those grants from the two interventions that have been rigorously assessed is that they adopt intermediary outcomes. Some reasons for that are (i) weak technical capacity, (ii) limited monitoring systems to measure performance indicators correctly, and (iii) the desire to gradually strengthen an accountability culture, starting from outputs to transition to outcomes. The existence of several performance grants to schools highlights the relevance of the topic and the urgency for more research on its effectiveness.

Among the two impact evaluations, one found mixed results, and the other (that combined incentives with capacity-building) identified positive effects. As evidence is limited, aspects discussed in this section should be carefully interpreted as indicative of promising design features of performance incentives for schools, but more research is needed. The study with positive impact combined the incentive with technical support to schools, a feature pointed as a promising channel by research in correlated topics, such as unconditional grants and school-based management programs (Lee & Medina, 2019; Snilstveit et al., 2016).

In Jakarta, a performance grant had positive learning effects in lower secondary schools and a little negative impact in primary schools. The government of Jakarta introduced in 2015 a performance-based component into a pre-existent school grant to raise education efficiency. The grant awarded the top 25% of schools with an additional 20% of funding per student.²⁴ Since the program was introduced to almost all schools, the analysis adopted quasi-experimental techniques. It found an average small negative effect for primary schools and an average positive effect for lower secondary schools. For lower secondary schools, the incentive increased students' performance by 2.6 percentage points in the program's first year and 4.6 and 4.3 percentage points in the second and third years, respectively. However, schools that received the grant were those that already had higher performance and had to make fewer improvements to win the grant. In primary education, the program had a small negative impact on high-performing schools and a small positive effect on low-

²³ Other studies explore the effects of school grants whose distribution depends on the selection or approval of a resources' usage plan but without a results-based component (Carneiro et al., 2020; Garcia-Moreno et al., 2019; Romero et al., 2021). RBF to schools have also been applied in other settings (as presented in box 2) but without an analysis that enables us to establish causality between the incentive mechanisms and the education outcomes.

²⁴ The rank formula combined students' average performance over the previous two years and the percentage point increase in this same period, and it considered the education level (primary and lower secondary) and schools' district to foster equality of opportunities.

performing ones, which reduced inequality. The evaluation also identified no additional effect for schools after they received the grant (Al-Samarrai et al., 2018).

In Ceará, Brazil, a school tournament was combined with principal peer-to-peer mentoring increased learning in low-performing schools without negative effects for high-performing ones. The state of Ceará implemented from 2007 onwards a set of education reforms centered on literacy at the right age.²⁵ Among those, it established in 2009 a performance-based incentive mechanism for schools that also combined peer mentoring for principals (*Prêmio Escola Nota 10*). The program monetarily rewards the top 150 schools according to their performance on the Education Quality Index (EQI), which mainly combines students' learning levels and learning gains in grade 2. To receive the totality of the grant, top-performing schools must support a low-performing school throughout one year and maintain or improve its own result. Low-performing schools also receive a grant (50% upfront and 50% conditional on enhancing their learning outcomes to a minimum threshold). An impact evaluation conducted with REACH funding identified a positive impact of 0.18 SD in the education quality index of low-performing participant schools without harming top-performing schools, which contributes to reducing the performance gap among schools. Nevertheless, the evaluation did not dissociate the effect of the performance-based incentive and the peer-monitoring component (Goldemberg et al., 2021). While further research is needed to disentangle these effects, evidence from another intervention that combined results-based incentives with pedagogical support identified that both components' impact is almost double the results of the incentive alone (Lautharte et al., 2021).²⁶

Still, more research is needed on the effects of RBF on schools. Overall, it is vital to expand the evidence around performance-based school grants. More research on the long-term effects is also needed. Evidence from the health sector suggests that grants can improve organizational constraints, which has a sustained impact in the long run (Bernal et al., 2018). Likewise, the intervention in Ceará is promising from a political economy perspective, as it enhances schools' practices without dismissing principals.

²⁵ It comprised the decentralization of primary and lower secondary education to municipal governments, a technical assistance program that provided pedagogical support to municipalities, incentive mechanisms and strong monitoring systems. Details of this reform are presented in the governments' section.

²⁶ Even for interventions with non-RBF grants, providing capacity building to schools has strengthened grants' impacts in most cases (Blimpo et al., 2015; Garcia-Moreno et al., 2019; Gertler et al., 2012; Romero et al., 2021).

Box 2 - There is more to RBF to schools than what is documented in the literature

RBF to schools has been applied in many settings, and some of them are very recent, limiting the analysis of their impacts. Below are some examples that highlight the diversity in their design:

Incentives in the form of tournaments:

- Governments in Tanzania and Malaysia provided monetary incentives to top-performing primary and secondary education schools (Mhagama, 2020).
- In Ethiopia, the Ministry of Education (MoE) awarded the top 10 percent primary schools with better transition rates between grades 1 and 2 and in grade 5 completion rates ([Ethiopia General Education Quality Improvement Program for Equity – 2018-2025](#)).
- In Sri Lanka, the MoE established two conditional grants to primary and secondary schools. One rewards the top 500 schools with higher learning outcomes, and the other is conditional on achieving teachers' professional development goals ([General Education Modernization Project, 2018-2024](#)).
- In Senegal, a multi-donor development project ([Quality Improvement and Equity of Basic Education, 2014-2021](#)) had three RBF layers. One provided top-up rewards to high-performing schools and technical assistance to the bottom 20%.

Grants upon achieving results:

- A performance-based grant will be scaled-up in Mozambique, based on lessons a pilot in three provinces that took place with REACH support. The pilot rewarded teacher attendance, transparency in school-grant management, involvement in the school council, and students' reading skills ([Improving Learning and Empowering Girls, 2021-2025](#)).
- In Cameroon, with REACH funding, the MoE piloted a school grant based on improvements in student retention, teacher attendance, financial transparency, community satisfaction, and textbook usage. Thirty percent of the grant was destined to bonus for teachers and headteachers, and the remaining part was invested in schools, according to their action plans. Positive outcomes led the government to scale up the program to 3,000 schools (World Bank, 2019a, 2019b).
- The Democratic Republic of Congo considers payments to schools linked to the quantity and quality performance indicators (Global Partnership for Education, 2020).
- In Madagascar, the MoE conditioned grants to the submission of improvement plans, the change in the academic calendar, and the improvement of students' graduation rates ([Madagascar Basic Education Support Project, 2018-2023](#)).
- In Bangladesh, a World Bank operation considered results-based incentives for students and schools in secondary education. School grants were conditional on accountability requirements and performance targets, such as student attendance, retention, and learning outcomes ([Transforming Secondary Education for Results Operation, 2018-2022](#)).
- In Malawi, the MoE transfers grants to schools, conditional on improvements in retention rates, especially among girls (Global Partnership for Education, 2020).
- In Nepal, the District Education Office has a performance-based grant to community schools that incentivizes monitoring systems' and textbooks' usage, improvements in time spent on teaching, attendance and retention rates, and learning outcomes ([School Sector Development Program, 2017-2022](#)).

4.3. Design and implementation features of incentives for schools

A summary of the design and implementation features for school grants is presented in Annex 2, based on the two studies of results-based incentives to schools and the analysis of similar literature conducted by Lee and Medina (2019). Teacher incentives have already elicited some of these issues, such as those referring to pre-conditions, accountability mechanisms, and adverse behaviors, since school grants are similar to group-based teacher incentives. Overall, performance-based school incentives have mostly focused on student learning and giving schools monetary rewards. The reward formulas try to promote equity, while the program in Ceará also has a peer-monitoring component to support low-performing schools.

4.4. Conclusion on incentives for schools

Robust evidence on school incentives is currently restricted to two studies on performance-based school grants that use a tournament design. These grants are awarded based on learning outcomes.

Although limited, evidence on RBF for schools seems promising, especially when the incentive is combined with capacity-building interventions, such as peer-mentoring to principals of high- and low-performing schools. The intervention in Ceará is also promising from a political economy perspective, as it enhances schools' practices without dismissing principals.

The lack of robust evidence does not mean that performance grants are rare. In fact, there are several incentives to schools that have been implemented in development projects, most of them adopting intermediary outcomes. While most interventions' design does not allow causality analysis, it highlights the relevance of the topic and the need to expand the evidence base, including assessing its long-term effects.

5. Subnational governments and education administrators

5.1. Why results-based financing for subnational governments and education administrators?

Incentives for subnational governments and education administrators can help prioritize education outcomes in the policy agenda and among different actors involved in the education provision. Subnational governments and education administrators or institutions are involved with the education policy either through a direct role (i.e., local and regional supervisors, municipal and regional education offices) or indirectly, as actors in charge of public administration (i.e., mayors, heads of villages, and provinces). Also called as meso-level actors or meso-level education management, these positions are not composed by frontline workers in the education sector, but can motivate, guide, and support teachers and principals, holding them accountable for educational outcomes. They also provide a means for education systems to regularly monitor student enrollment, retention, and learning. Estimates that 84 percent of the world's children live in countries where subnational governments are the leading providers of pre-university education only reinforce meso-levels' relevance to leverage education outcomes (Al-Samarrai & Lewis, 2021).

Incentives for subnational governments and education administrators can have several designs. For example, they can take the form of performance bonuses, conditional grants, performance-based contracts, or intergovernmental transfers. Incentives can focus on access and use of services (i.e., student enrollment, inputting data in monitoring systems, involving the community in school management decisions) or to improve services quality (incentives focused on student retention and learning outcomes). They can also cascade the effects of incentives to other education stakeholders (as incentives to regional offices and schools to leverage learning).

RBF for subnational governments and education administrators is relatively new to the education sector. A qualitative study funded by REACH identified that RBF to meso-level actors is predominantly from the health sector, and most research in the education field relates to higher education mechanisms implemented in the United States. RBF initiatives reported in developing countries are often focused on primary education and improving management processes, school access, student participation, and learning outcomes, with a predominance of interventions in Sub-Saharan Africa and South Asia (Terway et al., 2021).

5.2. What results can incentives for subnational governments and education administrators achieve?

As incentives can focus on different actors, this section presents the evidence based on the two types of stakeholders incentivized. It focuses firstly on education administrators, as regional education offices and supervisors, and secondly on subnational governments, as villages, municipalities, and provinces.

5.2.1. Incentives targeting regional education administrators

An intervention in Brazil and another in India combined incentives for teachers and education officers at regional levels. They found positive results, although it was not possible to disentangle the impact of education administrators. RBF for meso-levels provides the opportunity to align incentives between actors in the education system. At the time of this writing, there were only two impact evaluations of interventions targeting regional or school supervisors but they also included teachers among the treatment groups and were unable to isolate the effect of meso-level actors.²⁷ The bonus program for teachers and school principals in São Paulo rewarded regional education bureaus according to average regional results, and increased learning outcomes in grade 5 (Lépine, 2021; Oshiro & Scorzafave, 2015). In India, an intervention funded by REACH rewarded regional supervisors for improving teachers' attendance. Supervisors received training²⁸ and non-financial incentives in two treatment groups, but one group also incentivized teachers, as an attempt to balance power asymmetries between supervisors (who have short-term contracts) and teachers (who are civil servants). The group with incentivized teachers had 15% higher attendance rates (World Bank, 2021).

In Peru and Ethiopia, incentives for regional and local education offices aimed at improving management processes and monitoring systems to lay the foundations for learning accountability. In Peru, the MoE linked additional budget to regional and local offices according to the achievement of a set of targets ("*Compromisos de desempeño*") related to improvements in data, management and pedagogical processes, such as hiring teachers and distributing textbooks on time. The reward was proportional to the number of targets achieved and the money could be used for any educational expenditure. To support education offices, the MoE simplified purchasing processes and modernized data systems, introducing dashboards that allowed just-in-time monitoring. Since its introduction in 2014, there have been significant improvements in the areas targeted by the program. In 2007, learning outcomes were included as targets. Although it was not possible to isolate the effect of incentives on learning, between 2014-2016 there was an increase in test scores between 9-18% in

²⁷ Establishing causality between incentives and performance for education regional offices and supervisors can be significantly challenging. The main challenge is establishing a control group, either ex-ante or ex-post. On the one hand, it is politically difficult to set an incentive only to a subgroup of offices and treat the other offices as control groups. On the other hand, if the policy is implemented to all, it is hard to find a comparison group, even if the policy is implemented in only one province, as the context of one province can be very different from another. This happens in places where education provision is decentralized, where the role of regional supervisors/offices can vary substantially between provinces.

²⁸ Focused on motivating RPs and improving their ability to conduct classroom observations, provide feedback to teachers, and use technology for data entry.

reading comprehension, and between 11-26% in math (Correa Miranda, García Medina & Ugarte Vera. 2017. *in* Terway et al., 2021). In Ethiopia, a pilot RBF results-based aid cascaded incentives for the MoE and regional education bureaus to increase the number of additional sitters and passers on the 10th-grade national exam, improving learning monitoring. There was an increase in students taking the test in this period, but it was not possible to attribute it to the RBF mechanism (Cambridge Education, 2015).

RBF to education administrators is currently being applied in Benin, Nigeria, and Tanzania to improve student enrollment and retention, especially for closing gender gaps. These interventions highlight interesting incentive designs, but their effects will only be available in forthcoming years. In Benin, two incentives for sub-regional inspectors were introduced, one tied to improving the quality of their support to schools (including pedagogical support), and the other to stimulating schools to raise student enrollment and retention, especially among girls (Global Partnership for Education, 2020). In Nigeria, the MoE has linked an intergovernmental transfer for state education boards to increases in student enrollment, especially among girls and rural students ([Better Education Service Delivery for All](#), 2017-2022). In Tanzania, the MoE established two performance-based transfers to local government authorities. One releases funds to schools that improve the pupil-teacher ratio (35-50:1). The other rewards local governments with the greatest improvement in completion rates in primary and lower secondary and in girls' transition rates from primary to secondary ([Education Program for Results](#), 2015-2022).

5.2.2. Incentives targeting subnational governments

Rigorously assessed interventions in Brazil and Indonesia show mixed impacts of incentivizing local governments. However, they do highlight the effect of providing pedagogical support and the importance of targeting a limited number of indicators. Although incentives for local governments have been applied in other settings, such as Senegal and India, at the time of this review, there were only two publicly available robust evaluations. An intergovernmental transfer to villages in Indonesia was attached to improvements in 12 indicators ranging from health to education, but it had positive outcomes only in health and in the short term. The high number of performance indicators led villages to focus on areas they perceived easier to change, such as hiring more midwives at the expense of teachers, similar to what happened with some teacher bonus policies (Barrera-Osorio & Raju, 2017; Bellés-Obrero & Lombardi, 2021; McEwan, Patrick J. & Santibanez, Lucrecia, 2005). In Ceará, Brazil, another intergovernmental transfer was attached to learning outcomes. Almost simultaneously, the state also provided pedagogical support to municipalities to help them raise outcomes. It found that providing technical assistance almost doubled the effects of RBF, which is in line with results from another intervention that combined school tournaments with peer-monitoring (Goldemberg et al., 2021).

In Indonesia, attaching transfers to performance on 12 indicators (8 on health and 4 on education) had a positive effect only on health and in the short term. The intervention introduced a performance-based grant to villages aimed at improving a set of 12 indicators for health and education. The performance mechanism consisted of allocating 20 percent of an annual block grant

according to villages' relative performance on each of the indicators monitored by the program. There were four educational indicators: student enrollment and student attendance for primary and for lower secondary levels. A randomized control trial showed that the program contributed to greater initial performance on health indicators, but non-incentivized villages caught up at the endline. On education, although student enrollment increased, there was no significant difference between incentivized and non-incentivized villages. Additionally, the incentives led to a reallocation of funds away from education supplies toward health (Olken et al., 2014). Further research on the education indicators show that the initiative increased community and parental participation in school management, which can be seen as intermediate outcomes towards increased enrollment (Aizawa, 2019).

In the state of Ceará, Brazil, combining results-based fiscal incentives with pedagogical support almost doubled programs' impact. The state of Ceará made impressive gains in learning outcomes in just over a decade, despite its low socioeconomic conditions. Ceará is not a privileged locale: it has the fifth-lowest GDP per capita out of 26 states, departed from very low education outcomes, and has a per-student expenditure significantly low compared to other subnational governments in the country. Its success lies in a set of policy reforms, including an incentive mechanism for mayors to raise municipal revenues by improving learning outcomes (Loureiro et al, 2020).²⁹ Ceará innovated in fiscal transfers by making 18 percent of consumption taxes due to municipalities dependent on their learning outcomes. The change in the redistribution of the consumption tax generated a significant incentive for mayors and local government staff, in addition to education actors.³⁰ The fiscal incentive was combined with a structured pedagogy program that provided technical assistance to municipal governments to address heterogeneous capacity in implementing education policies.³¹ Comparing municipalities at the border of Ceará state in different milestones of the policies,³² an impact evaluation funded by REACH identified that students in grade 9 exposed to the RBF mechanism scored 0.15 SD higher in mathematics and Portuguese, which is equivalent to an additional 3 months of learning (Lautharte et al., 2021). When the incentive was combined with pedagogical support, impacts for grade 9 almost doubled and impacts for grade 5 were positive and significant. The gains of combining the incentive and TA for both grades are equivalent to 5 months of learning in Portuguese and 3 months in mathematics on the top of the effects of the incentive alone. Furthermore, despite a 1.2% increase on per-student expenditure in the first year of the policy, learning outcomes were

²⁹ Ceará has also (1) devolved the management of primary and lower secondary education to municipalities, granting them high autonomy in the design and implementation of the education policy, (2) established a regular learning monitoring system, and (3) provided pedagogical assistance to municipal networks to support those with lower technical capacity.

³⁰ In Brazil, there is administrative autonomy among federal, state and municipal governments, and primary and lower secondary education are mainly under municipal responsibility, creating ideal conditions for incentives to mayors/municipalities.

³¹ The pedagogical assistance program included a set of initiatives in structured pedagogy, such as providing scripted materials and training for teachers, textbooks for students and strengthening the pedagogical use of learning assessments. For details, see two reports funded by REACH: Loureiro, Cruz, Lautharte & Evans (2020). *The State of Ceara in Brazil is a Role Model for Reducing Learning Poverty*. World Bank, Washington, DC. World Bank and Loureiro, Alves, Cruz, Assunção & Cardoso (2020). *Technical Assistance for Local Governments to Improve Education Outcomes: An implementation guide inspired by the case of Ceará, Brazil*. World Bank.

³² The impact evaluation uses low-stakes exams and explores different policy milestones, as both the fiscal incentive and the pedagogical support started in 2008 with a focus on the literacy cycle (grades 1-2), and in 2012 the focus has shifted to grade 5. The pedagogical support was also expanded to lower secondary education (grade 9).

attained without an increase in municipal public spending.³³ Going on fourteen years, this incentive scheme is one of the longest running interventions for education administrators. Currently, Ceará is the Brazilian state with the highest increase in the national education quality index in both primary and lower secondary education, with 10 municipalities being among the top 20 national ranking, and most of its 10-year-old kids able to read and understand a simple text, as measured by World Bank's learning poverty indicator (Loureiro et al., 2020).

Several pedagogical practices seem to have contributed to Ceará's positive results, suggesting that complementary pedagogical interventions strengthen RBF's impact. The impact evaluation highlighted four potential reasons for the positive results seen in Ceará. The first relates to good practices in choosing school principals, as Ceará municipalities are more likely to adopt formal selection processes (10–20 percentage points (pp) in comparison to border municipalities pertaining to other states) instead of making political appointments. Secondly, school principals and teachers were more likely to enroll in training and more often reported that training was useful in their daily work (7.8pp for 5th grades and 12pp for 9th grades), which reflects the state's initiative to provide a structured pedagogy program with regular training and focus on classroom practice. Third, the pedagogical support seems to have contributed to better provision of instruction material, as teachers in Ceará are 10-25pp less likely to report lack of textbooks in the school and 9.5pp more likely to report that textbooks are good or great resources. Lastly, in Ceará, teachers were more likely to cover at least 80% of the school curriculum, which again is likely to be related to the structured pedagogy program (Lautharte et al., 2021). Together, these mechanisms point to the relevance of aligning the incentive mechanism to initiatives that strengthen school management and pedagogy, by providing adequate instruction material and training. These aspects are commonly found in structured pedagogy programs, which are considered one of the most promising interventions to leverage student learning (Evans & Mendez Acosta, 2021).

Other countries have performance-based transfers, although they are often more modest in size and do not have, yet, a study of their impacts. In Uganda and Colombia, performance-based intergovernmental transfers in education account for less than 10% of the education budget of subnational governments (Al-Samarrai & Lewis, 2021). In Brazil, a recent amendment to the federal constitution introduced determined that (1) all states must transfer between 10-35 percent of the consumption tax due to municipalities according to education outcomes. The other distributes to municipalities a part of a federal transfer earmarked to education³⁴ according to improvements in education outcomes (Emenda Constitucional nº 108, 2020). Also, in Brazil, the MoE has recently approved a new high school curricular reform that considers a full-time school shift. To incentivize state governments (the main providers of secondary school) to implement full-time school shifts, the

³³ In terms of incentive-design, the initial incentive rule gave more weight to learning gains to stimulate low-performing municipalities to respond to the incentive. As learning outcomes have increased, the incentive rule has increased the importance of learning levels by considering the share of students with adequate learning in relation to the whole distribution of students. On learning inequality, the incentive-mechanism seems to have increased learning gaps between high- and low-performing students, which was softened when the incentive rule started penalizing municipalities with substantial shares of students with low scores. At the municipal level, in the medium-term, Ceará's reforms benefited poor municipalities and also municipalities that initially lost resources when the redistribution rule started prioritizing education outcomes (Brandão, 2014).

³⁴ The transfer corresponds to the federal top-up to the National Fund for Education Development (FUNDEB).

MoE offered additional monetary resources, conditional on the submission and approval of an implementation plan. These plans must meet a set of requirements, including establishing an operational unit with staff dedicated to supporting the policy's implementation at the state level ([Support to Upper Secondary Reform in Brazil Operation](#), 2018-2023). In Senegal, a multi-donor development project ([Quality Improvement and Equity of Basic Education](#), 2014-2021) had three layers of RBF: incentives for regional authorities, district authorities, and schools. For the two first groups, incentives were established in the form of performance-based contracts, where money is transferred upon achievement of results. District offices were incentivized to improve management practices, the quality of support given to schools, and student attendance. Regional offices were incentivized to strengthen the monitoring and evaluation process of district performance indicators and learning outcomes, and retention rates.

5.3. Design and implementation features of incentives for subnational governments and education administrators

RBF for meso-levels can have multiple designs reflecting the existent variety in how education systems are organized, but a pre-condition is having some level of autonomy in at least some aspects of education policies' design and implementation. Setting incentive mechanisms involves adjustments to contextual challenges, such as multiple principle-agent relationships, the need for operational funds to achieve results, the lack of a monitoring system to set a pure performance-based indicator, and different autonomy levels of incentivized actors. One might think that these adjustments distort RBF initiatives but, according to specialists involved in the policy design, they actually contribute to strengthening a monitoring and accountability culture and align policy efforts, which lays the ground for future and more ambitious RBF designs (Terway et al., 2021).

The evidence available provides insightful considerations about meso-level incentives' design (detailed in Annex 2). On the metrics, rewarding percentage improvement can engage those actors with low baseline performance. Some interventions have also tried to address inequality gaps by choosing indicators of marginalized groups (i.e. number of girls enrolled in schools) or providing technical support in pedagogical and managerial aspects to promote fair play. A qualitative study on RBF to education administrators proposes six recommendations for the design and implementation processes, as summarized in box 3.

Box 3 - Recommendations for designing and implementing RBF to meso-level actors

Recommendation 1: Conduct a situation analysis to understand the educational challenges that results-based financing can help solve, identify any existing experience with results-based financing, and understand whether there is a culture conducive to results-based management. The latter can be done through creative processes, such as games played on mobile phones, to understand individuals' perceptions and factors that facilitate and constrain the use of RBF (World Bank, 2020).

Recommendation 2: Agree on a shared theory of change (or a clear results-chain) with all stakeholders, especially those setting the incentive and those who will be incentivized. This process is also crucial to ensure sufficient autonomy for the incentivized actors to achieve results.

Recommendation 3: Define results targets that can be transparently and objectively measured.

Recommendation 4: Design an incentive structure that cascades financial and non-financial incentives at organizational and individual levels. It does not mean establishing organizational and individual rewards but identifying what motivates individuals and organizations to pursue the established targets.

Recommendation 5: Provide capacity-building support and technical assistance to foster a culture of accountability. This can be done firstly with a small government team that will be responsible for designing the incentive - as seen in a World Bank initiative in Morocco (World Bank, 2019c) - and later with the implementation actors - as seen in the case of Ceará.

Recommendation 6: Ensure sustainability by setting legislation to regulate the RBF mechanism or with a plan to phase out external funding over time.

Adapted from (Terway et al., 2021)

6.4. Conclusion on RBF for subnational governments and education administrators

Practical experience of RBF with meso-levels has shown that incentives frequently target intermediate outcomes, such as improvements in management processes, data systems, and increases in student enrollment. These adjustments reflect contextual challenges but are perceived as a significant contribution to strengthening a monitoring and accountability culture and aligning policy efforts, contributing to laying the ground for future and more ambitious RBF designs.

Robust impact evaluations of RBF to meso-levels have focused on interventions aiming to improve student learning. They combined incentives for meso-levels with incentives for teachers or pedagogical support. While they had a positive impact, the limited number of robust analyses prevent us from making conclusive considerations. A highlight is the state of Ceará, in Brazil, with a long-lasting incentive policy for municipal governments that is aligned with other education provisions as technical and pedagogical support, adequate instruction material and training, and strengthened school management.

6. Concluding remarks

RBF has been applied to the education sector as a way to improve service delivery, among other things. As education provision is fragmented among different actors (such as teachers, textbook providers, and education bureaus that establish curricular and learning assessment standards), it is sometimes difficult to know what or who to incentivize in order to improve learning. This paper has shown that RBF can be used as a financing modality to strengthen accountability and service delivery by promoting a focus on the actions necessary to improve learning and reinforcing the capacity of education systems to measure and track progress.

The economic and education scenario arising from the COVID-19 pandemic raises the salience of efficiency and effectiveness in education spending. Since the onset of COVID-19, the education budget has shrunk in 65% of developing countries (Al-Samarrai et al., 2021). Additionally, school closures and deficiencies in remote learning provision generated a massive learning loss, increasing the share of children in learning poverty (World Bank et al., 2021). This scenario urges more efficient use of resources to recover learning, assure children the right to education, and foster human capital.

The evidence base of RBF in education continues to increase at a fast pace. Still, there is a significant imbalance in the literature, with more research readily available on teachers and student-family incentives. The newest evidence on these two actors sheds light on whether RBF contributes to student learning and the long-term effects of interventions. Design and implementation aspects continue to have great relevance for the success of initiatives, something that is almost intrinsic to public policy. Different interventions show that aligning incentives of more than one agent involved in education can be more effective than a scheme targeting a single actor. Likewise, combining incentives with appropriate support helps engage actors, especially those starting from a low baseline.

As per each type of intervention analyzed, the takeaways and areas for future research are:

Teachers:

- Incentives designed to improve teacher attendance can raise attendance when strong monitoring systems are in place. In addition, incentives whose primary purpose was to increase student learning have been shown to increase test scores (12 out of 15 interventions had positive outcomes), but design and implementation matter for effectiveness. A qualitative analysis of programs' characteristics indicates that interventions tend to be successful when: i) schools have minimum learning resources, ii) the reward rule clearly puts learning as the main outcome indicator, and iii) teachers are able (or receive support) to set specific targets for their students and articulate them to learning strategies.
- So far, teacher incentives have been largely implemented as a bonus policy instead of changing teachers' contracts. Pay for performance contracts can sort teachers more responsive to incentives, but evidence on selecting more qualified teachers is mixed.
- On the design, it is not clear if pay-for-percentile is the best metric, as positive evidence came from Math teachers in China. In fact, most of the positive evidence, both from experiments and scaled-up interventions, use simpler and group-based designs. Programs implemented by

governments and on a large scale include multiple outcome indicators but those with a positive impact on learning are the ones that had a substantial part of the incentive attached to student test scores. Additionally, given the heterogeneity among schools, it is not trivial to design a large-scale mechanism in which all schools perceive they can have a fair competition. To be sustainable and fully effective, incentives typically require an accountability mechanism that helps shape teachers' behavior in order to achieve long-term positive effects.

- On gaps, it is still unclear if individual incentives are better than group incentives since there is an unbalance in the evidence: experiments have focused on individual incentives, whereas large-scale interventions adopted group-based designs. It would be interesting to test a pay-for-percentile scheme on a scale. Also, little is known about the long-term effects of teacher incentives.

Students and families:

- Incentives can foster schooling and learning, but the design rule generally targets one over the other. Conditional cash transfers can increase enrollment, attendance, and completion rates, especially in primary and lower secondary education. In high school, short-term average impacts on schooling are mixed, though there are positive effects for the most vulnerable students. In the long run, children exposed to CCTs during primary education have increased upper secondary enrollment and completion, probably due to the increased salience of education returns. Evidence on learning effects is mixed, with small positive impacts.
- Merit-based incentives tied to students' test scores are still growing, with promising results. Since evidence on these programs is not extensive and encompasses various designs, it is difficult to generalize findings on which design works better, highlighting the need for more research on the topic. One more salient aspect in current research is that grouping students with different learning levels seem more effective than individual rewards.

Schools:

- Robust evidence on school incentives is currently restricted to two studies on performance-based school grants that use a tournament design and reward based on learning outcomes. Although limited, evidence on RBF for schools seems promising, especially when the incentive is combined with capacity-building interventions, such as peer-mentoring to principals of high- and low-performing schools.
- The lack of robust evidence does not mean that performance grants are rare. In fact, there are several incentives to schools that have been implemented in development projects, generally adopting intermediary outcomes. This highlights the relevance of the topic and the need to expand the evidence base, including assessing its long-term effects.

Subnational governments and education administrators:

- It is challenging to assess impacts of incentives for subnational governments and education administrators (also called meso-level actors) since most policies are implemented at scale. Robust impact evaluations have focused on interventions aiming to improve student learning that combined meso-level incentives with teacher incentives or pedagogical support. Although there are documented positive impacts, the limited number of analyses prevents any broader conclusions. A highlight is the state of Ceará, in Brazil, with a long-lasting incentive

policy aligned with other education provisions such as technical and pedagogical support, adequate instruction material and training, and strengthened school management.

- Practical experience has shown that incentives frequently target intermediate outcomes, such as improvements in management processes, data systems, and increases in student enrollment. These adjustments reflect contextual challenges but are perceived as a significant contribution to strengthening a monitoring and accountability culture and aligning policy efforts, contributing to laying the ground for future and more ambitious RBF designs.

7. References

- Abadzi, H. (2007). *Absenteeism and Beyond: Instructional Time Loss and Consequences*. World Bank. <https://doi.org/10.1596/1813-9450-4376>
- Aguilar, A. A., Barnard, C., & De Giorgi, G. (2019). *Long-Term Effects of Prospera on Welfare* (SSRN Scholarly Paper ID 3450898). Social Science Research Network. <https://papers.ssrn.com/abstract=3450898>
- Aizawa, T. (2019). Impacts of the Community Block Grant Programme on School Resources, Environment and Management in Indonesia. *Education Economics*, 27(5), 521–545.
- Akresh, R., & de Walque, D. (2013). *Cash Transfers and Child Schooling: Evidence from a Randomized Evaluation of the Role of Conditionality* (No. 6340; Policy Research Working Paper, p. 57). World Bank. <https://openknowledge.worldbank.org/handle/10986/13127>
- Al-Samarrai, S., Cerdan-Infantes, P., Bigarinova, A., Bodmer, J., Vital, M., Antoninis, M., Barakat, B., & Murakami, Y. (2021). *Education Finance Watch 2021* [Text/HTML]. World Bank. <https://documents.worldbank.org/en/publication/documents-reports/documentdetail/226481614027788096/Education-Finance-Watch-2021>
- Al-Samarrai, S., & Lewis, B. (2021). *The Role of Intergovernmental Fiscal Transfers in Improving Education Outcomes*. World Bank. <https://doi.org/10.1596/978-1-4648-1693-2>
- Al-Samarrai, S., Shrestha, U., Hasan, A., Nakajima, N., Santoso, S., & Wisnu, A. (2018). *Indonesia: Can Performance-Based School Grants Improve Learning?* (No. 124424; pp. 1–8). The World Bank. <http://documents.worldbank.org/curated/en/121371521520039305/Indonesia-Can-Performance-Based-School-Grants-Improve-Learning>
- Attanasio, O., Sosa, L. C., Medina, C., Meghir, C., & Posso-Suárez, C. M. (2021). *Long Term Effects of Cash Transfer Programs in Colombia* (Working Paper No. 29056; Working Paper Series). National Bureau of Economic Research. <https://doi.org/10.3386/w29056>
- Baez, J. E., & Camacho, A. (2011). *Assessing the Long-Term Effects of Conditional Cash Transfers on Human Capital: Evidence from Colombia*. The World Bank. <https://doi.org/10.1596/1813-9450-5681>
- Baird, S., Ferreira, F. H. G., Özler, B., & Woolcock, M. (2014). Conditional, Unconditional and Everything in Between: A Systematic Review of the Effects of Cash Transfer Programs on Schooling Outcomes. *Journal of Development Effectiveness*. <https://openknowledge.worldbank.org/handle/10986/18085>
- Baird, S. J., Chirwa, E., de Hoop, J., & Özler, B. (2013). *Girl Power: Cash Transfers and Adolescent Welfare. Evidence from a Cluster-Randomized Experiment in Malawi* (Working Paper No. 19479; Working Paper Series). National Bureau of Economic Research. <https://doi.org/10.3386/w19479>
- Baird, S., McIntosh, C., & Özler, B. (2011). Cash or Condition? Evidence from a Cash Transfer Experiment. *The Quarterly Journal of Economics*, 126(4), 1709–1753. <https://doi.org/10.1093/qje/qjr032>
- Baird, S., McIntosh, C., & 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. <https://doi.org/10.1016/j.jdeveco.2019.04.004>
- Barham, T., Macours, K., & Maluccio, J. A. (2017). *Are Conditional Cash Transfers Fulfilling Their Promise? Schooling, Learning, and Earnings after 10 Years* (SSRN Scholarly Paper ID 2941523). Social Science Research Network. <https://papers.ssrn.com/abstract=2941523>

- Barrera-Osorio, F., de Barros, A., & Filmer, D. (2018). *Long-Term Impacts of Alternative Approaches to Increase Schooling: Evidence from a Scholarship Program in Cambodia*. World Bank, Washington, DC. <https://doi.org/10.1596/1813-9450-8566>
- Barrera-Osorio, F., & Filmer, D. (2016). Incentivizing Schooling for Learning: Evidence on the Impact of Alternative Targeting Approaches. *Journal of Human Resources*, 51(2), 461–499. <https://doi.org/10.3368/jhr.51.2.0114-6118R1>
- Barrera-Osorio, F., Linden, L. L., & 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. <https://doi.org/10.1257/app.20170008>
- Barrera-Osorio, F., & Raju, D. (2017). Teacher Performance Pay: Experimental Evidence from Pakistan. *Journal of Public Economics*. <https://doi.org/10.1016/j.jpubeco.2017.02.001>
- Bastagli, F., Hagen-Zanker, J., Harman, L., Barca, V., Sturge, G., & Schmidt, T. (2016). *Cash transfers: What does the evidence say? A rigorous review of programme impact and of the role of design and implementation features* (p. 300). Overseas Development Institute. <https://cdn.odi.org/media/documents/11316.pdf>
- Behrman, J. R., Parker, S. W., & Todd, P. (2019). *Impacts of PROSPERA on Enrollment, School Trajectories, and Learning*. 42.
- Behrman, J. R., Parker, S. W., Todd, P. E., & Wolpin, K. I. (2015). Aligning Learning Incentives of Students and Teachers: Results from a Social Experiment in Mexican High Schools. *Journal of Political Economy*, 123(2), 325–364. <https://doi.org/10.1086/675910>
- Bellés-Obrero, C., & Lombardi, M. (2021). Teacher Performance Pay and Student Learning: Evidence from a Nationwide Program in Peru. *Economic Development and Cultural Change*, 714012. <https://doi.org/10.1086/714012>
- Benhassine, N., Devoto, F., Duflo, E., Dupas, P., & Pouliquen, V. (2015). Turning a Shove into a Nudge? A “Labeled Cash Transfer” for Education. *American Economic Journal: Economic Policy*, 7(3), 86–125. <https://doi.org/10.1257/pol.20130225>
- Bernal, Sebastian, & Celhay. (2018). *Is Results-Based Aid More Effective than Conventional Aid?: Evidence from the Health Sector in El Salvador*. Inter-American Development Bank. <https://publications.iadb.org/en/results-based-aid-more-effective-conventional-aid-evidence-health-sector-el-salvador>
- Berry, J. (2015). Child Control in Education Decisions: An Evaluation of Targeted Incentives to Learn in India. *Journal of Human Resources*, 50(4), 1051–1080. <https://doi.org/10.3368/jhr.50.4.1051>
- Berry, J., Kim, H. B., & Son, H. (2019). *When Student Incentives Don't Work*. 59.
- Birdsall & Savedoff. (2011). *Cash on Delivery: A New Approach to Foreign Aid*. Center for Global Development. <https://www.cgdev.org/publication/9781933286600-cash-delivery-new-approach-foreign-aid>
- Blimpo, M. P. (2014). Team Incentives for Education in Developing Countries: A Randomized Field Experiment in Benin. *American Economic Journal: Applied Economics*, 6(4), 90–109. <https://doi.org/10.1257/app.6.4.90>
- Blimpo, M. P., Evans, D., & Lahire, N. (2015). *Parental Human Capital and Effective School Management: Evidence from The Gambia* [Working Paper]. World Bank. <https://doi.org/10.1596/1813-9450-7238>

- Bo, Finan, & Rossi. (2013). Strengthening state capabilities: The role of financial incentives in the call to public service. *The Quarterly Journal of Economics*.
- Brandão, J. B. (2014). *O rateio de ICMS por desempenho de municípios no Ceará e seu impacto em indicadores do sistema de avaliação da educação*.
<http://bibliotecadigital.fgv.br/dspace/handle/10438/13149>
- Emenda Constitucional nº 108, no. Constituição Federal de 1988 (2020).
<https://legislacao.presidencia.gov.br/atos/?tipo=EMC&numero=108&ano=2020&ato=30dkXWE1UMZpWTcc0>
- Breeding, M., Béteille, T., & Evans, D. (2021). *Teacher Pay-for-Performance: What Works? Where? And How?* World Bank. <https://documents1.worldbank.org/curated/en/183331619587678000/pdf/Teacher-Pay-for-Performance-What-Works-Where-and-How.pdf>
- Brollo, F., Maria Kaufmann, K., & La Ferrara, E. (2020). Learning Spillovers in Conditional Welfare Programmes: Evidence from Brazil. *The Economic Journal*, 130(628), 853–879.
<https://doi.org/10.1093/ej/ueaa032>
- Brown, C., & Andrabi, T. (2021). *Inducing Positive Sorting through Performance Pay: Experimental Evidence from Pakistani Schools*. 83.
- Bruns, B., & Luque, J. (2015). *Great Teachers: How to Raise Student Learning in Latin America and the Caribbean*. World Bank. <https://doi.org/10.1596/978-1-4648-0151-8>
- Cabrera, & Webbink. (2016). *Do higher salaries yield better teachers and better student outcomes?*
<http://www2.um.edu.uy/jmcabrera/Research/Funding%20poor%20schools.pdf>
- Cahyadi, N., Hanna, R., Olken, B. A., Prima, R. A., Satriawan, E., & Syamsulhakim, E. (2020). Cumulative Impacts of Conditional Cash Transfer Programs: Experimental Evidence from Indonesia. *American Economic Journal: Economic Policy*, 12(4), 88–110. <https://doi.org/10.1257/pol.20190245>
- Cambridge Education. (2015). *Evaluation of the pilot project of results based aid in the education sector in Ethiopia*. 70.
- Carneiro, P., Koussihouédé, O., Lahire, N., Meghir, C., & Mommaerts, C. (2020). School Grants and Education Quality: Experimental Evidence from Senegal. *Economica*, 87(345), 28–51.
<https://doi.org/10.1111/ecca.12302>
- Chang, F., Brooks, C. D., Springer, M. G., Liu, H., & Shi, Y. (2021). The effect of participation in a performance pay program on teacher opinions toward performance pay in rural China. *Education Economics*, 0(0), 1–27. <https://doi.org/10.1080/09645292.2021.1924623>
- Chang, F., Wang, H., Qu, Y., Zheng, Q., Loyalka, P., Sylvia, S., Shi, Y., Dill, S.-E., & Rozelle, S. (2020). The impact of pay-for-percentile incentive on low-achieving students in rural China. *Economics of Education Review*, 75, 101954. <https://doi.org/10.1016/j.econedurev.2020.101954>
- Chaudhury, N., Hammer, J., Kremer, M., Muralidharan, K., & Rogers, F. H. (2006). Missing in Action: Teacher and Health Worker Absence in Developing Countries. *Journal of Economic Perspectives*, 20(1), 91–116. <https://doi.org/10.1257/089533006776526058>
- Chen, Glewwe, Kremer, & Moulin. (2001). *Interim Report on a Teacher Attendance Incentive Program in Kenya*. NBER.
http://users.nber.org/~dlchen/papers/Interim_Report_on_a_Teacher_Attendance_Incentive_Program_in_Kenya.pdf

- Chewla, Pellicer, & Maboshe. (2019). Teachers pay and educational outcomes: Evidence from the rural hardship allowance in Zambia. *South African Journal of Economics*, 87.
- Chhabra, E., Najeeb, F., & Raju, D. (2019). *Effects Over The Life Of A Program: Evidence From An Education Conditional Cash Transfer Program For Girls*. World Bank, Washington, DC.
<https://doi.org/10.1596/1813-9450-9094>
- Cilliers, J., Kasirye, I., Leaver, C., Serneels, P., & Zeitlin, A. (2018). Pay for locally monitored performance? A welfare analysis for teacher attendance in Ugandan primary schools. *Journal of Public Economics*, 167(C), 69–90.
- Clist, P. (2019). Payment by results in international development: Evidence from the first decade. *Development Policy Review*, 37(6), 719–734. <https://doi.org/10.1111/dpr.12405>
- Clist & Verschoor. (2014). *The Conceptual Basis of Payment by Results*. University of East Anglia.
<https://www.gov.uk/research-for-development-outputs/the-conceptual-basis-of-payment-by-results>
- Contreras, D., & Rau, T. (2012). Tournament Incentives for Teachers: Evidence from a Scaled-Up Intervention in Chile. *Economic Development and Cultural Change*, 61(1), 219–246.
<https://doi.org/10.1086/666955>
- Costa, L. O., & Carnoy, M. (2015). The Effectiveness of an Early-Grade Literacy Intervention on the Cognitive Achievement of Brazilian Students. *Educational Evaluation and Policy Analysis*, 37(4), 567–590.
<https://doi.org/10.3102/0162373715571437>
- Crespo, C. (2019). *Cash for Grades or Money for Nothing? Evidence from Regression Discontinuity Designs*. 54.
- Cruz, L., & Loureiro, A. (2020). *Achieving World-Class Education in Adverse Socioeconomic Conditions: The Case of Sobral in Brazil*. World Bank. <https://openknowledge.worldbank.org/handle/10986/34150>
- Damon, A., Glewwe, P., Wisniewski, S., Sun, B., Sverige, & Expertgruppen f??r bist??ndsanalys. (2016a). *Education in developing countries—What policies and programmes affect learning and time in school?* Expertgruppen f??r bist??ndsanalys (EBA).
- Damon, A., Glewwe, P., Wisniewski, S., Sun, B., Sverige, & Expertgruppen f??r bist??ndsanalys. (2016b). *Education in developing countries—What policies and programmes affect learning and time in school?* EBA. <https://www.oecd.org/derec/sweden/Rapport-Education-developing-countries.pdf>
- Das, J., Dercon, S., Habyarimana, J., Krishnan, P., Muralidharan, K., & Sundararaman, V. (2013). School Inputs, Household Substitution, and Test Scores. *American Economic Journal: Applied Economics*, 5(2), 29–57. <https://doi.org/10.1257/app.5.2.29>
- De Walque, D. B. C. M., & Valente, C. (2018). *Mozambique: Can information and incentives increase school attendance?* (No. 128153; pp. 1–8). The World Bank.
<http://documents.worldbank.org/curated/en/704081531204680698/Mozambique-can-information-and-incentives-increase-school-attendance>
- Dom, C., Fraser, A., Holden, J., & Patch, J. (2021). *Results-Based Financing (RBF) in the Education Sector: Country-Level Analysis—Final Synthesis Report (English)*. World Bank.
<https://documents1.worldbank.org/curated/en/250441637213208744/pdf/Results-Based-Financing-RBF-in-the-Education-Sector-Country-Level-Analysis-Final-Synthesis-Report.pdf>
- Draeger, E. (2021). Do conditional cash transfers increase schooling among adolescents?: Evidence from Brazil. *International Economics and Economic Policy*. <https://doi.org/10.1007/s10368-021-00505-6>

- Duflo, E., Hanna, R., & Ryan, S. P. (2012). Incentives Work: Getting Teachers to Come to School. *American Economic Review*, 102(4), 1241–1278. <https://doi.org/10.1257/aer.102.4.1241>
- Duque, V., Rosales-Rueda, M., & Sanchez, F. (2019). How Do Early-Life Shocks Interact with Subsequent Human Capital Investments? Evidence from Administrative Data. In *Working Papers* (No. 2019–17; Working Papers). University of Sydney, School of Economics. <https://ideas.repec.org/p/syd/wpaper/2019-17.html>
- 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. <https://doi.org/10.1016/j.jdeveco.2019.102392>
- Evans, D. K., Gale, C., & Kosec, K. (2021). *The Educational Impacts of Cash Transfers for Children with Multiple Indicators of Vulnerability* (No. 563; Working Paper). Center for Global Development. <https://www.cgdev.org/sites/default/files/educational-impacts-cash-transfers-children-multiple-indicators-vulnerability-revised-march2021.pdf>
- Evans, D. K., & Mendez Acosta, A. (2021). Education in Africa: What Are We Learning? *Journal of African Economies*, 30(1), 13–54. <https://doi.org/10.1093/jae/ejaa009>
- Evans, D. K., & Yuan, F. (2019). *What We Learn about Girls' Education from Interventions that Do Not Focus on Girls* (No. 513; Working Paper). Center for Global Development. <https://www.cgdev.org/publication/what-we-learn-about-girls-education-interventions-do-not-focus-on-girls>
- Filmer, D., Habyarimana, J., & Sabarwal, S. (2020). Teacher Performance-Based Incentives and Learning Inequality. *Policy Research Working Paper*, No. 9382, 35.
- Filmer, D., & Schady, N. (2009). *School Enrollment, Selection and Test Scores* (No. 4998; Policy Research Working Paper, p. 43). World Bank. <https://openknowledge.worldbank.org/bitstream/handle/10986/4190/WPS4998.pdf>
- Filmer, D., & Schady, N. (2014). The Medium-Term Effects of Scholarships in a Low-Income Country. *Journal of Human Resources*, 49(3), 663–694. <https://doi.org/10.3368/jhr.49.3.663>
- Friedman, W., Kremer, M., Miguel, E., & Thornton, R. (2016). Education as Liberation? *Economica*, 83(329), 1–30. <https://doi.org/10.1111/ecca.12168>
- Gaduh, A., Pradhan, M., Priebe, J., & Susanti, D. (2021). *Scores, Camera, Action: Social Accountability and Teacher Incentives in Remote Areas*. The World Bank. <https://doi.org/10.1596/1813-9450-9748>
- Ganimian, A., & Murnane, R. (2014). *Improving Educational Outcomes in Developing Countries: Lessons from Rigorous Impact Evaluations* (No. w20284; p. w20284). National Bureau of Economic Research. <https://doi.org/10.3386/w20284>
- García, S., Harker, A., & Cuartas, J. (2019). Building dreams: The short-term impacts of a conditional cash transfer program on aspirations for higher education. *International Journal of Educational Development*, 64, 48–57. <https://doi.org/10.1016/j.ijedudev.2018.12.006>
- García-Moreno, V., Gertler, P., & Patrinos, H. A. (2019). *School-Based Management and Learning Outcomes: Experimental Evidence from Colima, Mexico*. 29.
- Gazeaud, J., & Ricard, C. (2021). *Conditional Cash Transfers and the Learning Crisis: Evidence from Tayssir Scale-up in Morocco*. 55.

- Gertler, P. J., Patrinos, H. A., & Rubio-Codina, M. (2012). Empowering parents to improve education: Evidence from rural Mexico. *Journal of Development Economics*, 99(1), 68–79. <https://doi.org/10.1016/j.jdeveco.2011.09.004>
- Gilligan, D. O., Karachiwalla, N., Kasirye, I., Lucas, A. M., & Neal, D. (2019). Educator Incentives and Educational Triage in Rural Primary Schools. *Journal of Human Resources*, 1118. <https://doi.org/10.3368/jhr.57.1.1118-9871R2>
- Glewwe, P., Ilias, N., & Kremer, M. (2010). Teacher Incentives. *American Economic Journal: Applied Economics*, 2(3), 205–227. <https://doi.org/10.1257/app.2.3.205>
- Glewwe, P., & Muralidharan, K. (2016). Chapter 10 - Improving Education Outcomes in Developing Countries: Evidence, Knowledge Gaps, and Policy Implications. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the Economics of Education* (Vol. 5, pp. 653–743). Elsevier. <https://doi.org/10.1016/B978-0-444-63459-7.00010-5>
- Global Partnership for Education. (2020). *An Early Stage Review of Country Program Designs and Implementation Experiences with GPE's Variable Part Financing Mechanism (2015–2019)*. Global Partnership for Education. <https://www.globalpartnership.org/sites/default/files/document/file/2020-05-review-of-design-and-implementation-experiences-of-gpe-variable-part-financing.pdf>
- Goldemberg, D., Bacalhau, P., & Junior, I. J. L. (2021). *Can peer mentoring coupled with incentives affect school turnaround? Evidence from Ceará state in Brazil*. 17.
- Gustafsson-Wright, & Boggild-Jones. (2019). *Paying for Education Outcomes at Scale in India*. Center for Universal Education at Brookings. <https://files.eric.ed.gov/fulltext/ED602924.pdf>
- Hirshleifer, S. R. (2017). *Incentives for Effort or Outputs? A Field Experiment to Improve Student Performance*. 47.
- Holanda, M., Barbosa, M., Cruz, L., & Loureiro, A. (2021). *Implementing a Results-Based Financing Mechanism for Subnational Governments to Improve Education Outcomes: An Implementation Guide Inspired by the Case of Ceará, Brazil*. World Bank. <https://documents1.worldbank.org/curated/en/561471606111232725/pdf/Implementing-a-Results-Based-Financing-Mechanism-for-Subnational-Governments-to-Improve-Education-Outcomes-An-Implementation-Guide-Inspired-by-the-Case-of-Ceara-Brazil.pdf>
- Kremer, M., Miguel, E., & Thornton, R. (2009). Incentives to Learn. *Review of Economics and Statistics*, 91(3), 437–456. <https://doi.org/10.1162/rest.91.3.437>
- Lautharte, I., de Oliveira, V. H., & Loureiro, A. (2021). *Incentives for Mayors to Improve Learning: Evidence from State Reforms in Ceará, Brazil*. The World Bank. <https://doi.org/10.1596/1813-9450-9509>
- Leaver, C., Ozier, O., Serneels, P., & Zeitlin, A. (2021). Recruitment, effort, and retention effects of performance contracts for civil servants: Experimental evidence from Rwandan primary schools. *ArXiv:2102.00444 [Econ, q-Fin]*. <http://arxiv.org/abs/2102.00444>
- Lee, L. J. D., & Medina, O. (2019). *Results-Based Financing in Education: Learning from What Works* (No. 133932; pp. 1–95). The World Bank. <http://documents.worldbank.org/curated/en/915061548222619389/Results-Based-Financing-in-Education-Learning-from-What-Works>
- Lépine, A. (2021). Teacher Incentives and Student Performance: Evidence from Brazil. *Education Finance and Policy*, 1–55. https://doi.org/10.1162/edfp_a_00339

- Li, T., Han, L., Zhang, L., & Rozelle, S. (2014). Encouraging classroom peer interactions: Evidence from Chinese migrant schools. *Journal of Public Economics*, *111*, 29–45. <https://doi.org/10.1016/j.jpubeco.2013.12.014>
- Loureiro, A., Cruz, L. R. da, Lautharte, I., & Evans, D. (2020). *The State of Ceara in Brazil is a Role Model for Reducing Learning Poverty*. World Bank. <https://openknowledge.worldbank.org/handle/10986/34156>
- Loyalka, P., Sylvia, S., Liu, C., Chu, J., & Shi, Y. (2019). Pay by Design: Teacher Performance Pay Design and the Distribution of Student Achievement. *Journal of Labor Economics*, *37*(3), 621–662. <https://doi.org/10.1086/702625>
- Mbiti, I., Muralidharan, K., Romero, M., Schipper, Y., Manda, C., & Rajani, R. (2019). Inputs, Incentives, and Complementarities in Education: Experimental Evidence from Tanzania*. *The Quarterly Journal of Economics*, *134*(3), 1627–1673. <https://doi.org/10.1093/qje/qjz010>
- Mbiti, I., Romero, M., & Schipper, Y. (2019). *Designing Effective Teacher Performance Pay Programs: Experimental Evidence from Tanzania* (No. w25903; p. w25903). National Bureau of Economic Research. <https://doi.org/10.3386/w25903>
- Mbiti, I., & Schipper, Y. (2021). Teacher and Parental Perceptions of Performance Pay in Education: Evidence from Tanzania. *Journal of African Economies*, *30*(1), 55–80. <https://doi.org/10.1093/jae/ejaa012>
- McEwan, Patrick J., & Santibanez, Lucrecia. (2005). Teacher and Principal Incentives in Mexico. In *Incentives to Improve Teaching: Lessons from Latin America*. Emiliana Vegas. <https://openknowledge.worldbank.org/handle/10986/7265>
- Mhagama, M. (2020). An empirical Overview on the Implementation of Big Results Now Initiative in Tanzania and its Efficacy on Academic Performance in Secondary Schools. *Journal of Advances in Education and Philosophy*, *04*(01), 1–7. <https://doi.org/10.36348/jaep.2020.v04i01.001>
- Mitnick, B. M. (1975). *The Theory of Agency: A Framework* (SSRN Scholarly Paper ID 1021642). Social Science Research Network. <https://doi.org/10.2139/ssrn.1021642>
- Molina Millán, T., Barham, T., Macours, K., Maluccio, J. A., & Stampini, M. (2019). Long-Term Impacts of Conditional Cash Transfers: Review of the Evidence. *The World Bank Research Observer*, *34*(1), 119–159. <https://doi.org/10.1093/wbro/lky005>
- Molina Millán, T., Macours, K., Maluccio, J. A., & Tejerina, L. (2019). *The long-term impacts of Honduras' CCT program: Higher education and international migration* (Working Paper IDB-WP-907). IDB Working Paper Series. <https://doi.org/10.18235/0001670>
- Molina Millán, T., Macours, K., Maluccio, J. A., & 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. <https://doi.org/10.1016/j.jdeveco.2019.102385>
- Muralidharan, K. (2012). *Long-Term Effects of Teacher Performance Pay: Experimental Evidence from India*. 43.
- Muralidharan, K., & Sundararaman, V. (2011). *Teacher Performance Pay: Experimental Evidence from India* (Working Paper No. 15323; Working Paper Series). National Bureau of Economic Research. <https://doi.org/10.3386/w15323>
- Olken, B. A., Onishi, J., & Wong, S. (2014). Should Aid Reward Performance? Evidence from a Field Experiment on Health and Education in Indonesia. *American Economic Journal: Applied Economics*, *6*(4), 1–34. <https://doi.org/10.1257/app.6.4.1>

- Oshiro, C., & Scorzafave, L. (2015). Impacto Sobre o Desempenho Escolar do Pagamento de Bônus aos Docentes do Ensino Fundamental do Estado de São Paulo. *Revista Brasileira de Economia*, 69. <https://doi.org/10.5935/0034-7140.20150010>
- Perez-Alvarez, M., Priebe, J., & Susanti, D. (2020). *Teacher Accountability and Pay-for-Performance Schemes in (Semi-) Urban Indonesia: What Do Education Stakeholders Think?* World Bank. <https://doi.org/10.1596/33376>
- Rocha, R. H., Menezes-Filho, N., & Komatsu, B. K. (2018). *Avaliando o Impacto das Políticas de Sobral* (No. 35; Policy Paper, p. 33). Insper Centro de Políticas Públicas. <https://www.insper.edu.br/wp-content/uploads/2018/10/Avaliando-o-Impacto-das-Poli%CC%81ticas-de-Sobral.pdf>
- Romero, M., Bedoya, J., Yanez-Pagans, M., Silveyra, M., & de Hoyos, R. (2021). *School Management, Grants, and Test Scores: Experimental Evidence from Mexico*. The World Bank. <https://doi.org/10.1596/1813-9450-9535>
- Sabarwal, S., & Abu-Jawdeh, M. (2018). *What Teachers Believe: Mental Models about Accountability, Absenteeism, and Student Learning* [Working Paper]. World Bank. <https://doi.org/10.1596/1813-9450-8454>
- Sanchez Chico, A., Macours, K., Maluccio, J. A., & Stampini, M. (2020). Impacts on school entry of exposure since birth to a conditional cash transfer programme in El Salvador. *Journal of Development Effectiveness*, 12(3), 187–218. <https://doi.org/10.1080/19439342.2020.1773900>
- Schipper, Y., & Pradhan, M. (2022). *RBF interventions in education: Do they increase inequality of outcomes? - A literature review*. World Bank.
- Snilstveit, B., Stevenson, J., Menon, R., Phillips, D., Gallagher, E., Geleen, M., Jobse, H., Schmidt, T., & Jimenez, E. (2016). *The impact of education programmes on learning and school participation in low- and middle-income countries: A systematic review summary report* (3ie Systematic Review Summary No. 7). International Initiative for Impact Evaluation (3ie).
- Social Finance. (2022). *Learning lessons for education from the use of results-based financing (RBF) in health: A review of the literature*. World Bank.
- Terway, A., Burnett, N., & Dreux Frotte, M. (2021). *Results-based financing in education for sub-national government and school administrators—A Conceptual Framework and Practical Recommendations* (Working Paper No. 12). NORRAG. <https://resources.norrag.org/resource/689/results-based-financing-in-education-for-sub-national-government-and-school-administrators-a-conceptual-framework-and-practical-recommendations>
- Unesco. (2014). *Teaching and learning: Achieving quality for all; EFA global monitoring report, 2013-2014*. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000225660>
- United Nations. (2015). *Transforming our world: The 2030 Agenda for Sustainable Development*. https://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E
- Whetten, J., Fontenla, M., & Villa, K. (2019). Opportunities for higher education: The ten-year effects of conditional cash transfers on upper-secondary and tertiary enrollments. *Oxford Development Studies*, 47(2), 222–237. <https://doi.org/10.1080/13600818.2018.1539472>
- World Bank. (2018). *World Development Report 2018: Learning to Realize Education's Promise* (Doi: 10.1596/978-1-4648-1096-1). World Bank. <http://www.worldbank.org/en/publication/wdr2018>
- World Bank. (2019a). *Cameroon: Can School Grants and Teacher Incentives be Used to Increase School Access and Improve Quality?* (Results-Based Financing in Education | Evidence Note). World Bank.

<https://documents1.worldbank.org/curated/en/304191556557313186/pdf/Cameroon-Can-School-Grants-and-Teacher-Incentives-be-Used-to-Increase-School-Access-and-Improve-Quality.pdf>

World Bank. (2019b). *Cameroon: Results-Based Grants and Teacher Incentives Improve Performance at the School Level* (Results-Based Financing in Education | Impact Notes). World Bank.

<https://documents1.worldbank.org/curated/en/440131576835439403/pdf/Cameroon-Results-Based-Grants-and-Teacher-Incentives-Improve-Performance-at-the-School-Level.pdf>

World Bank. (2019c). *Marocco: Supporting the Design of Performance-Based Contracts to Improve Results in Education*. World Bank.

<https://documents1.worldbank.org/curated/en/813581576839048549/pdf/Morocco-Supporting-the-Design-of-Performance-Based-Contracts-to-Improve-Results-in-Education.pdf>

World Bank. (2020). *Can Understanding How Middle Managers Make Decisions Help Design Effective Results-Based Financing Mechanisms in Education?* World Bank, Washington, DC.

<https://doi.org/10.1596/33636>

World Bank. (2021). *India: Can Results-Based Incentives Encourage Teachers to Attend School?* World Bank.

<https://openknowledge.worldbank.org/bitstream/handle/10986/35964/India-Can-Results-Based-Incentives-Encourage-Teachers-to-Attend-School.pdf?sequence=1&isAllowed=y>

World Bank, UNESCO, & UNICEF. (2021). *The State of the Global Education Crisis: A Path to Recovery*. World Bank, UNESCO, and UNICEF. <https://documents.worldbank.org/en/publication/documents-reports/documentdetail/416991638768297704/the-state-of-the-global-education-crisis-a-path-to-recovery>

Annex 1 – Summary table of teacher incentives focused on student learning

Intervention year	Country	Design	Metric	Tournament	Focus	Outcome indicator	Impact overall	Estimates	Type	Paper
1993	Mexico	Individual	Average scores	No	Learning, career	Student learning, teacher test scores	Null	-	Large scale	(McEwan & Lucrecia, 2005)
1996	Chile	Group incentive	Levels and gains	Yes	Learning	Learning, school level indicators, equity	Positive	0.14 SD in language, 0.25 SD I math (high stakes test)	Large scale	(Contreras & Rau, 2012)
1998	Kenya	Group incentive	Absolute scores, gains	Yes	Learning	Learning	Positive	0.14 SD (high stakes)	RCT	(Glewwe et al., 2010)
2005	India	Group vs individual	Gains	No	Learning	Learning	Positive	After 2 years: 0.28 SD in math, 0.16 SD in language. After 5 years: 0.54 SD math, 0.35 SD language (high stakes)	RCT	(Muralidharan & Sundararaman, 2011)
2008	Mexico	Individual (teacher vs student)	Threshold	No	Learning	Learning	Positive (combined and students-only).	After 3 years: 0.23 SD students-only 0.57 SD teachers + students	RCT	(Behrman et al., 2015)
2008	Brazil	Group incentive	Levels (target)	No	Learning	Learning, teacher attendance	Positive for 5th grade	After 2 years: 0.42 SD math, 0.14 SD language (low stakes)	Large scale	(Oshiro & Scorzafave, 2015)
2010	Pakistan	Group (teachers vs principals)	Gains	No	Learning, enrollment	Learning, enrollment, test participation	Null	-	RCT	(Barrera-Osorio & Raju, 2017)
2013	Tanzania	Individual (teacher vs student)	Gains	Yes	Learning	Learning	Positive for teachers	0.15 SD in math and language (teachers+students year 1) & 0.13SD (teachers year 2)	RCT	(Filmer et al., 2020)
2013	Tanzania	Individual vs school grants	Threshold	No	Learning, inputs	Learning	Positive (combined and only bonus)	High stakes: 0.21 SD (teacher bonus), 0.36 SD (bonus + grants) Low stakes: 0.23 SD (bonus + grants)	RCT	(Mbiti, Muralidharan, et al., 2019)
2013	China	Individual	Levels, gains, vs percentile	Yes	Learning	Learning	Percentile positive, levels and gains null	0.15 SD in math (percentile group)	RCT	(Loyalka et al., 2019)

2014	China	Individual	Percentile	Yes	Learning, low-achieving students	Learning	Positive	0.1 SD in math (avg students) and 0.15 SD bottom 30%	RCT	(Chang et al., 2020)
2015	Tanzania	Individual	Threshold vs percentile	Yes	Learning	Learning	Positive in both metrics	*Levels: 0.19 SD language, 0.14 SD in math *Percentile: 0.1 SD language, 0.09 SD math *Treatment effect of levels was 0.09 SD higher and statistically significant than percentile	RCT	(Mbiti, Romero, et al., 2019)
2015	Peru	Group incentive	Absolute scores	Yes	Learning, dropout	Learning, dropout, school management	Null	-	Large scale	(Bellés-Obrero & Lombardi, 2021)
2016	Indonesia	Individual	Threshold	No	Learning, social accountability	Learning, teacher attendance	Positive (SAM & SAM+Cam)	0.2 SD (SAM+Cam), 0.11 SD (SAM)	RCT	(Gaduh et al., 2021)
2016	Uganda	Individual	Percentile	Yes	Learning, dropout	Learning, dropout	Positive only for high-performing students with books	0.17 SD on questions covering textbooks' content	RCT	(Gilligan et al., 2019)

Annex 2 – Summary on design and implementation features

The discussion on the effects of incentives to specific actors already elicited some design and implementation features that contribute to a program’s impact. This annex gathers the main concerns raised by policy makers and researchers when designing and implementing RBF mechanisms and organized the evidence available for each topic and incentivized actor.

1. Teachers

As explained in Lee and Medina (2019), incentives for teachers can reward based on individual or group performance:

- **Group incentives:** may favor cooperation between teachers since student learning is beyond teachers’ control, but also induce free-riding which may reduce incentive’s impacts.
- **Individual incentives:** connect with a person’s effort, but may be challenging to scale-up since it requires calculating the reward for each teacher.

On the metrics, rewards can consider levels, gains or rank.

- **Levels:** provide an objective measure to teachers, which may help them to design strategies to improve student learning. It can involve a single target, as the number of students that passed an exam, or a multiple threshold, as learning scales. The disadvantage is that teachers have an incentive to focus on students that are closer to meeting the target(s), which may jeopardize low-achieving pupils far from the threshold(s).
- **Gains:** try to favor the opposite, since students with very low levels will have a greater percentage increase than students with already good performance, but once students start improving their performance, the incentive may become less attractive.
- **Rank:** reward the top performers, usually a defined number, which makes it cost-predictable.
- **Pay-for-percentile:** A hybrid method that conciliates gains and rank features. Students are tested at the beginning of the year and gathered according to their levels. At the end of the year, students are tested again and teacher bonuses are calculated based on students’ ranking position.

Table 1 - Design features of teacher incentives tested in different interventions

Pre-conditions	Technical capacity, resources and political will are three preconditions for successful incentive mechanisms (Breeding et al., 2021).
Size	The size of the incentives does not seem to have a large impact on the size of the effect, and should match the country's context. As identified in 18 studies, incentive size ranges from 1% to 20% of teachers’ annual salary, with median at 4%, and interventions with no effects had incentive size above the average of 7% (Barrera-Osorio & Raju, 2017; Bellés-Obrero & Lombardi, 2021; McEwan, Patrick J. & Santibanez, Lucrecia, 2005).

<p>Type of incentive (cash/in-kind)</p>	<p>Monetary incentives are the most common type, although two interventions have used in-kind incentives (Filmer et al., 2020; Glewwe et al., 2010).</p> <p>Most of the interventions are designed as bonus programs, while some have changed teachers' contracts or salary raise (Brown & Andrabi, 2021; Leaver et al., 2021; McEwan, Patrick J. & Santibanez, Lucrecia, 2005) and one treated low-performance as a penalty, reducing the allowance pre-destined to teachers (Gaduh et al., 2021).</p>
<p>What to incentivize</p>	<p>Incentivizing attendance can improve attendance if strong accountability mechanisms are in place and if attendance has significant relevance in the reward formula. Effects on learning are promising.</p> <p>Incentivizing learning outcomes points to positive effects on learning outcomes, as, out of the 15 interventions whose reward formula focuses on learning, 12 have some positive effects. Effectiveness seems to be related to assuring schools have minimum pedagogical resources, the reward rule clearly puts learning as the main outcome indicator, and teachers understand the mechanism well enough to reshape pedagogical strategies and improve its performance.</p> <p>Introducing performance-pay in teachers' contracts can help sort teachers who are more responsive to incentives and impacts on learning are promising.</p>
<p>Who to incentivize: individual or group based</p>	<p>Both individual and group-based incentives have been shown to have positive effects, although randomized control trials generally adopt individual incentives (especially after Muralidharan and Sundararaman (2011) found higher impacts on individual over group-incentives. In contrast, three out of four interventions implemented on a scale use group incentives.</p>
<p>Which metrics to use: level, gains, and percentiles</p>	<p>Unclear. Pay-per-percentile has been successful in China among Math teachers (Chang et al., 2020; Loyalka et al., 2019) and in interventions looking at sorting effects of performance-based contracts (Brown & Andrabi, 2021; Leaver et al., 2021) but not so successful in Uganda and Tanzania.</p> <p>From an implementation perspective, incentives attached to learning levels or ranks are easy to administer, since they require only one exam at the end of the school year, while learning gains and pay-for-percentile require a baseline and end line. Levels and gains are also easier to communicate, as they are frequently used in many areas (Lee & Medina, 2019). For instance, large scale interventions tend to use levels: Chile had positive outcomes with tournaments but not Peru (Bellés-Obrero & Lombardi, 2021; Contreras & Rau, 2012).</p> <p>As seen in São Paulo, Brazil that established learning targets for each school, setting a threshold may need adjustments over time to assure engagement: if targets are too easy there is no point in establishing the mechanism. If targets are too hard, teachers may not respond to the incentive. Overall, there seems to exist a trade-off between design complexity and ease of use, which involves both teachers' understanding and implementation issues.</p>

Learning inequality	<p>Not all interventions report results on learning inequality and the existent evidence have different designs, preventing us from reaching a conclusion on which metric works better. The pay-for-percentile design tries to direct teachers' efforts towards all students and has a positive impact in China (Loyalka et al., 2019). Another experiment in China that also adopted a pay-for-percentile scheme gave extra bonus for leveraging results of low-performing students (Chang et al., 2020). Nevertheless, in Tanzania, pay-for-percentile led teachers to focus on high-performing students, while the threshold scheme had learning gains more equally distributed (Mbiti, Romero, et al., 2019). A tournament in Tanzania that rewarded schools with highest learning gains (which presumably favors schools with low performing students) only impacted student learning in schools with higher baseline performance (Filmer et al., 2020).</p>
Accountability mechanisms	<p>Although the focus of many interventions is student learning, it is important to monitor program implementation and results with multiple indicators (Breeding et al., 2021). This also helps teachers' buy-in. In Pakistan, learning outcomes were higher among the group in which performance was assessed by supervisors (and included other indicators as attendance and teaching practices) than assessing it only through test scores (Barrera-Osorio & Raju, 2017). In Indonesia, the treatment group that had performance monitored by cameras and by the community had better outcomes (Gaduh et al., 2021). Other successful interventions have also combined classroom observations and unannounced visits to check teachers' attendance (Leaver et al., 2021; Muralidharan & Sundararaman, 2011).</p>
Control for adverse behaviors	<p>The most frequent adverse behaviors reported are cheating, teaching to the test, test manipulation, free-riding and moral hazard (Breeding et al., 2021), and considering design and implementation features to control adverse behavior can raise the incentive's effectiveness. Some examples of strategies to prevent those behaviors are: in Tanzania, a picture was taken of each student to prevent replacing students in the test. Also, as the incentive was threshold-based, only students enrolled at the baseline could take the test to prevent adding students with higher performance (Mbiti, Romero, et al., 2019). In India, to prevent teachers from discouraging low-performing students to take the exam, in case students had a negative outcome, the reduction in teachers' bonus was capped to -5%, and this same amount was established for students that missed the test (Muralidharan, 2012).</p>
Sustainability	<p>The sustainability of a program requires legal procedures, establishing a solid monitoring system, having technical capacity to operate the incentive program and having fiscal space for the payments (Breeding et al., 2021). Transitioning from NGO-led to government-led interventions may also compromise programs' effects. In Indonesia, the first year of the intervention was implemented by a NGO and the second year by the government. Results for the second year are lower (Gaduh et al., 2021).</p>

Source: Lee & Medina (2019) with additions based on the papers cited in this report.

2. Students and families

As design features of incentives for student learning were explicitly discussed in the main text, this annex focuses only on conditional cash transfers.

Table 2 - Design issues of CCTs

Conditionality	<p>Both conditional and unconditional transfers have effects of similar magnitude in general, but some unconditional interventions have smaller effects on education outcomes (S. Baird et al., 2014). In Morocco, labeled cash transfers yielded similar results in school participation than CCT (Benhassine et al., 2015). On monitoring conditionality, when neighbors or friends receive a warning about their non-compliance, positive effects are also noticed among their peers (Brollo et al., 2020).</p> <p>In Colombia, conditioning part of the transfer to high school re-enrollment or completion improved programs' effectiveness. However, linking transfers to tertiary enrollment promoted enrollment in lower-quality tertiary institutions in the medium term but not in the long term (Barrera-Osorio et al., 2019).</p>
Information and labeling	<p>Information treatments (for example, providing an attendance report card) can have positive effects that complement the transfer itself (De Walque & Valente, 2018; Duque et al., 2019). Labeling cash transfers as educational transfers have positive outcomes, increasing the perceived importance of education (Benhassine et al., 2015). The perception of a conditionality (which sometimes may not even exist) and well-communicating the desired behavior change enhances the program's effectiveness.</p>
Who to incentivize	<p>There seem to be no significant differences between fathers and mothers (with some exceptions). Giving kids part of the transfer may increase the magnitude of the effect (De Walque & Valente, 2018).</p>
Share of students enrolled	<p>The lower the initial share of students enrolled, the higher the effect on enrollment (Ganimian & Murnane, 2014).</p>
Size of transfer	<p>Larger transfers do not always cause larger effects, diminishing returns (Filmer & Schady, 2014; Ganimian & Murnane, 2014).</p>
Timing of transfer	<p>Aligning transfers with school fees' deadline helps to maximize programs' education impacts. Delaying part of the payment and making it conditional on next grade enrollment increases retention (Barrera-Osorio et al., 2019). There is limited evidence that more frequent student rewards improve learning outcomes more than one large reward (Ganimian & Murnane, 2014).</p>
Age and grade of recipient	<p>CCTs improve enrollment, dropout, completion and transition rates in both primary and lower secondary education. Evidence for programs targeting high school students is mixed. Long-term analyses of children benefitting from CCTs since early grades point to higher secondary school completion rates. Outcomes improve with larger exposure to cash transfers.</p>
Poverty level	<p>The poorer the beneficiaries, the larger the impact, especially on schooling. For learning, the impact seems more prominent among those with previous higher performance (Evans et al., 2021).</p>

Source: Lee & Medina (2019) with additions based on the papers cited in this report.

3. Schools

Some design issues related to school incentives were already elicited in teachers' table, such as those referring to pre-conditions, accountability mechanisms, and adverse behaviors. The table below tries to discuss what is more specific to school grants based on the two robust impact evaluations available and on the analysis of similar literature conducted by Lee and Medina (2019).

Table 3 - Summary on design features of performance-incentives to schools

Pre-conditions	As for teacher incentives, RBF to schools requires technical capacity to design and implement the mechanism, financial resources and political will, a robust monitoring system (including regular learning assessments), and a defined process to transfer awards. As seen in a pilot project in Mozambique, monitoring implementation and verifying performance progress in a context of low capacity is a challenge, pointing to the need to simplify verification means and performance indicators.
Size	Existent studies provide different metrics. In Jakarta, the performance bonus was equivalent to 20% of the standard unconditional bonus received by schools. In Ceará, the bonus ranged from less than 5% to 10% of the school budget. In both cases, the bonus was set on a per-student basis and linked to the number of students taking the learning assessment.
Type of incentive	In both cases, monetary incentives were adopted, although programs attribute considerable importance to the social status of being recognized as a top-performing school.
What to incentivize and which metrics to use	The evidence so far has focused on learning outcomes , both levels, and gains. Learning gains, the percentage point improvement, try to stimulate low-performing schools particularly. Learning levels include some mechanism to encourage sustained performance. In Jakarta, the formula considers average examination performance over two years. In Ceará, part of the grant to top-performing schools is conditional on maintaining or improving performance.
Who to incentivize	School grants are group-based per definition, while the mechanism per se does not always include indicators directly related to all school dimensions or all school actors. Attaching the reward rule to learning outcomes assumes that better student performance results from several activities, such as decent infrastructure, satisfied nutritional needs, adequate teaching, and a healthy school environment.

Learning inequality	In both studies analyzed, schools competed for the resources, but the incentive design included mechanisms to mitigate learning inequality. In Jakarta, the rank was calculated for each district to account for socioeconomic disparities, and the reward formula included learning gains to incentivize the participation of low-performing schools. In Ceará, three features tried to account for learning inequality. First, the rank formula gave more weight to learning gains, benefiting low-performing schools. Second, the program paired low-performing schools with high-performing ones to increase turnaround opportunities. Third, low-performing schools received 50% of their grants at the beginning of the school year, enabling them to make small purchases that could strengthen their pedagogical practices.
Accountability mechanisms	In Ceará, the school grant is part of a broad technical assistance program to help municipalities improve learning outcomes, and the program has multiple accountability mechanisms. Some examples are (i) a set of diagnostic evaluations implemented twice a year, (ii) a matrix listing all literacy competencies per student, allowing teachers to map individual progress, (iii) a routine of regular meetings with principals and the heads of municipal education departments to discuss learning outcomes, and (iv) classroom observations and visits to schools.
Adverse behaviors	Many adverse behaviors listed for group incentives for teachers focusing on student performance apply for school results-based incentives, such as teaching to the test, test manipulation, free-riding, and moral hazard. Additionally, there is some evidence from non-performance school grants that the possibility of receiving extra funds may discourage parents from investing in education (Das et al., 2013).
Long-term effects	Research on this aspect is needed. Evidence from health interventions suggests that grants can improve organizational constraints that have a sustained effect in the long run (Bernal et al., 2018). The intervention in Ceará is also promising since it enhances schools' practices without dismissing teachers.

Source: Lee & Medina (2019) with additions based on the papers cited in this report.

4. Education administrators

Table 4 - Summary table on design features of incentives for education administrators

Pre-conditions	The autonomy of meso-level actors in designing and implementing at least certain aspects of the education policy is a pre-condition for performance-based incentives. As detailed in the teachers' and schools' sections, technical capacity, political will, and monitoring systems are also vital. In Morocco, the World Bank, with REACH funding, provided training to strengthen government officials' capacity to design and enact performance contracts (World Bank, 2019c).
Size	The size varies greatly with the design. For instance, the project in Ethiopia considered a per-student rate starting at £50, with higher amounts for underrepresented groups, with a total amount of up to £10 million per year. According to the number of indicators achieved, Peru established a minimum and maximum amount. In Indonesia, block grants averaged US\$ 8.5 and 13.5 thousand dollars per village in the first and second years of the program. In Ceará, 18% of the value-added tax due to municipalities was attached to education results. In the first years of the policy, this transfer represented up to 70 percent of total revenue for some small cities with the greatest improvements. In Benin, the grant considered a fixed amount. In São Paulo, the grant varied with average results.
Type of incentive	RBF to meso-levels can take the form of performance bonuses, conditional grants, performance-based contracts, or intergovernmental transfers.
What to incentivize	Access and use of services (i.e., student enrollment, especially among underrepresented groups, inputting data in monitoring systems, involving the community in school management decisions) or improvements in services' quality (incentives focused on student retention and learning outcomes, or reducing teacher-per-pupil ratio).
Which metrics to use	Metrics vary with the focus of the incentive: 1) Average: São Paulo rewarded according to the regional average in learning outcomes. 2) Threshold: India gave certificates according to teacher attendance (70, 80, or 90 percent attendance). In Peru, education offices had to achieve a set of targets with different weights. At the beginning of the policy, offices received the money proportional to the number of targets achieved. Nowadays, some targets consider partial payment proportional to improvements. 3) Improvements: In Ethiopia, part of the transfer was proportional to the percentage increase in the number of students sitting and passing the test.
Who to incentivize	Incentives in meso-level positions can focus on stakeholders directly involved in education (i.e., local and regional supervisors, municipal and regional education offices) or actors in charge of public administration (i.e., mayors, heads of villages, and provinces). Incentives can also occur in a cascade of incentives, aligning efforts among teachers, schools, and administrative actors.

Addressing inequality	<p>Some RBF interventions have tried to close inequality gaps by incentivizing indicators directly related to marginalized groups, such as giving more weight to girls' enrollment and retention (as in Benin, Nigeria, and Tanzania).</p> <p>Technical support in pedagogical or managerial aspects is seen as a relevant component to assure that incentivized agents have conditions to respond to incentives and mitigate inequalities in a technical capacity. This was seen in government-led initiatives in Peru and Ceara and development projects, as in Ethiopia and India.</p> <p>In Ceará, the first reward formula gave more weight to learning gains to make the incentive more attractive to smaller and poorer municipalities and mitigate inequalities between cities. Moreover, to prevent disparities between students in the same town, the formula also adjusted average grades for each city by the standard deviation of student grades so that higher averages obtained at the cost of learning inequalities are penalized. However, research indicates that this mechanism was not totally effective, as estimates identified effects three to four times higher for students from students in the top 0.2 quantiles compared to the bottom 0.2. This inequality was partially overtaken by a change in the reward formula that penalized municipalities with students performing below the minimum threshold (Lautharte et al., 2021).</p>
Adverse behaviors	<p>In Nepal and Bangladesh, an innovative initiative funded by REACH used a game to identify principals' and district officers' perceptions that could constrain the use of RBF. Some of those findings include (i) the perception that a school can do very little to help students if parents don't get involved in children's education, (ii) the tendency to prioritize students with more potential, and the feeling of having limited ability to help all students to learn, (iii) the reluctance to hold teachers to high standards. These aspects can be seen as warnings of adverse behavior, and measures to mitigate them should be considered when designing the incentive (World Bank, 2020).</p> <p>In Ceará, in addition to preventing the focus on the top-performing students, the incentive also accounted for the possibility of sending only the best students to take the exam. Thus, the rank formula adjusted average grades by the ratio between the number of students that took the exams and those enrolled at the beginning of the school year (Holanda et al., 2021).</p> <p>The intervention in Indonesia has shown that mixing incentives for achieving results in different sectors (as health and education) can cause reallocation of resources to areas that are perceived to be easier to improve outcomes.</p>
Long-term effects or sustainability	<p>A key aspect for sustainability is setting legislation to regulate the incentive, as seen in Peru, São Paulo, Ceará, and Indonesia. Introducing performance rules for existent transfers also fosters sustainability as it does not require extra funds (besides those related to monitoring performance indicators). Ceará has the longest incentive policy (14 years of existence) whose impacts have been rigorously assessed, and it is the state in Brazil with the highest increase in education performance in the last decade.</p>

Source: Author's elaboration based on the papers cited in this report.