What Are We Learning from Impact Evaluation Studies about Teacher Incentives and Education Quality

Received: 29 October 2023; Revised: 14 January 2024; Published: 02 February 2024

Jin Chi

Faculty of Education, Beijing Normal University, China
Email: chij@bnu.edu.cn
ORCID: https://orcid.org/0000-0002-6980-3149

Eduardo Velez Bustillo

Graduate School of International Cooperation Studies, Kobe University, Japan
Email: evelezbustillo@hotmail.com
ORCID: https://orcid.org/0000-0002-7355-8940

Abstract:

The quality of teachers is a determinant to the quality of the education system. Despite the significant increase in resources invested in teacher incentives, there is little robust empirical evidence on their effectiveness, especially showing attribution. This paper examines the international literature using Impact Evaluation (IE) studies to see what lessons can be learned, focusing specifically on a key education input: teacher incentives as a policy to motivate teachers and improve education quality in middle- and low-income developing countries. The review shows there are limited number of IE studies on teacher incentives at the preschool, primary and secondary education levels applying robust experimental design and the results are mixed. When incentives are used properly they can have a positive impact on the quality of education. However, certification itself is not enough as a way to improve the quality of the education system; it is critical to develop reforms explicitly geared towards quality. Finally, although incentives can reduce teacher absenteeism, they do not always improve the quality of education outcomes.

Keywords: teacher incentives, impact evaluations, education quality, education system

1. Introduction

The quality of an education system is synonymous to the quality of its teachers. Empirical evidence in countries of different economic development levels shows this correlation (Barber & Mourshed, 2007; Lee & Koh, 2020). Studies in developing countries (low- and middle-income countries) have also repeatedly found the relevance of teachers to the quality of the education system. For example, the difference between a weak and good teacher on student test scores is estimated at 0.36 standard deviations (SDs) in Uganda (Buhl-Wiggers et al., 2017) and 0.54 SDs in Pakistan (Bau & Das, 2017). This means that having a good teacher versus a weak one during a school year is equivalent to a difference of more than two years of schooling (Evans & Yuan, 2018). It has also been shown that the impact of good teachers goes beyond cognitive achievement. Chetty, Friedman, and Rockoff (2014) found that in the United States, teachers' value-added in a single academic year explains behavioral outcomes in later life.

In a world that is changing rapidly with technological development as well as knowledge production and innovation, the demand of the labor market is reacting accordingly. However, education systems are not keeping up with the new demand. Countries need to keep improving human capital to remain competitive in this changing world. Thus, the education sector needs to synchronize elements of the system like accountability, monitoring and evaluation, as well as the quality of teacher's education and education materials.

The premises of this study are that the quality of teachers is critical to the quality of the education system and that investment in human capital is one of the most effective ways for countries to improve their national welfare and reduce poverty in the long term. Thus, identifying lessons that improve teachers' incentives may help countries use scarce education resources efficiently. The objectives of teacher incentive programs are to help teachers increase effort and gain effectiveness in teaching, which are strongly associated with student learning outcomes (Hanushek & Rivkin, 2006; Glewwe & Kremer, 2006). Interestingly, despite the significant increase in resources invested in teacher incentives in developing countries, there is little robust empirical evidence on their effectiveness, especially when showing attribution.

There is much at stake. An effective teacher impacts the lives of students by making it possible for them to learn and gain skills. An ineffective teacher not only wastes their limited time of potential learning but also fails to facilitate students with the competency they should have and the learning skills in subsequent years to live successful lives (Darling-Hammond, et. al., 2017).

In many low- and middle-income countries, a significant number of teachers are not effective in supporting students with the skills needed[1]. Therefore, it is important to analyze the teacher incentive

[1] This is particularly important because there are growing expectations for teachers increasingly emphasizing the importance of balancing expert content delivery with nurturing the social-emotional skills that are also important for students' lifelong success. In addition, teachers are expected to play several roles ranging from content expert, curriculum developer, and pedagogue, to social worker, psychologist, mentor, and motivator. Every teacher has dimensions of this set of skills and they can potentially improve them through many tools. Its objective is to align the support provided to individual teachers to their unique challenges and needs.

programs and learn from the successful ones. Although most of these programs are not evaluated following rigorous methodologies, several solid studies allow us to identify patterns of successful programs.

How to design teacher incentive programs is not straightforward. Across the world, there are programs delivered by governmental entities or managed by NGOs, with ample variations in the contents. Incentives can include monetary and non-monetary dimensions, such as recognition and prestige, salary differentials, job stability, pensions and benefits, professional growth, adequate infrastructure and teaching materials, active working environment, school-based management, among others (Vegas, 2005; OECD, 2009). An empirical analysis based on production functions or correlational analysis shows that their effectiveness is highly mixed, with programs producing a spectrum from weak positive to no impact.

In this paper we are using Impact Evaluation (IE) studies employing Randomized Control Trials (RCT) by the International Initiative for Impact Evaluation, 3ie. The methodology is briefly presented in section II. The empirical evidence will be described in section III. The last sections will discuss the lessons learned and conclusions drawn.

2. Methodology

This review examines the international literature using IE studies to see what lessons can be learned from robust empirical research in developing countries, focusing specifically on a key education input: teacher incentives as a policy to motivate teachers and improve the quality of education in middle- and low-income developing countries. Lately IE studies of Randomized Control Trials (RCT) and a variety of statistical algorithms, including Propensity Score Matching (PSM), difference-in-differences, instrumental variables, regression discontinuity, and others (Gertler et al., 2016), have been used to establish attribution in the relationship between programs improving teachers' skills/behavior and student outcomes (cognitive and non-cognitive ones).

RCT is a trial in which subjects are randomly assigned to one of two groups: one (the experimental group) receiving the intervention that is being tested, and the other (the comparison/control group) receiving an alternative (conventional) treatment. The two groups are then followed up to see if there are any differences between them in the outcome. RCTs are the most stringent way of determining whether a cause-effect relation exists between the intervention and the outcome (Sibbald, 1998). To know how effective a program is, it needs to have a comparison group. Without comparison, it would limit the understanding what would have happened without the program. The only way to have an equitable comparison group is random assignment. Thus, this review uses randomized evaluations to measure impact, because they provide the most credible and reliable way to learn what works and what does not.

In this way, it helps to identify strong empirical evidence of successful education programs in increasing incentives for teachers to do their jobs better and in impacting students' learning outcomes. Considering policy impact at scale, the selection of the evidence is based on the rigor of analysis as well as the potential for the use of the evidence for policymaking. The main source of information used

in this study is found in the work done by the International Initiative for Impact Evaluation, 3ie, the specific repository that classified several thousands of worldwide IE studies. According to 3ie, the examination of the rigor should validate attribution through confounding factors, selection bias, spillover effects, contamination of the control, and impact heterogeneity, etc., so as to gain solid evidence. To present the cases to the reader, we use an edited version of the abstracts for the selected studies as published in the 3ie Development Evidence Portal. Aside from the international development studies in English and Spanish, a search was conducted to the database of research literature in Chinese, including CNKI, CQVIP, Wanfang Data, with the keywords "impact evaluation" and "teacher incentive". Not one single study met the criteria to be included in this study. Thus, the results mainly show available IE studies on the subject 3ie selected that follow RCT and used a rigorous methodology in their analysis. To complement the analysis, and for comparison purposes, other reviews are also examined (Guerrero et al., 2012; Aslam, et al, 2016; Popova, et al., 2019). Summaries of the IE studies included in the paper are presented in the Annex.

The study focuses on a set of policies among the most relevant in increasing teachers' effectiveness in education systems according to the global studies (OECD, 2009; UNESCO, 2015; World Bank). They look at motivating teachers to perform through incentives like rewards not only based on seniority but on effectiveness.

3. Empirical Evidence

The study presents the findings of the motivation to perform policies that have been evaluated using IE studies. The theory of change is straightforward: programs designed to motivate teachers to perform are interventions that may influence teachers' behavior in the classroom, and through them, students' learning outcomes. The interventions may include monetary and non-monetary incentives, types of recruitment, teacher-deployment, etc. Among intermediate outcomes, there are indicators of teaching quality, including measures such as teacher credentials, effort, time on task, absenteeism, etc. And students' cognitive and non-cognitive learning is the indicator of the final outcome. Some of these indicators are not ideal, but in the absence of any more effective measures, they are among the ones used often in the research on teacher effectiveness.

In many countries around the world, authorities have introduced policies on systemic reforms in pedagogy, education governance, and teacher incentives, to increase the effectiveness of investment for education improvement. Some have argued that performance-related payments for teachers are a good way to increase effectiveness. The policy has supporters who argue that salaries should not be defined by seniority, education attainment and training but by education achievement. Others, however, argue that education achievement is too complex to attribute to teachers' performance. Questions remain about the real impact of this policy dimension on education quality. The study includes one IE study on certification in Indonesia, which is indirectly related to the incentive policy. This is important because education systems around the world spend billions of dollars annually on motivation to help teachers meet the diverse needs of their students with limited results in low- and middle-income developing countries. Identifying the more effective way may help improve the efficient use of

resources to improve the quality of education systems. Below the IE evidence is presented by education levels

3.1 Early childhood development and pre-school

Only one IE study is included using an RCT design in the Karamoja Region in Uganda. Gilligan, and Shalini (2016) used UNICEF-supported early childhood development (ECD) centers to examine the cognitive impacts of cash through a stimulation pathway (play, mental engagement, education) that lasted one year and found that cash improved preschool teachers' capacity and resulted in higher preschool attendance, children experiencing higher quantity and quality of exposure to stimulation. Incidentally, they found that food had no significant impact on these outcomes.

The study considered which contextual factors might lead to its limited effects in relation to cash and found that cash transfers led to gains in cognitive scores, but not food transfers. Limited impact on non-cognitive scores occurred with either modality. Improved diet was found to have contributed to improvement in cognitive performance. Among those receiving cash transfers, meat or egg and dairy consumption increased significantly and moderate and/or severe anemia decreased (with weak significance). The researchers also found evidence of increased stimulation through increased participation in the ECD centers. Cash transfers also led to significant increases in the number of days the ECD centers were open and the number of days children attended. Further, cash recipient households were found to increase the value of gifts they gave to volunteer ECD leaders.

3.2 Primary and secondary education

We start with a study in Kenya by Duflo and colleagues (Duflo, Dupas, & Kremer, 2015). They examined a program that randomly selected school committees to hire additional teachers on an annual contract renewable conditional on performance. This contract was outside normal Ministry of Education civil-service channels and was at one-quarter of the normal compensation level. Despite a reduction in class size from 82 to 44 on average for students randomly assigned to stay with existing classes, test scores did not increase significantly. In contrast, scores improved for students assigned to be taught by the locally-hired contract teachers. One reason, as observed in many countries implementing similar mechanisms (Jimenez & Sawada, 1998), may be that contract teachers when hired by the community, had low absence rates, while centrally-hired civil-service teachers working in schools where contract teachers were randomly assigned, reduced their effort. Civil-service teachers also captured rents for their families, with approximately 1/3 of contract teacher positions going to relatives of existing teachers. A governance program that empowered parents within school committees reduced both forms of capture. The best contract teachers obtained civil service jobs over time, and they estimate large potential dynamic benefits from supplementing a civil service system with locally-hired contract teachers brought in on a probationary basis and granted tenure conditional on performance (Duflo, Dupas, & Kremer, 2015). Their results suggest that in the presence of weak institutions, increases in resources may be undermined by the behavioral responses of existing providers.

However, local governance offers the potential to translate increased resources into better outcomes. In particular, programs devolving authority to hire teachers on short-term contracts have the potential both to improve test scores in the short term and to improve the quality of the teacher workforce in the long run. In this light, it is worth considering a related study. A large-scale randomized evaluation program examined contract teachers in India (Muralidharan & Sundararaman, 2011), who were not graduates from teacher training colleges and were ineligible to become civil-service teachers. After 2 years of implementation of the program, the evaluation showed that students in incentive schools performed significantly better than those in control schools by 0.27 and 0.17 SDs in math and language tests, respectively. The program was highly cost-effective, and incentive schools performed significantly better than other randomly chosen schools that received additional schooling inputs of a similar value. In addition, the results further suggest that the gains are broadly distributed among all students, but those children in the more remote areas do appear to receive more benefits (Muralidharan & Sundararaman, 2013). Contract teachers were also much less likely to be absent from school than civil-service teachers (18% vs. 27%). Combining the experimental reduction in the school-level pupil-teacher ratio (PTR) induced by the provision of an extra contract teacher, with high-quality panel data estimates of the impact of reducing PTR with a regular civil-service teacher, they showed that contract teachers were not only effective at improving student learning outcomes, but that they were no less effective at doing so than regular civil-service teachers who were more qualified, better trained, and paid five times higher salaries.

Up to here, an improvement in test scores is observed in two places with different cultural and institutional contexts. Another study to compare is one in India where Duflo, Hanna and Ryan (2012) used a randomized experiment and a structural model to test whether monitoring and financial incentives can reduce teacher absence and increase learning in India. In treatment schools, teachers' attendance was monitored daily using cameras, and their salaries were made a nonlinear function of attendance. The analysis showed that teacher absenteeism in the treatment group fell by 21 percentage points relative to the control group, and the children's test scores increased by 0.17 SDs. The estimated structural dynamic labor supply model revealed that teachers responded strongly to financial incentives. The model is used to compute cost-minimizing compensation policies.

In Pakistan, Barrera-Osorio and Raju (2017) provided evidence from the first three years of a randomized controlled trial of a government-administered pilot teacher performance pay program. The program offered yearly cash bonuses to teachers in a sample of 600 public primary schools with the lowest mean student exam scores in the province. The bonus was linked to the change in the school's average student exam scores, the change in the school's enrollment, and the level of student exam participation in the school. Bonus receipts and sizes were randomly assigned across schools according to whether or not the teacher was the school's head. The program increased student exam participation rates in the second and third years and increased enrollment in grade 1 in the third year. They did not find that the program increased student exam scores in any year. Mean impacts were similar across program variants. The authors argued that the absence of positive impacts on test scores may be due to weaknesses in the program's incentive structure and/or limitations in the program's administrative data.

Back to Africa, there is an IE study of the Gambian hardship allowance (Pugatch & Schroeder, 2014), which provided a salary premium of 30-40% to primary school teachers in remote locations,

based on the distribution and characteristics of teachers across schools. A geographic discontinuity in the policy's implementation and the presence of common pre-treatment trends between hardship and non-hardship schools provided sources of identifying variation. They found that the hardship allowance increased the share of qualified (certified) teachers by 10 percentage points. The policy also reduced the pupil-qualified teacher ratio by 27, or 61% of the mean, in recipient schools close to the distance threshold. Further analysis suggests that these gains were not merely the result of teachers switching from non-hardship to hardship schools. With similar policies in place in more than two dozen other developing countries, the study provides evidence of their effectiveness.

In a very different context, Contreras and Rau (2012) evaluated the effects of the introduction of group monetary incentives for teachers, based on a school performance tournament in Chile. This is particularly relevant since it is the only scaled-up incentive program for teachers in the world. They evaluated the tournament effect, that is, the effect of introducing the incentive scheme on all participant schools, both winning and losing. They compared public and private subsidized schools to private fee-paying schools following a matched difference-in-difference and using three different empirical approaches. They finally explored the heterogeneous impact of the treatment through the distribution of the probability of winning. The results indicate a positive and significant tournament effect, especially for schools that are very likely to win.

Similarly, in Mexico, Behrman and colleagues (Behrman, et. al., 2012) did an evaluation of the impact of three different performance incentive schemes using data from a social experiment that randomized 88 Mexican high schools with over 40,000 students into three treatment groups and a control group. They followed treatments providing individual incentives for performance on curriculum-based mathematics tests to students only, providing individual incentives to teachers only, and the final treatment giving both individual and group incentives to students, teachers and school administrators. Their study of the impact estimates revealed the largest average effects for treatment three, smaller impacts for treatment providing incentives to students only, and no impact for the treatment that provides incentives only to teachers.

Continuing with Latin America, Brazilians decided to implement a program, after bonus payment schemes to teachers had been implemented overseas as a tool for education quality improvement. The program evaluated the effect of a bonus program for teachers in the state of Sao Paulo, over the proficiency of students (Oshiro et al., 2015). Propensity Score Matching and Difference-in-Differences methods were used to measure the impact of the program. The study found positive and significant effects on math tests (0,42 SDs) and reading (Portuguese) tests (0,14 SDs) for the fourth grade between 2007 and 2009, but there was a sharp decay in these effects in 2011. They didn't find any effects for the eighth grade of elementary school and, in some specifications and control groups, the results were negative in 2009 and 2011. These results are robust to different control groups.

For certification, there was a large program in Indonesia aimed to certify all teachers by 2015. The program was rolled out at a rate of approximately 200,000 teachers each year. An evaluation of the certification program was implemented, in terms of its impact on student learning outcomes (World Bank, 2016). The results of the analysis were discouraging: despite its massive fiscal implications, the certification program did not result in improvement in student learning outcomes. The report provides

clues into how to gradually transform the system into one that can yield higher returns in educational performance going forward. It emphasizes the importance of a system that rewards useful demonstrated competencies, such as minimum levels of subject-matter knowledge, rather than loose proxies of quality such as bachelor's degrees or seniority alone (which is essentially what the evaluated certification program does). The report also highlights the need for reforms in the pre-service system of teacher training and teacher hiring; it presents policy options for sizable and lasting changes in education quality.

The study shows that to improve the quality of the education system, it is required that they develop reforms explicitly geared towards quality. These eventual reforms may work best, if steps in teachers' career progressions, or professional (re)certification, linked to salary incentives, are tied directly to their ability to demonstrate useful competencies, one of such competencies being minimal levels of subject matter proficiency (World Bank, 2016).

4. Discussion

This section addresses findings that may be of interest to policymakers and practitioners about policies to motivate teachers to perform by changing incentives. Unlike qualitative findings in interpreting the causes, the rigorous impact evaluations tackle the attribution, and address the replicability of the changes and sustainability of impacts at scale. A word of caution is needed regarding the limited number of IE studies, although there has been an increasing recognition of the value of impact evaluation in informing evidence-based policy decisions since the 1980s. This does not allow us to generalize the policy for any countries to apply them, however the findings point out lessons countries should look at when designing their own education sector developing programs.

Although most IE studies show that targeted cash incentives have a positive effect in increasing learning outcomes, including when incentives go to teachers and students like in Mexico, results are mixed with some cases finding an impact in some years or levels of education but not in others. Only one study, in Pakistan, did not improve test results but it improved other intermediate outcomes. Monetary incentives also positively improve the allocation of teachers in vulnerable schools.

Incentives and time on task. Contract teachers, part of a second set of programs to motivate teachers to perform, also seems to impact the allocation of teachers making sure that quality teachers go to schools in isolated low-income areas. Some studies also have demonstrated a positive effect on learning, apparently by increasing time on task and increasing teachers' attendance. In fact, according to IIEP, teacher absenteeism is endemic in developing countries and it is not uncommon to find that a significant percentage of the teachers in a given school, mainly in low-income and isolated areas, may be absent or if present at the school, out of the classroom. Time on task is critical for learning so the idea that teacher attendance is likely to be critical to achieve student learning goals has received significant attention from policymakers and education system stakeholders. The review of the available robust IE studies on incentives showed that they help reduce absenteeism. However the topic itself is of most importance, so it is presented in a summary of a related study conducted about a decade ago (Guerrero et al., 2012). They produced a review of evaluations that dealt with teacher absenteeism and the main conclusions are in Box 1 below

Box 1. Bridge between Incentives and Absenteeism

The study by Guerrero and colleagues identified the evidence on the effectiveness of interventions intended to increase teacher attendance in developing countries (using both quasi-experimental and experimental design and it encompasses nine studies, including seven RCTs, one quasi-experiment using matching methods and one study using the Heckman two-step method. Four studies were from Africa (one in Madagascar and three in Kenya), three from Asia (all India) and two from Latin America (Peru and El Salvador) and concluded that a combination of better monitoring and powerful incentives are effective in tackling teacher absenteeism. However, having a teacher in the classroom does not appear to be sufficient to improve student achievement. All these are corroborated by more recent studies.

Although improving attendance was not straightforward, results provided evidence that a combination of better monitoring and powerful incentives seems effective in tackling teacher absenteeism. In that sense, it was interesting to note that the four interventions that succeeded in improving teacher attendance had some form of monitoring.

Most interventions aimed to increase teachers' attendance by offering monetary incentives; only one tried to alter their workload to increase their satisfaction and eventually reduce absenteeism. However, factors such as improving the work environment or promoting professional development could also be considered as tools to improve teacher absenteeism. At the school and educational system level, there was an emphasis on creating and strengthening monitoring systems, but the impact of other relevant variables, such as group norms, school principal leadership and teacher administrative duties, could be tested in rigorous impact evaluations.

Overall, it would seem that having a teacher in the classroom is an important but insufficient prerequisite for improving achievement. The quality of the pedagogical processes within the classroom also needs to be considered (included studies did not provide information on what happened during the increased time of attendance).

It highlights the need for more studies of teacher-attendance interventions and their effects on teacher absenteeism and student learning. It also calls for research to address questions of how, where and why teacher-incentive programs succeed or fail in increasing attendance and improving student achievement and how the extra time that teachers spend in classrooms is used. Future research should therefore also focus on interventions to improve pedagogical activities.

Teacher certification. This has been a policy of choice in many countries around the world. Mandatory minimum education levels are established, with the idea of implementing minimum quality standards. It assumes that recognizing teachers who meet those requirements would improve the status of the teaching profession, and in consequence, the caliber of new entrants into the teaching labor market will increase as well. The studies show that results are mixed. The case of Indonesia shows no effect of teacher certification on teacher quality and this intervention shows no effect on student achievement. The study found that because the certification was based on factors other than merit, any potential positive impacts were muted. The authors also highlighted that policies dealing with improving teachers' salaries must be based on proven competencies required of an effective teacher in order for these policies to demonstrate results (World Bank, 2016). Previous cases where governments have increased salaries without conditions linked to effective teaching have resulted in wastage (Vegas, 2005).

Interventions involving the time for teaching and learning have a significant effect on student performance as well. Conn (2014) found that in the context of Sub-Saharan Africa, the greater length of time in class resulted in better test scores (0.412 SDs) in mathematics. Our sample of studies shows using incentives to increase student learning motivation has a greater effect on learning outcomes than incentivizing teachers (0.288 SDs versus 0.075). School or system accountability interventions that increase school-level and system-level accountability, such as providing information on school performance or funding to the community, have a large effect on student performance (0.147 SDs).

In a world of rapid change, education systems must ensure that they provide teachers with the needed skills to prepare students for jobs that might not yet exist, to use technologies that have not yet been invented, and to solve problems that are still unknown. In this context, learning and development outcomes matter (OECD, 2018). Many of education's benefits depend on the skills that students develop in school. As workers, people need cognitive and socioemotional skills to be productive and innovative. As parents, people need literacy to read to their children or to interpret medication labels, and numeracy to budget for their futures. As citizens, people need literacy and numeracy, as well as higher-order reasoning abilities, to evaluate politicians' promises. As community members, they need the sense of agency that comes from developing mastery (World Bank, 2018). Although the empirical literature on the impacts of education has focused much more on schooling, specifically years of formal education, than on quality, specifically learning, mounting evidence supports this intuition. Even after controlling for schooling, empirical studies find that skills in the adult population affect outcomes, including social mobility, health, and earnings among others (World Bank, 2018).

In this context, how best a teacher can do is a challenge that we have to face when designing education programs based on the lessons learned from empirical evidence. The evidence presented in this review is descriptive and the review is not designed to provide specific advice on which interventions are more or less appropriate in particular contexts, but rather, to summarize what is known in response to the main topic on teacher incentives and quality of education.

5. Conclusion

There are no silver bullets to solve all problems, and context is critical to designing pertinent education programs. Given the decades of investment in different policies without proper empirical analysis, policymakers and educators should support innovations that have proven to be effective.

Based on the findings, strengthening the teaching workforce and improving the quality of the teachers will require improving the classroom performance of individual teachers; this needs to be a specific objective of the incentive programs. This review shows that teacher incentive programs have some gains in student learning. Conditional Cash Transfers may increase enrollment but have small effects on outcomes. Direct community hiring improves teachers' capacity but has limited impact on student achievement. Monitoring and incentives reduce absenteeism, improving results. Hardship allowance improves the presence of qualified teachers. Incentives to different stakeholders (teachers, students and administrators) are more effective than incentives to teachers alone. Finally, regarding certification, it is recommended that for countries to improve the quality of the education system, it is critical to develop reforms explicitly geared towards quality. Certification needs to be linked to the teachers' ability to demonstrate useful competencies, including subject matter proficiency.

When looking at the available empirical literature, the number of IE studies using experimental design is very limited even in developed countries. A recent meta-analysis of 196 evaluations of education interventions found just half a dozen on incentives in developed and developing countries dealing with incentives to parents, students or teachers (Fryer, 2017).

In the case of teacher incentive policies, it is worth doing new studies covering intervention areas that seem promising. Future studies should use mixed method designs to examine the effects of interventions as well as process, implementation and contextual factors that influence final outcomes to help explain heterogeneity and inform future programming. This coincides with the recommendation by Aslam and colleagues (Aslam et. al., 2016), that most IE studies remain context-specific. Like their findings on teachers' effectiveness, the evidence covered in this review does not give us sufficiently detailed information to uncover whether it is specific design features and/or contextual factors (and how much of each) that have driven the causality observed in the studies. Given the variation in the design of reforms and contexts, future research can consider both in detail, so that it will be possible to disentangle the independent and joint effects of each.

Given the limited number of studies, it is clear that more research on teacher absenteeism and its effect on student outcomes is needed. Establishing how, where and why teacher incentive programs succeed or fail in increasing attendance and improving student achievement remains an important priority. Hence, future research needs to focus on interventions and studies that would not only increase teachers' presence but would also help to use the time in pedagogical activities so that student achievement is indeed improved.

Declarations and Acknowledgement:

The authors declare no conflict of interest.

This paper was financed by the Faculty of Education, Beijing Normal University. The views expressed here are those of the authors and should not be attributed to BNU.

The authors would like to thank the editors and reviewers for their insightful comments and suggestions, which have greatly improved the quality of this manuscript.

References

Aslam, M., Rawal, S., Kingdon, G., Moon, B., Banerji, R., Das, S., Banerji, M., & Sharma, S.K. (2016). Reforms to Increase Teacher Effectiveness in Developing Countries. London: EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London.

Barber, M., & Mourshed, M. (2007). How the World's Best-Performing School Systems Come Out on Top. London, United Kingdom: McKinsey.

Barrera-Osorio, F., & Raju, D. (2017). Teacher Performance Pay: Experimental Evidence from Pakistan. *Journal of Public Economics* Vol. 148 pp:75-91.

Bau, N., & Das, J. (2017). The Misallocation of Pay and Productivity in the Public Sector: Evidence from the Labor Market for Teachers. Policy Research Working Paper, 8050. World Bank, Washington, D.C.

Behrman, J.R., Parker, S.W., Todd, P.E., & Wolpin, K.I. (2012). Aligning Learning Incentives of Students and Teachers: Results from a Social Experiment in Mexican High Schools. PIER Working Paper 13-004.

Buhl-Wiggers, J., Kerwin, J. T., Smith, J. A., & Thorton, R. (2017). The impact of teacher effectiveness on student learning in Africa. Research on Improving Systems of Education (RISE) Working Paper.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review* 104(9): 2633-2679.

Conn, K.M. (2014). Identifying Effective Education Interventions in Sub-Saharan Africa: A meta-analysis of rigorous impact evaluations. Columbia University Theses Doctoral.

Contreras, D., & Rau, T. (2012). Tournament Incentives for Teachers: Evidence for a Scaled-UP Intervention in Chile. *Economic Development and Cultural Change* 61 (1): 219-246.

Darling-Hammond, L., Hyler, M. E., Gardner, M. (2017). *Effective Teacher Professional Development*. Palo Alto, CA: Learning Policy Institute. https://doi.org/10.54300/122.311.

Duflo, E., Dupas, P., & Kremer, M. (2015). School Governance, Teacher Incentives, and Pupil-Teacher Ratios: Experimental Evidence from Kenyan Primary Schools. *Journal of Public Economics* 123: 92-110.

Duflo, E., Hanna, R. & Ryan, S.P. (2012). Incentives Work: Getting Teachers to Come to School. *American Economic Review* 102 (4): 1241-78.

Evans, D.K., & Yuan, F. (2017). Economic returns to interventions that increase learning. Background paper, World Development Report 2018, World Bank, Washington D.C.

Fryer, Jr, R. G. (2017). The production of human capital in developed countries: Evidence from 196 randomized field experiments. In *Handbook of Economic Field Experiments* 2: 95-322.

Gertler, P.J., Martinez, S., Premand, P., Rawlings, L.B., & Vermeersch, C. M. J. (2016). *Impact Evaluation in Practice*, second edition. Washington, DC: Inter-American Development Bank and World Bank.

Gilligan, D.O., & Shalini, R. (2016). The Effect of Transfers and Preschool on Children's Cognitive Development in Uganda. 3ie Series Report 32.

Glewwe, P., & M. Kremer. (2006). School, teachers, and education outcomes in developing countries. In: Hanushek E.A, & Welch, F. (Eds.), *Handbook of the Economics of Education*. Vol. 2. Elsevier B.V, Amsterdam

Guerrero, G., Leon, J., Zapata, M., Sugimaru, C., & Cueto, S. (2012). What works to improve teacher attendance in developing countries? A systematic review. London: EPPI- Centre, Social Science Research Unit, Institute of Education, University of London.

Hanushek, E.A., & S.G. Rivkin. (2006). Teacher quality. In Hanushek, E.A., and F. Welch (Eds.), *Handbook of the Economics of Education*. Vol 2. Elsevier B.V., Amsterdam

Lee, S., & Koh. A. (2020). Lessons for Latin America from Comparative Education: South Korea's Teacher Policy. Department of Research and Chief Economist. Discussion Paper IDB-DP-No.744. Inter-American Development Bank.

Muralidharan, K., & Sundararaman, V. (2011). Teacher Performance Pay: Experimental Evidence from India. *Journal of Political Economy* 119(1), 39-77.

Muralidharan, K. & Sundaraman, V. (2013). Contract Teachers: Experimental Evidence from India, National Bureau of Economic Research Working Paper Series, NBER Working Paper No.19440.

OECD. (2009). *Evaluating and Rewarding the Quality of Teachers. International practices*. OECD Publishing, Paris.

OECD. (2018). The future of Education and Skills. Education 2030. OECD Publishing, Paris.

Oshiro, C.H., Scorzafave, L.G., & Dorigan, T.A. (2015). Impacto sobre o Desempenho Escolar do Pagamento de Bônus aos Docentes do Ensino Fundamental do Estado de São Paulo. *Revista Brasileira de Economia* 29(2).

Pugatch, T., & Schroeder, E. (2014). Incentives for Teacher Relocation: Evidence from the Gambian Hardship Allowance. *Economics of Education Review* 41, 120-136.

Popova, A., Evans, D.K., Breeding, M.E., & Arancibia, V. (2019). Teacher Professional Development around the World: The Gap between Evidence and Practice. Center for Global Development. Working Paper 517, Washington D.C: Center for Global Development.

Sibbald, R.M. (1998). Understanding controlled trials: Why are randomised controlled trials important? *BMJ*, 316:201

Snilstveit, B., Stevenson, J., Phillips, D., Vojtkova, M., Gallagher, E., Schmidt, T., Jobse, H., Geelen, M., Pastorello, M., & Eyers, J. 2015. Interventions for Improving Learning Outcomes and Access to Education in Low- and Middle Income Countries: A Systematic Review. 3ie Systematic Review 24.

IIEP. Teacher Absenteeism. Teacher absenteeism | Education | IIEP Policy Toolbox (unesco.org).

UNESCO. (2015). Teacher Policy Development Guide. Paris: UNESCO.

Vegas, E. (2005). *Incentives to Improve Teaching: Lessons from Latin America*. Washington, DC: World Bank.

World Bank. It All Start with a Good Teacher. http://documents1.worldbank.org/curated/en/555631541081210478/pdf/131639-BRI-teachers-PUBLIC-Series-World-Bank-Education-Overview.pdf

World Bank. (2018). World Development Report 2018: Learning to Realize Education's Promise. Washington, DC: Oxford University Press and The World Bank.

World Bank. (2016). Indonesia Teacher Certification and Beyond: An Empirical Evaluation of the Teacher Certification Program and Education Quality Improvements in Indonesia. Jakarta: World Bank.

About the Authors:

Dr. CHI Jin is an Associate Professor in the Institute of Teacher Education Research, Faculty of Education at Beijing Normal University. Her work focuses on quality learning, teacher policy in education systems, child development, and rural women and girls' education and development. Prior to this role, she led the Gender Equality and Women Leadership program at UNESCO INRULED. She has committed to international services to promote education and development locally and globally. Dr. Chi has been an Echidna Global Scholar at Brookings Institution and a Visiting Scholar at Georgetown University. Dr. Chi holds a Ph.D. from the Faculty of Education at The University of Hong Kong.

Eduardo Vélez, PhD (Sociology, University of Illinois). His main interest is the area of Economics of Education. Former World Bank Education Sector Manager for Latin America and the Caribbean and for East Asia and the Pacific; Human Development sector leader for China, Mexico, and Colombia; and Principal Education Specialist for Eastern and Southern Africa. He was an Adjunct Professor at Georgetown University and at Peking University, and has been a Visiting Professor at Kobe University, Brown University, the University of Connecticut and at the Universidad de los Andes in Bogotá, Colombia. He has published eleven books and near 100 articles in professional journals and research reports.

Annex. IE Studies -- Motivate Teacher to Perform

Authors	Country	Outcomes	Results
Gilligan and Shalini, 2016	Uganda	Cognitive and non-cognitive	Cash transfers increase pre-school attendance, and student non-Cognitive (stimulation)
Duflo, Dupas and Kremer, 2015	Kenya	Student and teacher achievement	Improve teacher's capacity after direct community hiring and teacher tenure based on performance. Mixed student achievement
Muralidharan and Sundararaman, 2011 Muralidharan and Sundararaman, 2013	India	Student learning	Local community hiring improves student learning and teacher's time on task. Cost-effective
Duflo, Hanna and Ryan, 2012	India	Student learning Teacher absenteeism	Incentives reduce teacher absenteeism and improve student learning. Cost-effective
Barrera-Osorio and Raju, 2017	Pakistan	Student learning	Performance pay programs did not have an effect
Pugatch and Schroeder, 2014	The Gambia	Qualified teacher	Incentive salaries improve the quality of teachers in remote areas
Contreras and Rau, 2012	Chile	Student achievement	Group economic incentives have mixed results. Positive tests in some schools
Berhman, Parker, Todd, and Wolpin, 2012	Mexico	Student achievement	Incentives when to students, teachers and administrators, improve student achievement. No improvement when teachers only or administrators only
Oshiro, Scorzafave, and Dorigan, 2015	Brazil	Student achievement	Bonus payments to teachers produced mixed results. Improve math and language for some grades and some years
World Bank, 2016	Indonesia	Student achievement	Teachers' accreditation had no impact on student quality because it was not linked to quality requirements