Teacher Incentives in Public Schools:

Do they improve learning in Tanzania?

Highlights

- Twaweza's KiuFunza II program implemented two teacher incentive systems (Stadi (levels) and Mashindano (gains)) as an experiment in Grades 1-3 in a representative sample of public primary schools in Tanzania. The incentives were linked to student test scores.
- After two years of treatment (2015 and 2016), both incentive systems improve learning outcomes. The Stadi (levels) incentive program is easier to communicate and implement while it is as effective as Mashindano (gains).
- The effect sizes estimated are between 0.06-0.17 SD, depending on subject and test setting. The low-stakes data estimates translate into a learning impact of one-third of a year of schooling (added to the *business as usual* learning progress).
- There is no effect, negative or positive, on grades or subjects that are not tested or incentivized.
- For the average KiuFunza teacher, the incentive payment amounts to less than half of a net monthly salary (3.5 percent of the annual salary).

1. Introduction

All children should learn to read and calculate in the first years of primary school. Teachers are key agents and the most costly resource employed to make this happen. In Tanzania, however, pupil learning progress is inadequate: Only 38 percent of children aged 9-13 are able to read and do arithmetic at Grade 2 level (Uwezo, 2017). Teacher motivation is low too: 47 percent of teachers were absent from their classroom during unannounced primary school visits in 2014 (World Bank, 2015).

Teacher performance pay links teacher motivation and learning: it offers a financial

This brief was written and produced by
Twaweza East Africa.

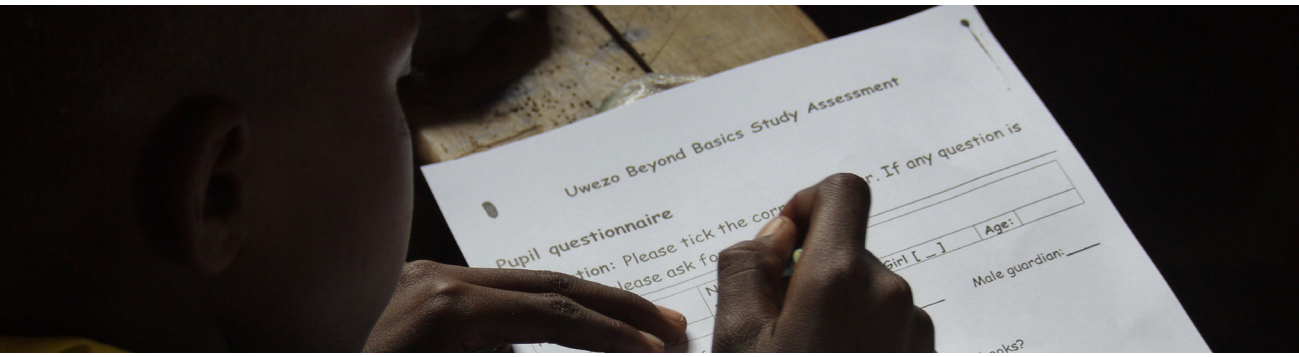
P. O. Box 38342, Dar es Salaam, Tanzania.
t: +255 22 266 4301 | e: info@twaweza.org |
www.twaweza.org/sauti

reward based on teacher performance, typically measured by their students' learning. This constitutes a fundamental change in human resource management in an environment in which well performing teachers are not recognized nor rewarded for their efforts.

Teacher performance pay systems have been shown to improve student learning in different settings (Bruns and Luque, 2014; Glewwe and Muralidharan, 2016), but important design questions remain. One trade-off is between more simple proficiency levels based incentives versus more complex systems based on learning gains that could be more effective but are harder to implement and to communicate to teachers (Neal, 2011).

In 2015-2017 Twaweza East Africa implemented a randomized performance pay trial in the early grades of a nationally representative sample of primary schools. This trial is labeled KiuFunza II and constitutes the second phase of KiuFunza, an experimental program to improve learning introduced by Twaweza in collaboration with J-PAL/IPA.[1] KiuFunza II implemented two different teacher performance pay systems. The first system is called Stadi (levels) and rewards teachers based on the number of students that reached specific proficiency levels. The second is called Mashindano (gains[2]) and rewards teachers based on their students' test score ranking relative to children with the same starting level. In theory, rewarding learning gains should produce better results, as it incentivizes teachers to improve learning across the entire student distribution (rather than focusing on students close to a proficiency threshold). However, implementing a system based on proficiency levels is easier.

The performance pay learning impact was studied in a sample of 180 schools across Mainland Tanzania (60 schools randomly selected into each of the two incentive pay programs, and 60 control schools). Both teacher performance pay systems improved student test scores. Interestingly, the evaluation finds that the simpler "levels" system was at least as effective in raising student learning as the more complex "gains" system. Furthermore, the levels scheme had a more equitable distribution of benefits, improving learning across all initial ability levels.



1 KiuFunza I (2013-2015) implemented three treatment arms: a full curriculum proficiency bonus system, a capitation grant and a combination arm. See Mbiti et al. (2016) for details.

2 In Kiswahili Mashindano literally means "competitions" but we refer to this treatment as "gains".

Background

In Tanzania, government expenditure on education equals 16 percent of the total government budget in 2015/16.[3] Like many other education systems, teachers are the most costly resource in Tanzanian schools: teacher wages accounted for 82 percent of the 2011-12 primary education budget.[4] Teacher goals are set in terms of the curriculum, inputs such as the curriculum based work plan and the time a teacher should spend teaching and preparing. While these system goals are set centrally, administrative responsibility for schools is decentralized to local governments that have limited capacity for school or class level monitoring. Day-to-day supervision of teachers takes place at the school level and external quality inspections are rare. These inspections are ex-post and usually do not address student learning.

A key input indicator that reflects motivation is whether a teacher is in class during school hours. Recent nationally representative studies, which used unannounced monitoring of primary school classrooms, estimated teacher classroom absence at 47 percent (World Bank, 2015); and at 65 and 59 percent among Grade 1-3 teachers (KiuFunza 2015 and 2016 data). Teachers who are at school but not in the classroom represent the larger share of this absence. Teachers do not fear dismissal or transfer because of low performance. Moreover, teachers express reservations about their profession: about 40 percent of lower grade teachers say they would not choose teaching again if they could start over.[5]

In short, there are serious and visible problems with teacher motivation and learning outcomes are falling far short of curriculum expectations. Classroom absence and low motivation may explain why increased budgets for education have not resulted in improved learning outcomes. These observations have led to an emerging sense in Tanzania that evidence is required to inform policies that may lead to teacher performance improvement. One of the leading ideas in this domain is teacher performance pay.

3 UNICEF, 2016.
4 Data from Joshi and Gaddis (2015).
5 Findings based on KiuFunza teacher surveys (unless stated otherwise).

2. Why teacher performance pay?

Reviews of evidence point to three broad families of interventions that improve learning: pedagogy that matches teaching to students' learning levels; detailed guidance for teacher pedagogy; and improving school governance, accountability and incentives (see recent systematic reviews, e.g. World Development Report 2018; Evans and Popova, 2016; Glewwe and Muralidharan, 2016). Teacher performance pay is one element of a range of policies that can improve governance and teacher incentives, including professional rewards (recognition, professional mastery), accountability pressure (threat of dismissal and managerial pressure), and financial incentives (bonus pay, pension and other benefits, career path reforms; see Bruns and Luque, 2014).

Teacher performance pay offers a number of practical and political advantages. It offers an incentive with no financial downside risk for teachers and is therefore easier for teachers to accept. It has a clear quid pro quo (cash for learning outcomes): the desired outputs can be precisely specified, linked to curriculum goals and compensated. Performance pay presents an opportunity to achieve learning results at low cost compared to base salaries and addresses wastage of public money caused by teachers not fulfilling their contractual obligations. The cash bonuses also offer flexibility as no long-term commitments need to be put into law. Effects can be expected in the short term *without recruiting and training new teachers*.

The fundamental innovation of teacher performance pay is that it recognizes and rewards teachers who take initiative and care for their pupils. A teacher performance pay system communicates a number of incentive signals. First, the system emphasizes student level learning and testing.

Second, it sends a message from the ministry about the importance of learning outcomes. Third, it provides a platform to account for and provide feedback about teacher performance. With good communication teachers' efforts become visible for managers and colleagues (as well as the teachers themselves) and will be talked about. And fourth, there is the financial reward to focus attention.

Performance pay is controversial too. The main criticisms are that teachers may not have the capacity to increase the desired learning outcomes, or that they are constrained by factors outside their control (e.g. low student attendance). Bonuses may cause jealousy among other colleagues in schools. Others argue that intrinsic motivation suffers when bonuses are offered, or that it triggers perverse effects such as cheating or abandoning poorly performing students. A number of these issues can be addressed by the incentive design.

Bruns and Luque (2014) are cautiously optimistic about the balance of evidence, particularly for bonus pay programs in developing countries that have " *.. produced more consistently positive results than in developed country settings to date.*" They argue that these programs will be most productive where other systems of monitoring and accountability are weak.

3. The KiuFunza incentive design

The KiuFunza incentive design has a number of noteworthy features. Incentives are offered to individual teachers of students in Standards I, II and III of treated primary schools. The incentives are calculated based on test results at the stream level and a teacher is paid only for the streams he/she is responsible for. In addition, head teachers receive 20 percent of what the teachers in their school earn. Student drop-out is penalized because it lowers the number of test taking students at endline and therefore the amount of bonus money that a teacher can earn.

The focal subjects are Kiswahili and English reading, and Maths. In KiuFunza I, the proficiency bonus was fixed at TZS 5,000 (about 3 USD), conditional on a student passing all the curriculum skills for one focal subject. With three bonus subjects, the maximum bonus was TZS 15,000 per student. Because of the fixed per student-pass bonus, the incentive budget for KiuFunza I was open ended. The strength of this system is the clarity and simplicity of the incentive offer. A drawback is that the full curriculum proficiency condition is too demanding for many students. Mbiti et al. (2016) show that test scores in incentive-only schools were higher in the high-stakes exam (used to calculate the bonus), but not in the low-stakes exams (used for research purposes only). The KiuFunza I low-stakes exam data do show a significant positive learning impact for the combination treatment arm.

The findings of the KiuFunza I trial suggested a basic design change: create an incentive condition that is reachable for pupils with lower skill levels. This resulted in two new types of incentive treatments in KiuFunza II, called Stadi (levels) and Mashindano (gains), which became the experimental arms of the KiuFunza II trial.

The levels treatment is implemented as a skill threshold system, paying a bonus directly to teachers for each student who passes grade-specific skills outlined in the national curriculum; for example, consecutive skills in the Kiswahili curriculum for Standard II are: can read syllables; can read words; can read sentences; can answer comprehension questions. The amount paid to teachers per skill is set to ensure that payouts are equal across each grade-subject combination. Because the amount paid out to teachers for each skill passed is dependent on how many students pass, the exact bonus rate is unknown before the end of the year tests. Payment amounts are calculated this way to ensure budget comparability across the levels and gains designs, and also in an effort to reward teachers proportionally to how many skills their students learn as well as how hard each skill is to learn.

The gains treatment arm rewards teachers for learning gains even when a student cannot meet the levels threshold, or is well beyond it. This intervention arm is implemented as a teacher tournament design, as described by Barlevy and Neal (2012). At the beginning of the school year students are grouped based on their initial levels of learning (based on the test scores from the previous year, and schools' historic test scores for students in Grade 1 who have not been previously tested). At the end of the school year students are tested again and ranked within each ability group; teachers are paid proportionally to their students' ranks.

This tournament in gains recognizes all learning improvements that occur, regardless of a student's initial learning achievements. This will be particularly important for teachers in poor communities with students who have initial learning levels that are very low and far from the passing threshold in the levels design, as well as for students with initial learning levels that are well above the passing threshold. The gains design thus has two theoretical advantages. First, because students "compete" within ability groups it incentivizes all teachers, regardless of their students' initial learning levels. Second, there are no performance thresholds in this system.

Rewards are offered continuously for improvements across the entire test score distribution of students, and teachers are not encouraged to focus on students near a learning threshold.

A budget of USD 150,000 for teacher incentives was split between the two treatment arms each year proportional to the number of students enrolled. As a result, the total prize in each treatment arm was approximately $3 per student. This was to ensure that the budgets of the two designs would be directly comparable. However, since the aggregate treatment budgets were fixed there was some uncertainty about the bonus size at the teacher level (since they would not be calculated until after student outcomes were measured).

Mashindano (gains) is more complex to understand than Stadi (levels), but it is also more equitable for the teachers since students compete at their own level. This is not the case for Stadi (levels), where teachers with higher level students (often in urban schools) tend to earn more. Twaweza decided to test both systems as a comparison: Stadi pays for levels and is easier to understand, while the more complicated Mashindano design pays for value-added and offers incentives along the full skill distribution.

4. The KiuFunza sample and implementation

The teacher incentive programs were evaluated using a randomized design. First, ten districts were randomly selected. The study sample of 180 schools was taken from the KiuFunza I field experiment (Mbiti et al., 2016) where all students in Grades 1, 2, and 3 had been tested at the end of 2014. These school level tests provided the necessary baseline student-level test score

information that we needed in order to implement the gains treatment. Within each district, we randomly allocated schools to one of our three experimental groups. Thus, in each district, six schools were assigned to the levels treatment, six schools to the gains treatment, and six schools served as our control group. In total, we have 60 schools in each experimental group in the research sample.[6] The sample was also stratified by treatment of the previous RCT and by an index of the overall learning level of students in each school. All of our specifications control for the three levels of stratification: district, treatment in the previous RCT, and overall school quality.

All interventions were implemented by Twaweza in partnership with EDI, a Tanzanian research firm, and local district partners. Within each intervention arm, information describing the program was distributed during baseline school visits via teacher meetings and FAQ leaflets and to parents via booklets. This baseline visit took place early in the school year (Figure 1). During these visits Twaweza asked each teacher in the Stadi (levels) and Mashindano (gains) schools whether they agreed to participate. All teachers signed up for the program and provided their grade-subject assignment and payment account details.[7] The implementation teams also conducted additional mid-year school visits to re-familiarize teachers with the program, gauge teacher understanding of the bonus payment mechanisms, and answer any remaining questions.

Figure 1: KiuFunza implementation cycle summary

At the end of the school year, all students in Grades 1, 2, and 3 in every school, including control schools, were tested in reading Kiswahili, reading English and Maths. As this test was used to determine teacher incentive payments, it is high-stakes (from the teacher's perspective). The tests were developed by Tanzanian education professionals, following a similar test development framework as the Uwezo Annual Learning Assessment that is widely used in East Africa. The test results were digitized and processed. By early April of the following year teachers received the bonus payment in their bank or mobile money account, according to their choice. This cycle of activities took place once in 2015, and once in 2016.

At various points during the implementation, teachers were asked questions to test their understanding of the program (since they would not be able to respond without understanding the program). These tests typically showed sufficient to high percentages answering question about the program correctly. The highest average score is 91 percent correct, the lowest 62 percent correct. Moreover, mean percent correct was higher in gains schools than in levels

6 On top of the 120 research intervention schools, 14 other non-research schools received one of the two bonus programs for field testing.

7 Once a teacher declined at first but later changed her mind and asked to be registered.

schools. Trust in Twaweza to calculate and pay the bonuses in a transparent manner was high, at 95 percent.

Teachers were also asked for their opinion about performance pay programs in general. Nine out of ten teachers (91%) support the idea of performance pay. When asked whether the government should include a performance-based bonus scheme in a future salary review, 63 percent of teachers say yes (while 37 percent prefer just a flat increase).

Table 1 shows the incentive amounts earned in 2016 (paid to teachers in 2017). The total budget *per student* is equal between the two interventions, but since the total number of students is slightly larger in the Stadi group the *total* budget amount is also slightly larger. Head teachers receive 20 percent of subject teacher earnings.

Table 1: KiuFunza incentive payments 2016 (in TZS)

Incentive	Teachers	Head Teachers	Totals
Mashindano	97,966,883	19,593,377	117,560,260
Stadi	111,888,984	22,377,797	134,266,780
TOTAL	209,855,867	41,971,173	**251,827,040**

The total value of all bonus payments to subject teachers was TZS 209,855,867. This amount was earned by 788 early grade teachers. Therefore, the average teacher bonus equals TZS 266,315 (about USD 127 in 2016). The average net monthly teacher salary in 2016 was TZS 637,790 so the average bonus is about 42 percent of that monthly wage or 3.5 percent of the annual salary. The mean annual increase in nominal salaries in this sample, based on teacher reports, is 11 percent. [8]

8 Salary is self-reported by teachers and represents total compensation per month (including allowances), excluding pension.

5. Impacts on student learning

Core results

Table 2 shows the second year impact of the incentives on student learning in Maths and Kiswahili using both the low-stakes (Panel A) and the high-stakes data (Panel B). Based on the low-stakes test data, both incentive systems raised math test scores by a modest (0.07 SD) but statistically significant amount (Panel A). We find statistically significant increases in Kiswahili for students in both the levels system (0.11 SD) and the gains system (0.06 SD), although the difference between the two systems is not significant.

To put this improvement in perspective, we calculate that the impact in equivalent years of schooling (EYOS) for Kiswahili, Stadi (levels), is 0.37 EYOS. In other words, children in Stadi (levels) schools improved their Kiswahili skills by 0.37 school years as a result of the incentives, in addition to what they would have learned without the incentives. For Maths, the learning improvement is about one-third of a school year in both treatment arms. For Kiswahili Mashindano (gains), the impact is about 20 percent of an EYOS.

Table 2: KiuFunza Learning impact (in terms of standard deviations)

Year 2 impact	Panel A Research data Low stakes		Panel B Intervention data High stakes	
	Math	**Kiswahili**	**Math**	**Kiswahili**
Levels (Stadi)	0.07*	0.11***	0.12***	0.17***
Gains (Mashindano)	0.07**	0.06*	0.10**	0.08*
Gains - Levels	0.00	-0.06	-0.02	-0.09*
Nr of students	4,869	4,869	59,755	59,755

Note: The final row in each panel labeled "Gains-Levels" shows the difference between the student learning gains from the "gains" design and the "levels" design, in standard deviations (SD).
Stars indicate significance levels: *p<0.1; **p<0.05; ***p<0.01.

As most of the existing literature on pay for performance uses a single high-stakes test to determine teacher rewards as well as evaluate the program, we present the treatment effects of our interventions using our high-stakes data in Panel B. Generally, the estimated treatment effects are larger compared to those estimated using the low-stakes data in Panel A. High-stakes test scores in year two increased between 0.10 and 0.12 SD in Maths, and 0.08 and 0.17SD in Kiswahili. However, the differences between the high- and low-stakes data impact estimates are not statistically significant.

The larger treatment effects found in the high-stakes data are likely driven by test-taking effort, where teachers have incentives to motivate their students to take the tests seriously. The

importance of student test-taking effort has been documented in other settings such as an evaluation of teacher and student incentives in Mexico city (Behrman et al., 2015).

Although the content (subject order, question type, phrasing, difficulty level) is consistent across low- and high-stakes tests, there are a number of important differences in the test administration. The low-stakes test took longer (~40 minutes) than the high-stakes test (~15 minutes). Not only were there more questions in each subject section, the low-stakes test included an 'other subjects' module at the end to test spillover effects. Moreover, the testing environment was quite different. Low stakes tests were administered by a data firm enumerator, who would take ten sampled children out of the classroom and test them one by one during a regular school day.

Children in Stadi (levels) schools improved their Kiswahili skills by 0.37 school years as a result of the incentives, in addition to what they would have learned without the incentives.

In the high-stakes test, all students in grades I-III were tested on an agreed test day. On this day, higher grades were given a day off, while Twaweza test teams administered the one-on-one tests in designated classrooms. One test out of ten test sets was randomly assigned to a student for additional test security. Each student listed at baseline had an individual pre-printed test form. Test forms were handled only by the administrators. Shortly after the tests, forms were digitized via optical character recognition software and grade tallying was automated. High stakes test taking and test data handling were tightly controlled, which suggests to us that test-taking effort rather than gaming explains the difference between the two data sets.

English

Starting in 2015 English was removed from the national curriculum in Grades 1 and 2. As a consequence the curriculum for Grade 3 changed. However, there was a lot of variation in how the curriculum changes were actually implemented by schools.[9] As a result we dropped English from the incentives in Grade 1 and 2 in 2016, but included Grade 3 English teachers in the incentives. To avoid confusion, we also communicated that the end of year English test in 2016 would still use the existing Grade 3 curriculum.

The English test score impacts thus need to be interpreted with care, against a background of non-uniformly implemented curriculum changes. The estimated test score impacts are presented in Table 3. We find positive point estimates for both low-stakes and high stakes data, with a significant coefficient of 0.19 SD for the gains system. As before, the difference between the coefficient estimates for levels and gains is not significant. Again, the high-stakes point estimates are higher (and more precisely estimated).

Table 3: KiuFunza Learning impact English (standard deviations)

Year 2 impact	Panel A Research data Low stakes	Panel B Intervention data High stakes
	English	**English**
Levels (Stadi)	0.11	0.26***
Gains (Mashindano)	0.19**	0.23***
Gains - Levels	0.08	-0.04
Nr of students	1,533	15,493

Note: The final row in each panel labeled "Gains-Levels" shows the difference between the student learning gains from the "gains" design and the "levels" design, in standard deviations (SD).
Stars indicate significance levels: *p<0.1; **p<0.05; ***p<0.01.

Other findings

One concern regarding teacher incentives that focus on certain subjects and grades is that teachers may cut back on non-incentivized subjects, and schools may shift resources to focal grades, away from other school grades. For example, schools may shift resources such as textbook purchases from higher grades to Grades 1, 2, and 3. Additionally, teachers may cut back on teaching non-incentivized subjects such as Science. On the other hand, if our incentive programs improve literacy and numeracy skills, they may promote student learning in other subjects.

9 Some schools stopped teaching English in 2015, while others stopped in 2016. There was also no official guidance on whether to use Grade 1 English materials in Grade 3 as there were no new books issued to reflect the curriculum changes.

Overall, we do not find a significant effect on Grade 4 test scores, although the point estimates are positive for the levels design, ranging from 0.04 to 0.13 SD. This suggests that schools did not shift resources from higher grades to the KiuFunza focal grades. There also was no significant effect on the non-incentivized science test scores, although the point estimates were generally positive.

While we find positive test score effects, data on teacher effort do not provide a clear explanation for these (as reported in other teacher incentive studies, e.g. Muralidharan and Sundararaman, 2011). We do not find differences in teacher attendance, either at school or in the classroom. We find some suggestive evidence of more help provided by teachers under the levels system in the first year but not in the second year. We also observe generally higher propensities to assign homework by teachers under the levels system compared to teachers in the gains system. These differences are statistically significant but the individual point estimates are not significant.

Effect heterogeneity

We use mean learning differences by initial student test score quintile to qualitatively assess the pattern of benefit distribution. In the first year of the program, improvements in learning benefitted the more advanced students across both incentive programs. However, learning benefits were generally more equitably distributed under the levels system than under the gains system.

For example, in the first year improvements in math test scores are concentrated in the highest quintile in the gains (pay for percentile) system: Maths teachers appear to focus a lot of attention on the top 20 percent of students, whereas teachers in the levels system focus on the top 60 percent of their class. In the second year of the program, for Maths we do not see such overt focus on top students. Learning benefits were more broadly distributed across all students for maths in both incentive designs, even reaching the bottom two quintiles.

In Kiswahili, learning gains reached the top four quintiles of students in year one and all quintiles in year two under the levels design. However, teachers in the gains design focused on the very best students, with only the top 20 percent improving their test scores across both years. These results suggest that, while the gains design in theory should motivate teachers to help students across the full initial skill distribution, this design actually had the opposite effect in practice.

We observe hardly any differences in learning impact across student or teacher characteristics. In other words, the learning improvements were spread out across different types of students and teachers. There were no differences in learning treatment effects by student gender, age or pre-school attendance. Significant heterogeneity is also absent in terms of teacher characteristics, including an index of teacher content knowledge, teacher gender, and age. Lastly, there were no significant differences in learning gains based on school facilities or distance to urban areas, but schools with higher student-teacher ratio benefited less in math in the gains design.

6. Policy relevance

The KiuFunza II impact study shows that relatively modest changes in compensation structure, which reward teachers on the basis of student test performance, can generate substantial improvements in learning outcomes in Tanzanian public primary schools. The impact estimates of 0.06-0.17 SD are comparable in magnitude to findings of other teacher incentive studies (including those in the review by McEwan, 2015; and the language estimate in Muralidharan and Sundararaman, 2011).

The impact findings are meaningful for policy because the learning improvements are attained at a fraction of the cost of typical 'business as usual' expansion in education spending, such as salary increases. Close to 50 percent of teachers in our study are not satisfied with their salary and promotion potential, but teacher salaries are over four times GDP per capita. Given the size of the workforce, increasing teacher salaries is very costly. More importantly, it is unlikely that simply raising salary levels will improve performance (e.g. De Ree et al., 2018). However, a majority of teachers approve of including some measure of performance pay in future salary reviews.

Another 'business as usual' measure is to increase the number of teachers, thus reducing the student teacher ratio (STR). A small number of studies summarized in Glewwe and Muralidharan (2016) show that reducing this ratio can improve students' time in school and their test scores. However, reductions in class size are typically expensive. For the period 2007/08 – 2011/12, Joshi & Gaddis (2015) calculate that the personnel emoluments component of recurrent education sector spending increased at an annual average rate of 13.4 percent. This growth rate represents substantial increases in the number of teachers. Given the size of classrooms and large student-teacher ratios, these increases are defensible but they are costly.

The KiuFunza results suggest that performance pay can motivate teachers, both experienced and new, to perform better and thus make more efficient use of the most costly resource in the education system. The size of the mean KiuFunza bonus suggests that adding the equivalent of three to four percent of nominal teacher wages as a performance bonus linked to student results has the potential to significantly improve learning outcomes.

References

Barlevy, G. and Neal, D. (2012), 'Pay for Percentile', *American Economic Review* 102(5), 1805-31.

Behrman, J. R.; Parker, S. W.; Todd, P. E. and 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.

Bruns, B. and Luque, J. (2015), Great teachers: How to raise student learning in Latin America and the Caribbean, World Bank Publications.

De Ree, J.; Muralidharan, K.; Pradhan, M. and Rogers, H. (2017), 'Double for nothing? experimental evidence on an unconditional teacher salary increase in Indonesia', *The Quarterly Journal of Economics*, forthcoming.

Evans, D. K. and Popova, A. (2016), 'What really works to improve learning in developing countries? An analysis of divergent findings in systematic reviews', *The World Bank Research Observer* 31(2), 242--270.

Glewwe, P. and Muralidharan, K. (2016), Improving education outcomes in developing countries: Evidence, knowledge gaps, and policy implications, *Handbook of the Economics of Education*, Vol. 5, Elsevier, pp. 653--743.

Joshi, A. R. and Gaddis, I. (2015), Preparing the next generation in Tanzania: Challenges and opportunities in education, World Bank Publications.

Mbiti, I.; Muralidharan, K.; Romero, M.; Schipper, Y.; Rajani, R. and Manda, C. (2016), 'Inputs, incentives, and complementarities in primary education: Experimental evidence from Tanzania', Unpublished Paper.

McEwan, P. J. (2015), 'Improving Learning in Primary Schools of Developing Countries: A Meta-Analysis of Randomized Experiments', *Review of Educational Research* 85(3), 353--394.

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

Neal, D. (2011), The design of performance pay in education, 'Handbook of the Economics of Education', Elsevier, pp. 495--550.

UNICEF (2016), Education Budget Brief Tanzania, FY 2011/12–FY 2015/16. https://www.unicef.org/tanzania/UNICEF-TZ-BB-Education-WEB(2).pdf, retrieved February 2018.

Uwezo (2017): Are Our Children Learning? Uwezo Tanzania Sixth Learning Assessment Report. Dar es Salaam: Twaweza East Africa

World Bank. 2018. World Development Report 2018 : Learning to Realize Education's Promise. Washington, DC.

World Bank (2015), 'Education Service Delivery in Tanzania', Technical report, Washington, DC.