Designing and Testing a Scalable Teacher Incentive Programme in Tanzania

by Youdi Schipper, Isaac Mbiti (both RISE Tanzania Country Research Team), and Mauricio Romero (RISE India Country Research Team)

Introduction

School participation in Tanzania has increased dramatically over the past two decades: primary school enrolment increased from 4.9 million in 2001 to 10.9 million in 2020. While 81 percent of primary-school-age children are currently enrolled, over the last ten years, the primary completion rate has dropped and remains below 70 percent since 2015 (data from UNESCO Institute for Statistics).[1]

Despite improvements in enrolment, indicators of foundational learning remain low. According to the 2020 report of the Standard Two National Assessment (STNA), conducted by the National Examinations Council of Tanzania (NECTA), in 2019 five percent of Grade 2 students pass the benchmark for reading proficiency ("Can correctly read exactly 50 words of the passage in one minute and with 80 percent or higher comprehension"). The report finds that 17 percent of students pass the benchmark (80 percent correct) of the addition and subtraction sub-tasks.

These outcomes are not the result of students' lack of academic aspiration: according to the RISE Tanzania baseline survey, 73 percent of Grade 2 and 3 students say they would like to complete secondary school or university (Figure 1).
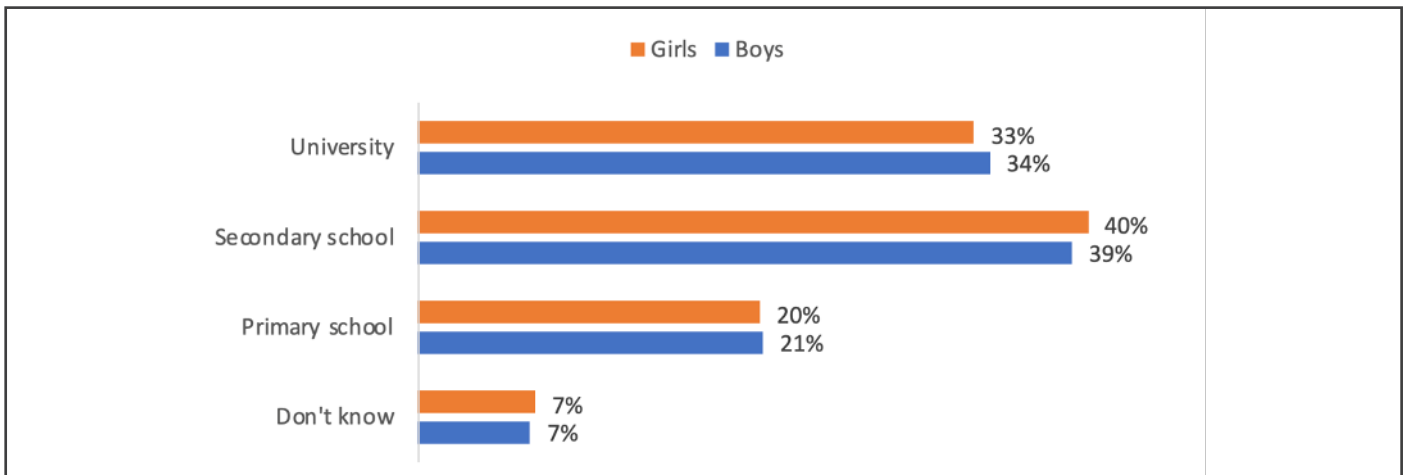
In a recent report, the Global Education Evidence Advisory Panel (World Bank, 2020) asked what programmes and policies are the most cost-effective instruments for addressing the learning crisis and improving learning for all children. The report creates three categories: the "great buys" category includes programmes that provide very low-cost but salient information on the benefits, costs, and quality of education. The "good buys" category includes programmes

Key Points

- Five percent of primary Grade 2 students in Tanzania pass the oral reading fluency benchmark on the Standard Two National Assessment and 17 percent pass the benchmark for addition and subtraction sub-tasks (NECTA, 2020).

- Low teacher morale and lack of independent assessments in the early grades help explain these low foundational pass rates. A large majority of teachers prefer teaching in upper rather than lower primary grades. Teachers in Grades 2 and 3 over estimate the foundational reading and numeracy skills of their students by more than 100 percent, when compared with results from an independent assessment.

- We describe results of KiuFunza 3, a teacher performance pay programme that offers individual cash rewards for core foundational skills to teachers in lower primary. The programme uses a simple incentive design based on results from an earlier study. To make the programme easier to scale, we further simplified implementation and sharply reduced the costs of high-stakes student assessments. Despite the lower cost, we find the incentives cause substantial improvements in foundational reading and numeracy.

- Despite potential concerns that bonus pay may lead teachers to focus on well-performing students, we do not find systematic differences in the treatment effects by baseline scores in any year.

Figure 1: What level of formal education would you like to complete?

Source: RISE Tanzania Baseline Survey, 2019.

that provide structured pedagogy, instruction targeted by learning level, merit-based scholarships and pre-school interventions. Finally, the category "promising but low-evidence" includes teacher accountability and incentive reforms.

KiuFunza, a teacher performance pay programme in Tanzania, fits this last category. KiuFunza (shorthand for Kiu ya Kujifunza or Thirst to Learn) provides test-score linked cash incentives to teachers in Grades 1, 2, and 3 to increase foundational literacy and numeracy outcomes for students. The programme is managed by Twaweza East Africa, a Civil Society Organization, and was set up to provide evidence on the impact of teacher incentives in a series of experimental evaluations. This note discusses the rationale for teacher incentives in Tanzania, the design elements of KiuFunza and preliminary results for the most recent phase of KiuFunza (this phase was implemented in 2019-2021 and the impact evaluation is part of the RISE Tanzania research agenda).
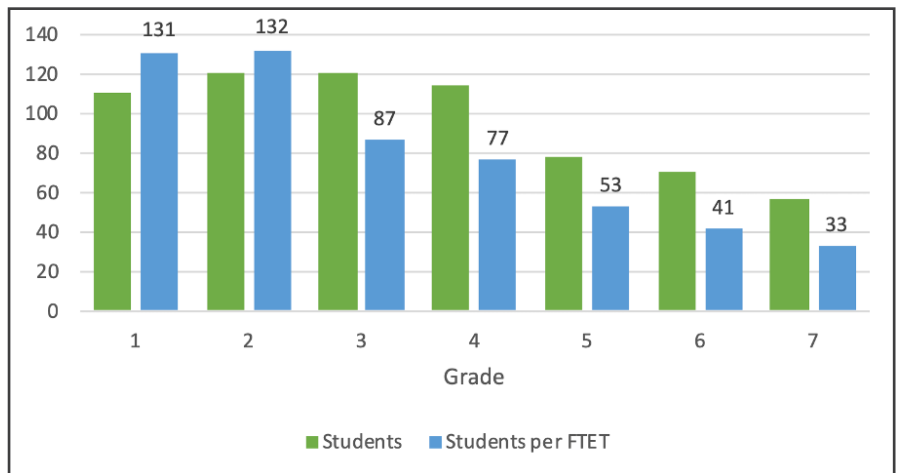
Rationale

For Twaweza, the idea to start a teacher incentives programme was based on three sets of evidence:

1. Evidence of the learning crisis that presents a lack of accountability to citizens for the large public expenditure in education, mainly on teacher salaries (World Bank, 2018; Bold et al., 2017; Bashir et al., 2018).

2. Evidence that Twaweza's Uwezo initiative, which performed large-scale assessments and informed communities about low learning levels, did not lead to community-based citizen engagement and improved learning at schools, although it did contribute to a focus on learning at national policy levels (see Lieberman et al., 2014; a similar result for India is found by Banerjee et al., 2010).

3. Evidence of the potential cost-effectiveness of performance incentives in primary education, and potential complementarity with existing large investments in education (see Glewwe and Muralidharan, 2016; Evans and Mendez Acosta, 2021).

The KiuFunza programme theory of change, supported by the experimental results, is that performance incentives linked to learning—implemented with good communication, fidelity and independent testing—will improve mastery of the basic skills tested. This theory of change is consistent with research in the economics of management (e.g., Bloom and co-authors [2015, 2014]). These studies show that effective organisations are successful in setting clear goals; in measuring progress towards those goals; and instituting employee level accountability and consequences for measured progress. This research includes management of schools: Bloom et al. (2015) find that across schools in different countries, having strong school accountability for student performance to an outside body predicts better pupil outcomes. KiuFunza sets basic literacy and numeracy skill targets, organises independent student performance assessments, and rewards progress towards these targets. The aim is to increase school and teacher level accountability and ownership with regard to learning results.

The programme targets teachers in Grades 1, 2 and 3 for two reasons. First, foundational skills instruction takes a central place in these grades; between 2010 and 2016, Tanzania introduced a simplified early grade curriculum focusing on the "3Rs" in Grades 1 and 2 (Komba and Shukia, 2021). Students who do not acquire foundational literacy by Grade 3 will struggle to pick up higher order skills in upper primary. Second, class sizes in these grades are large, both in an absolute sense and relative to upper primary. Figure 2 illustrates this by showing

Figure 2: Students per grade

Source: RISE Tanzania Baseline survey, 2019.

two data points for each grade: the first is the average number of students per grade enrolled; the second is students divided by the full-time equivalent teacher (FTET) capacity per Grade.[2] The latter is a grade-specific student-teacher ratio, adjusted for the fact that teachers assigned to more than one grade should not be counted as a full-time teacher in each of those grades.
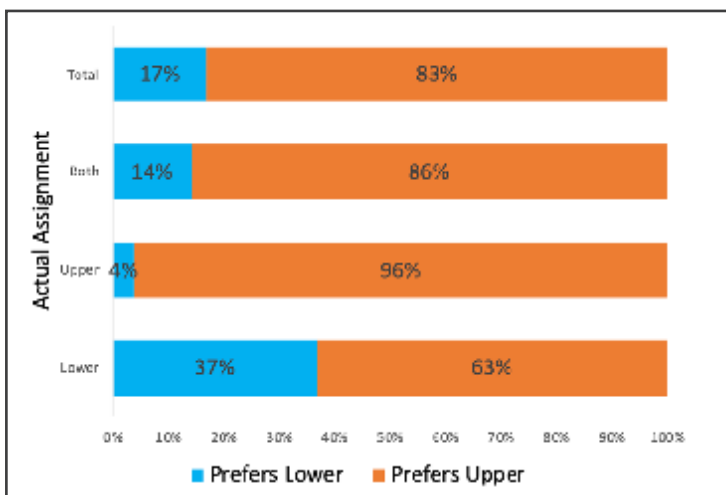
Note that Figure 2 provides averages of enrolled students and does not account for student absence. However, even if 20 percent of students are absent on average these numbers create a challenging workplace environment for teachers in the early grades. Large class sizes are mentioned by over half of lower grade teachers when asked about challenges to foundational instruction in the RISE baseline survey. Consistent with these findings, we find a substantial imbalance in terms of motivation: Figure 3 shows that 83 percent of all primary teachers prefer to only teach in Grades 4-7 (RISE baseline survey data, 2019), where student-teacher ratios are lower.

Figure 3: Teachers prefer upper primary (percentage of teachers, by current assignment)
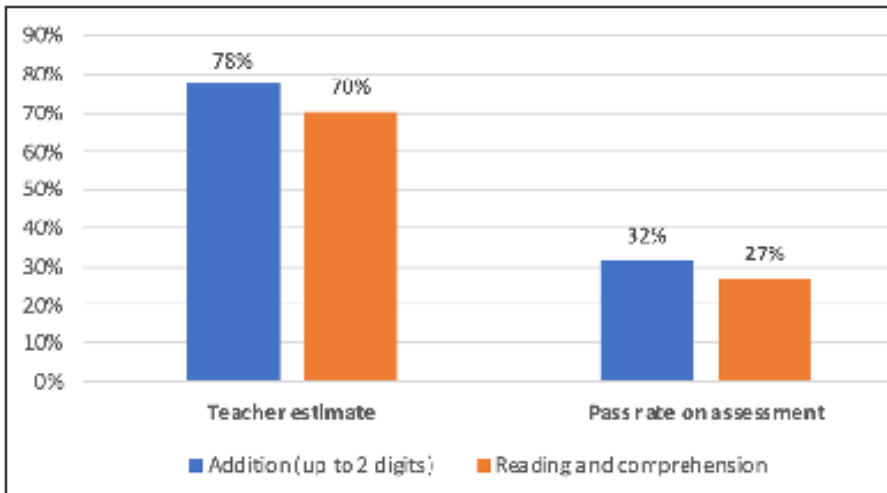

Source: RISE Tanzania Baseline survey, 2019.

In addition, the first three grades in Tanzania lack a national monitoring system for learning outcomes; such a system could provide a common understanding and acceptance of outcomes. A lack of acknowledgement of learning outcomes is illustrated by results from the RISE Tanzania baseline survey, where teachers in Grades 1, 2 and 3 were asked "What is the approximate share of pupils in your class that can read Kiswahili at Grade 2 level (for example a short story of five sentences); and answer comprehension questions?". A similar question was asked for Grade 2 level addition: "What is the approximate share of pupils in your class that can do Grade 2 level addition, for example 26+32?"

On average, teachers of these students estimated that 70 percent could read with comprehension,

---

[2] The FTET accounts for the fact that many teachers need to divide their time over more than one grade: formally, if N is the number of grades that a teacher is assigned to, each grade is assigned 1/N FTE for the teacher. For example, if a teacher is assigned to teach in three grades, each grade is assigned 1/3 FTE for that teacher. If a teacher is assigned to one grade only, that grade is assigned 1 FTE for this teacher. For each grade, we add the assigned FTEs from all teachers to calculate total FTEs for that grade. Teacher assignments are based on the full teacher roster provided by the head teacher.

Figure 4: Student performance (percentage passing)

Source: RISE Tanzania Baseline survey, 2019.

and that 78 percent could do addition (see Figure 4). In the same survey, enumerators administered a low-stakes survey assessment of these skills and found that only 27 percent of students in Grades 2 and 3 could read at Grade 2 level; and only 32 percent of these students could do addition.[3]

Finally, descriptive statistics from research surveys point to substantial problems with school level supervision and accountability. Teacher classroom attendance is not monitored regularly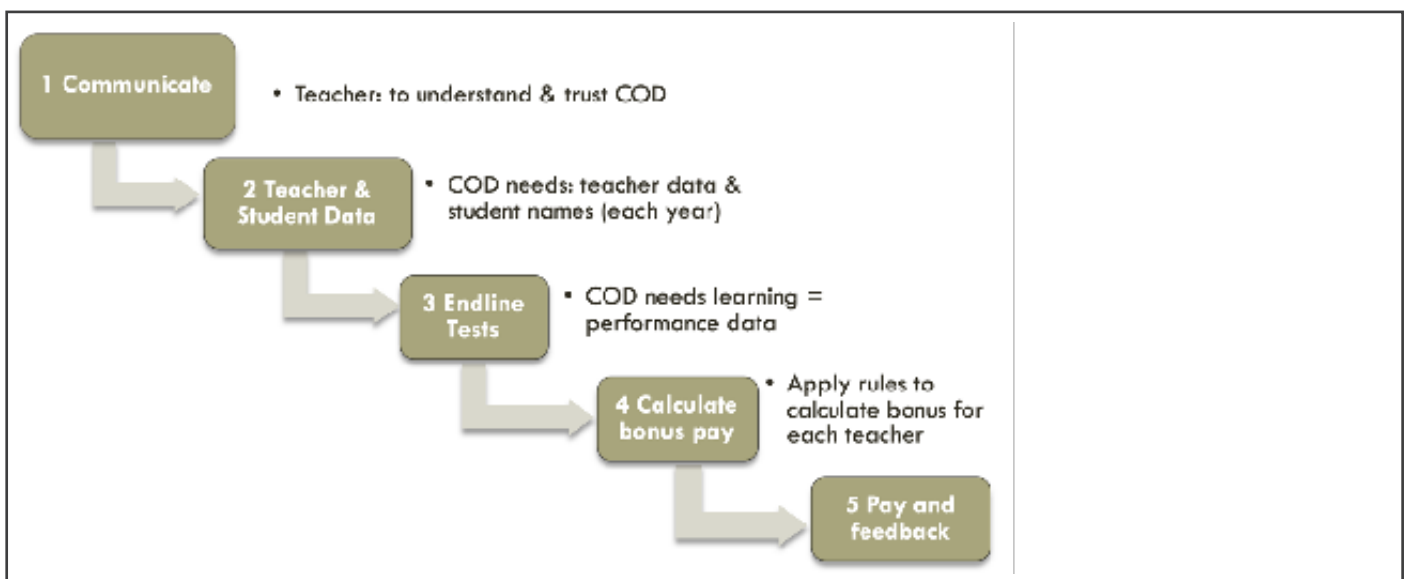, but results from several surveys show that it is below 50 percent (Bold et al., 2017; Mbiti et al., 2019b). Employee supervision is uneven, and often lacking: 41 percent of teachers indicate that a head or senior teacher never observes their classroom (RISE baseline survey).

Intervention design

The findings in the previous section point to the huge challenge of teacher performance improvement. According to Pritchett (2013), teacher management—the "implementation-intensive imposition" of norms and obligations—is among the hardest functions to organise in education, as it is transaction-intensive, discretionary, and high-stakes. The rationale for KiuFunza teacher performance pay is to provide a form of teacher management that focuses teacher effort on foundational learning. To maximise this focus and to keep the implementation complexity to a minimum, the bonus rewards are tied only to measures of foundational learning (rather than, say, a mix of learning, attendance, and lesson plan preparation).

The core intervention design of KiuFunza is simple (see Figure 5): at the start of the school year, programme staff visit a school, communicate the bonus offer, and enlist teachers; at the end of the school year they assess students;

Figure 5: KiuFunza implementation cycle.

---

[3] For the assessment, Grade 2 level was defined as adding up numbers between 1 and 20, without carrying over: 5+1, 11+4, 5+15.

then, based on test scores, payments are calculated; incentives are paid at the start of the next year; schools and administrators receive a report on student performance by grade-subject.

The necessary inputs for a teacher incentive programme are good communication tools; a trustworthy assessment system, with a trained and trusted assessment team; a good data management system; a budget to pay the bonuses and the staff. In addition, since the programme allocates cash bonuses, the programme requires auditable processes and procedures. Systems were developed to maintain integrity, both in the field (during assessments and data collection) and at the management level. At the field level, for example, a key ingredient is randomised selection of test sets. At the data management and payment stage all steps are documented so that data and financial auditors can link raw data to teacher transfers.

Several implementation parameters have remained constant since KiuFunza started in 2013:

- the focus on teachers and students in Primary School grades I-III;

- the focus on foundational literacy and numeracy skills in the high-stakes tests;

- the inclusion of the head teacher in the bonus offer;

- measuring performance at the school-grade-subject-classroom level;[4] and

- the absence of a downside risk for teachers.

Incentive design

The KiuFunza incentive budget is based on an expected teacher bonus equal to about 3.5 percent of annual teacher salary. These mean reward sizes are close to those in some other teacher incentive experiments, including Muralidharan and Sundararaman, 2011 (India); and Behrman et al., 2015 (Mexico). In our experience it is sufficiently high to attract teacher attention, and survey data show that the programme is popular with teachers (Mbiti and Schipper, 2021).

KiuFunza has tested different incentive designs (i.e., mappings between test results and teacher earnings). Mbiti et al. (2019a) describes and evaluates a standard threshold incentive design, where a student needs to pass a set of curriculum skills before the teacher earns a reward. Mbiti et al. (2019b) evaluates two designs that can provide rewards even if students do not pass the full curriculum test: a pay-for-percentile design and a design called "Stadi" (Levels) with payments for individual curriculum skills (e.g., reading a few sentences, subtraction).

The impact evaluations (Mbiti et al., 2019a and 2019b) show that all these designs improved learning based on high-stakes test data (at the time of writing we have only preliminary results for the most recent design, 2019-21). For the low-stakes (survey) test data, only the two designs in the second phase had a statistically significant impact on learning outcomes. Stadi was the most cost-effective of these two and this incentive design was selected for the third phase of KiuFunza (2019-21).

KiuFunza III design adjustments

The third phase of KiuFunza (part of the RISE impact evaluation) featured a number of other, more practical design changes. A first change was made to reduce student assessment costs: with large student populations in the early grades, these form a substantial part of overall implementation costs and thus present an impediment to scaling test-based accountability programmes. The student assessment approach was changed in two ways: in Grades 1 and 2, one-on-one census testing was replaced by one-on-one testing in a random sample of 15 students per grade (the consensus is that written tests are not practical before Grade 3). In Grade 3, one-on-one census testing was replaced by class-based census testing using a written test. Together, these changes lowered testing unit costs by about 60 percent compared with the previous phase.

---

[4] This is equivalent to the teacher personal level in all but very few cases where teachers share teaching responsibility.

Second, an "infrastructure bonus" was added to the incentive design, based on Government advice. The infrastructure bonus targets the bottom quartile of schools, by region, in terms of the Standard Four National Assessment. Third, government officers (from the School Quality Assurance department) were added to CSO implementation teams, to build more implementation capacity.

Fourth, KiuFunza removed the mobile money payment option for teachers, because these were much more likely to bounce back and require follow-up than bank payments. Fifth, the programme started to exploit the widespread ownership of smartphones, primarily through an app that captures test results for Grades 1 and 2 (in earlier phases test results were written on forms, then scanned). The app also has several other useful functions: it generates random numbers that are used for sampling; it randomly selects test sets, a feature that strengthens test integrity; it guides test administrators through the testing process; and it generates time and GPS stamps that help monitor field work.

Preliminary results

Overall, we find that KiuFunza III is effective at improving learning outcomes, as measured by both the programme (incentivised) assessments and the survey (non-incentivised) assessments. Focusing on the results of the non-incentivised survey assessment, we find significant test score increases in numeracy and reading (Kiswahili), especially during the second and third year of the programme. Using a composite learning measure, based on language (Kiswahili reading and writing) and numeracy scores, we estimate test score effects of 0.16 and 0.18 standard deviations, respectively, in years 2 and 3 of the experiment. These estimates translate to a learning improvement equal to one-third of an equivalent year of schooling (EYOS). Despite potential concerns that bonus pay may lead teachers to focus on well-performing students, we do not find systematic differences in the treatment effects by baseline scores in any year.

The incentivised programme assessment displays highly significant test score gains for both subjects too. The pass rate levels and impacts are larger than those for the non-incentivised tests, which we interpret as a "test-day effort" effect (see Mbiti et al., 2019b). Based on the incentivised test data the composite learning effects are 0.37 and 0.44 standard deviations. The finding that the programme is more effective after the first year is consistent with earlier phases. Our interpretation is that trust in the incentive offer is boosted by the receipt of the first-year bonus payments.

Conclusion

The evaluation results for 2015-16 (Mbiti et al., 2019b) and preliminary results for 2019-2021 suggest that the KiuFunza performance pay model is a cost-effective instrument to improve foundational literacy and numeracy. The implementation work itself has provided several practical results, including the design adjustments we described; the experienced regional level CSO and government implementers; and an understanding of the programme by government officials at different levels. Twaweza has received donor support to continue KiuFunza and is currently designing a limited scale-up that will target low-performing schools, using a coalition of regional CSOs to assess public school performance.

References

Banerjee, A.V., Banerji, R., Duflo, E., Glennerster, R., Khemani, S. 2010. Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation in Education in India. American Economic Journal: Economic Policy 2, 1–30. https://doi.org/10.1257/pol.2.1.1.

Bashir, S., Lockheed, M.E., Ninan, E., Tan, J.P. 2018. Facing Forward : Schooling for Learning in Africa (No. 130473). The World Bank.

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, 325–364.

Bloom, N., Lemos, R., Sadun, R., Scur, D., Reenen, J.V. 2014. Jeea-Fbbva Lecture 2013: The New Empirical Economics of Management. Journal of the European Economic Association 12, 835–876. https://doi.org/10.1111/jeea.12094

Bloom, N., Lemos, R., Sadun, R., Van Reenen, J. 2015. Does management matter in schools? The Economic Journal 125, 647–674.

Bold, T., Filmer, D., Martin, G., Molina, E., Stacy, B., Rockmore, C., Svensson, J., Wane, W. 2017. Enrollment without Learning: Teacher Effort, Knowledge, and Skill in Primary Schools in Africa. Journal of Economic Perspectives 31, 185–204. https://doi.org/10.1257/jep.31.4.185.

Evans, D.K., Mendez Acosta, A. 2021. Education in Africa: What Are We Learning? Journal of African Economies 30, 13–54. https://doi.org/10.1093/jae/ejaa009

Glewwe, P., Muralidharan, K. 2016. Chapter 10 - Improving Education Outcomes in Developing Countries: Evidence, Knowledge Gaps, and Policy Implications, in: Hanushek, E.A., Machin, S., Woessmann, L. (Eds.), Handbook of the Economics of Education. Elsevier, pp. 653–743. https://doi.org/10.1016/B978-0-444-63459-7.00010-5

Komba, A., Shukia, R. 2021. Accountability Relationships in 3Rs Curriculum Reform Implementation: Implication for Pupils' Acquisition of Literacy and Numeracy Skills in Tanzania's Primary Schools. RISE Working Paper Series 21/065.

Lieberman, E.S., Posner, D.N., Tsai, L.L. 2014. Does Information Lead to More Active Citizenship? Evidence from an Education Intervention in Rural Kenya. World Development 60, 69–83. https://doi.org/10.1016/j.worlddev.2014.03.014

Mbiti, I., Muralidharan, K., Romero, M., Schipper, Y., Manda, C., Rajani, R. 2019a. Inputs, Incentives, and Complementarities in Education: Experimental Evidence from Tanzania. Q J Econ 134, 1627–1673. https://doi.org/10.1093/qje/qjz010

Mbiti, I., Romero, M., Schipper, Y. 2019b. Designing Effective Teacher Performance Pay Programs: Experimental Evidence from Tanzania (Working Paper No. 25903), Working Paper Series. 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, 55–80. https://doi.org/10.1093/jae/ejaa012

Muralidharan, K., Sundararaman, V. 2011. Teacher Performance Pay: Experimental Evidence from India. Journal of Political Economy 119, 39–77.

Pritchett, L. 2013. Folk and the Formula: Fact and Fiction in Development. WIDER Annual Lecture 016. Helsinki: UNU-WIDER.

World Bank. 2018. World Development Report 2018: Learning to Realize Education's Promise. World Bank. https://www.worldbank.org/en/publication/wdr2018.

World Bank. 2020. Cost-Effective Approaches To Improve Global Learning: What does recent evidence tell us are "Smart Buys" for improving learning in low- and middle-income countries? Recommendations of the Global Education Evidence Advisory Panel. World Bank, Washington D.C.

Youdi Schipper is a researcher on the RISE Tanzania team and a research advisor at Twaweza East Africa, an initiative that works on education and issues concerning government transparency and responsiveness in Tanzania, Kenya and Uganda. He is a senior fellow at the Amsterdam Institute for Global Health and Development. His work and publications focus on education reform, impact evaluation and mobile data collection. He manages KiuFunza, a series of experimental evaluations of teacher performance pay designs in public primary schools in Tanzania.

Isaac Mbiti is a researcher on the RISE Tanzania team and an Associate Professor of Public Policy and Economics at the University of Virginia. He has published several papers on education issues in East Africa and has engaged with education policymakers and other stakeholders from the region.

Mauricio Romero is a Principal Investigator on the RISE India team and an assistant professor of economics at Instituto Tecnológico Autónomo de México (ITAM). He was born and raised in Colombia, where he earned a BA in economics (summa cum laude) and a BA in mathematics (cum laude) from Universidad de los Andes. He did a PhD in economics at the University of California, San Diego. Mauricio's work focuses on the bottlenecks that impede high-quality government provision of education, health care, and environmental protection. In conjunction with his empirical research agenda, he works on methodological issues in applied econometrics and statistics.

Please contact info@riseprogramme.org for additional information, or visit www.riseprogramme.org.

RISE is funded by: