RESULTS-BASED FINANCING
RBF EDUCATION | EVIDENCE

WORLD BANK GROUP

TANZANIA

Can a Simple Teacher Incentive System Improve Learning?

MARCH 2018

REACH co-funded an evaluation that examined whether a simple or more complex teacher performance pay system was more effective in increasing student learning.

≠

Evidence shows little correlation between level of teacher salaries and student learning.

Teacher bonuses ∞ Student performance = Learning outcomes

Teacher performance pay systems are an example of RBF that has been shown to improve student learning.

***The Results in Education for All Children (REACH) Trust Fund** supports and disseminates research on the impact of results-based financing on learning outcomes. The EVIDENCE series highlights REACH grants around the world to provide empirical evidence and operational lessons helpful in the design and implementation of successful performance-based programs.*

Despite several major reforms and significant new investments in public education over the last decade, student learning levels in East Africa remain low. Results-based financing (RBF) has been used in many developing countries in an attempt to incentivize teachers and other stakeholders to achieve better results. RBF mechanisms work by making financing conditional on achieving measurable results such as student test scores or other intermediate education outcomes. Teacher performance pay systems are one example of RBF that has been shown to improve student learning in many settings, although its results have been mixed. Education systems with limited administrative capacity currently face a tradeoff between adopting more complex incentive systems that may be more effective but are harder to implement and choosing simpler systems that are easier to implement but may be less effective.

The Results in Education for All Children (REACH) Trust Fund at the World Bank co-funded an evaluation that compared the effectiveness of

This note was adapted from Mbiti, Isaac, Karthik Muralidharan, Mauricio Romero, and Youdi Schipper (2017). Designing Teacher Performance Pay Programs: Experimental Evidence from Tanzania, (mimeo).

two different teacher performance pay systems in early primary schools in Tanzania. These performance pay systems are part of *KiuFunza*, an experimental teacher pay program introduced by Twaweza East-Africa, a civil society organization, in collaboration with the Abdul Latif Jameel Poverty Action Lab (J-PAL), Innovations for Poverty Action (IPA) and Economic Development Initiatives (EDI). The incentive design rewarded teachers based on the number of specific milestones (or proficiency) levels each of their students could achieve. The second was a more complex system that first grouped students by baseline test scores, and then rewarded teachers based on the rank ordering of each of their students within each group. Hence, this system rewards teachers based on the gains of their students within the structure of a rank order tournament. In theory rewarding learning gains through a rank order tournament with rankings determined within sets of similar students at baseline should produce better results because all teachers are rewarded regardless of their students' initial learning levels and because it should incentivize teachers to improve learning across the entire student distribution.

While the evaluation found that both systems raised test scores, despite the theoretical advantages of the more complex learning "gains" system, the simple learning "levels" system was at least as effective in raising student learning levels. Furthermore, the benefits of the simpler scheme were more equitably distributed across students from all five quintiles, while the more complex scheme primarily benefited students in the top quintile. These results highlight the critical importance of the design of RBF schemes. By rewarding teachers for student achievement at multiple learning levels rather than just one, the simple scheme overcame one of the disadvantages of similar proficiency levels-based systems with minimal added complexity.

CONTEXT

Tanzania invests 3.5 percent of its GDP in its education sector, which is below the Sub-Saharan Africa average of 4.5 percent. Student learning levels in the country remain low, with large majorities of children unable to read or do arithmetic at the required level.[1] While these challenges are well known, reforms have largely failed to improve these results.

One of the many issues in Tanzanian schools is that no one is held accountable or is incentivized to improve learning. Teachers are paid regardless of their attendance or performance. Even when teachers are in the school, they are often not in the classroom teaching. Teachers spend only 40 percent of their time on task doing instructional activities, according to a World Bank survey of service delivery indicators.[2] Teacher salaries and benefits account for almost two-thirds of Tanzania's education budget, while the average teacher in Sub-Saharan Africa earns almost four times per capita GDP.[3] Despite already high wages, the Ministry of Education (MoE) has faced sustained pressure from the teachers' union to increase teacher pay, with proponents arguing that this would motivate teachers to improve student learning. However, a large body of evidence has shown that there is little correlation between teacher compensation and student learning.[4/5/6] Without addressing teacher accountability and incentives, simply increasing the volume of resources is unlikely to be effective in raising the test scores of Tanzanian students.

In contrast, introducing teacher performance pay systems could give teachers an incentive to help

Poor accountability

Amount of time teachers are off task: 60%

their students to learn by linking their pay to their students' learning outcomes. Twaweza, an East African civil society organization, first developed and launched teacher incentive programs in Tanzania in 2013 under a broader umbrella program called *KiuFunza* ("thirst for learning" in Swahili). The first teacher incentive program had a simple design that would be relatively easy to implement at scale. Both of the incentive programs evaluated by this report were implemented by Twaweza East-Africa in partnership with EDI, a Tanzanian research firm, and local partners in each district.

WHY WAS THE INTERVENTION CHOSEN?

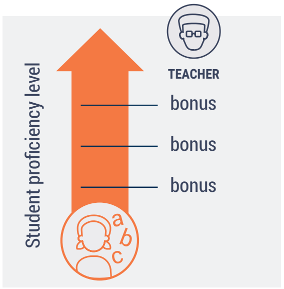
Teacher performance pay systems have been implemented in several developing countries, but evidence of the effectiveness of these programs is mixed. This heterogeneity is driven in part by large differences in the way in which the incentives were designed.[7] In general, incentives designed to reward teachers based on student learning gains have been more effective than systems that reward teachers based on simple learning levels. However, it is difficult to compare these results because of differences in the context, design, and budgets of the different schemes. There is little research that directly compares different systems. Furthermore, there may be tradeoffs between choosing incentive schemes that reward learning gains and choosing those that reward learning levels. Rewarding teachers based on a simple student proficiency level may penalize those teachers who serve students from disadvantaged backgrounds and encourage teachers to focus only on those students who are close to the threshold. On the other hand, rewarding teachers based on learning gains should in theory incentivize them to improve learning across the entire student distribution and should be more equitable for all teachers regardless of their students' initial levels. However, this type of scheme requires maintaining a complex database of students' performance to calculate the "value added" for each teacher, which is difficult for countries with limited administrative capacity like Tanzania to implement. These systems may also be more difficult for teachers to understand, which may weaken the incentive.

Therefore, the objective of this evaluation was to compare the effectiveness of two teacher incentive programs, both implemented in the same context and with the same budget. One scheme had a simple learning "levels" design that rewarded teachers based on the number of students who reached specific proficiency levels, and the second was a more complex scheme that rewarded teachers based on the average learning "gains" that their students achieved relative to their initial learning levels.

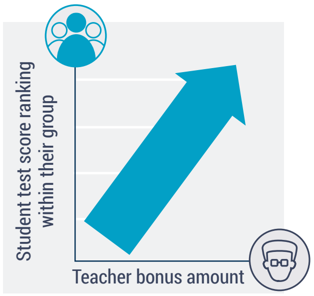
HOW DID THE INTERVENTION WORK?

In the simple learning “levels” scheme, students were tested at the end of the school year, and teachers were rewarded for the number of students who reached various levels of proficiency. Teachers were rewarded for students’ mastery of grade-specific and subject-specific skills, ranging from very basic to more advanced (for example, grade one students in Swahili were assessed on three skills—letters, words, and sentences). The total amount of money available for teacher bonuses was the same for each type of skill, so that more advanced skills that were achieved by fewer students led to higher teacher bonuses per student who reached the required level.

LEVELS



GAINS



In the learning “gains” scheme, students in all schools participating in the scheme were tested at the beginning of the school year and grouped according to their initial learning levels. At the end of the school year, the students were tested again and ranked within their group. Teachers were paid in proportion to their students’ ranking within each group. This incentive design has two theoretical advantages. First, it does not penalize teachers who serve disadvantaged students so it incentivizes all teachers to exert more effort, regardless of their students’ initial learning levels. Second, because the rewards are given for improvements across the entire distribution of students, teachers are encouraged to focus on all students rather than only on students near the learning thresholds. In theory, under certain circumstances this design can maximize learning gains across the entire student population.

The program focused on Math and Swahili teachers in grades one, two, and three. Both incentive designs had a fixed bonus pool of $75,000 split between each subject-grade combination, with an average bonus of $3 per student or roughly $125 per teacher. This was to ensure that the budgets of the two designs would be directly comparable. However, this also led to some uncertainty about the size of teachers’ payments since they could not be calculated until after student outcomes were measured. This may have affected how the teachers responded to the incentives. The program implementers provided information about each incentive program to schools and their communities at public meetings at the beginning of each school year. The implementers used culturally appropriate materials, examples, and analogies to convey the features of the program. They also revisited each school in the middle of the school year to refresh teachers’ knowledge of the program and test their understanding, which was generally considerable in the case of both incentive schemes.

The evaluation was implemented in 180 randomly selected schools across 10 districts in Tanzania, with 60 schools in each incentive scheme and 60 more in the control group. All students completed a baseline test, a “high stakes” endline test that was used to determine teacher bonuses and assess the program’s impact, and a “low stakes” endline test that was only used to measure the program’s impact. The two endline tests were similar except that the low stakes test covered was longer and covered a wider range of curricula concepts and learning domains, including material that was not incentivized. In addition, the enumerators collected data on the characteristics on schools, head teachers, individual teachers, and individual students as control variables and to measure the program’s impact across these characteristics.

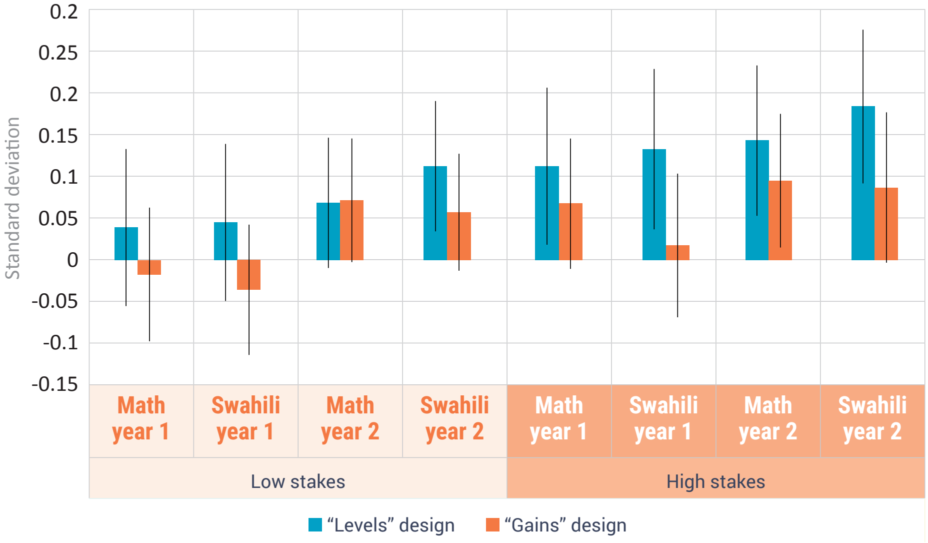
WHAT WERE THE RESULTS

Both incentive schemes significantly raised test scores. However, despite the theoretical advantages of the more complex learning "gains" design, the simpler learning "levels" design was at least as effective in raising student learning.

When the "low stakes" test scores were used to measure program impact, math test scores increased by 0.07 standard deviations (SD) under both systems in the second year of the evaluation. However, the Swahili test scores increased by 0.11 SD under the simple "levels" design compared to only 0.06 SD under the more complex "gains" design, a difference of 0.057 SD (p=0.16). Similarly, when the "high stakes" test scores were used to measure program impact, math scores increased by 0.142 SD in the "levels" design compared to 0.0910 SD in the "gains" design, a difference of 0.044 SD (p=0.31), and the Swahili scores increased by 0.187 SD in the "levels" design compared to only 0.098 SD in the "gains" design, a statistically significant difference of 0.093 SD (p=0.045). Overall, these gains of 0.06 to 0.187 SD are comparable in magnitude to the results from a recent meta-analysis on the use of teacher and student incentives as well as the results of other interventions to improve student test scores, such as computer-assisted learning, teacher training, reducing class size, providing instructional materials, and providing school grants.[8]

In addition, the learning gains from the simple "levels" design were more equitably distributed across all students. In the first year of the program, teachers in the scheme with the "levels" design focused on the top half of their class in math, while in Swahili the top four quintiles of students made substantial learning gains. However, in the scheme with the "gains" design only students in the top quintile improved their test scores, which suggests that teachers focused only on the very best students. In the second year, the gains in math were more broadly distributed across all students in both types of incentive schemes, even those in the bottom two quintiles. However, in the "gains" scheme, Swahili teachers seem to

Figure 1: Effects of Teacher Incentives on Student Test Scores



have continued to focus mostly on students in the top quintile, while all students achieved gains in the "levels" scheme, suggesting that teachers focused on all students. Therefore, despite the theoretical advantage of the "gains" design in motivating teachers to help all students across the distribution, the results suggest that this kind of incentive scheme actually had the opposite effect.

The learning gains were broadly distributed across various student, teacher, and school characteristics. There were no significant differences in students' learning gains by gender, age, or pre-school attendance. Likewise, there were no significant differences by teachers' gender, age, or content knowledge. Lastly, while there were no significant differences in learning gains based on school facilities or proximity to urban areas, schools with higher student-teacher ratios benefited less in math in the "gains" design.

Learning gains did not come at the expense of other subjects and grade levels. One potential concern about implementing teacher incentives only for some subjects and grade levels was that teachers might cut back the effort that they put into teaching non-incentivized subjects and that schools might shift resources away from other grades to grades one to three. On the other hand, it was possible that positive learning gains would spill over into other subjects or could persist over time to later grades. Overall, neither incentive scheme had a significant effect on grade four learning, although the point estimates were positive for the "levels" design, ranging from 0.04 to 0.13 SD. This suggests that schools did not shift any resources away from students in higher grades and that the learning gains made by grade three students in the first year of the scheme may have persisted when they moved into grade four in the second year. Likewise, there was no significant effect on science test scores for grades one to three, although the point estimates were generally positive, suggesting that the incentives may have had some positive spillover effects on other subjects.

Teachers exerted more effort in the simple "levels" design than in the "gains" incentive scheme, and the program results were not driven by any differences in teacher comprehension.

While there were no differences in teacher attendance, teachers in schools in the "levels" scheme were more likely to be on task, less likely to report that their students were disengaged, and assigned more homework. In the first year, teachers in the "levels" scheme were also more likely to provide extra help to their students. Furthermore, teachers were given comprehension tests to ensure that they understood the incentive program to which they were assigned. Their comprehension was generally good and roughly equal in both programs. In fact, the point estimates of teacher comprehension were higher for the "gains" scheme, so there is no evidence that the lower learning gains in that incentive design were driven by a lack of comprehension of the incentives by teachers.

WHAT WERE THE LESSONS LEARNED?

Previous research in Tanzania had found that the effectiveness of the "levels" incentive design was limited, particularly for students far above or below the threshold, when it only set one proficiency level for the whole curriculum.[9] To address the issue of teachers focusing only on students close to the threshold, the *KiuFunza*, version of the "levels" design that was evaluated here included multiple thresholds at various points along the student learning distribution, so that teachers could earn bonuses for helping a broad set of students. However, even with multiple thresholds, because the simple "levels" design did not consider students' initial learning levels, it still did not offer rewards to teachers for all students' improvements across the entire distribution of test scores. These results suggest that including multiple thresholds in the simple "levels" design overcame the limitations of the earlier incentives scheme. However, there are still many other potential variations of the incentive design that have not been tested. Continuing to experiment with small tweaks to the design could have big payoffs in terms of maximizing learning gains. While it may not be feasible to conduct randomized evaluations of several incentive designs, these design tweaks could be tested and compared using a series of smaller experiments, for example, using an "A/B test" approach in which two alternative designs are compared on an outcome that can be assessed quickly. These tests could be used to collect low stakes test scores over a short period of time or intermediate outcomes such as classroom observations.

Continuing to experiment with small tweaks to the design could have big payoffs in terms of maximizing learning gains.

When it comes to RBF in Tanzania, simpler is better, but further research will be needed to establish the most effective ways to use teacher performance pay systems.

CONCLUSION

A simple teacher incentive scheme that rewarded teachers based on the number of students who achieved specific learning levels improved learning at least as much as a more complex scheme that rewarded teachers based on learning gains. Furthermore, contrary to expectations, the simpler design also benefited a broader set of students, while the more complex scheme led teachers to focus primarily on the best students. Given the limited administrative capacity in Tanzania and other developing countries to implement complex RBF schemes, this kind of simple incentive design that rewards learning levels may be the most suitable to be implemented on a wide scale. However, within this simple incentive scheme, certain design features are critical, particularly the need to use multiple learning thresholds rather than just one threshold so that teachers can earn bonuses for learning achievements across the entire student distribution. Further research will be needed to establish the most effective ways to use teacher performance pay systems to narrow the learning gap and effectively target the most vulnerable students.

[1] Uwezo. (2017). *Are Our Children Learning? Uwezo Tanzania Sixth Learning Assessment Report.* Dar es Salaam: Twaweza East Africa (Tech. Rep.).

[2] World Bank. (2011). Service delivery indicators: Tanzania (Tech. Rep.). The World Bank, Washington D.C.

[3] World Bank. (2017). World Development Indicators. http://databank.worldbank.org/data/reports.aspx?source=world-development-indicators

[4] Kane, T. J., J.E. Rockoff, and D.O. Staiger, (2008). "What does certification tell us about teacher effectiveness? Evidence from New York City." *Economics of Education Review,* 27 (6), 615-631.

[5] Bettinger, E. P., and B.T. Long (2010). "Does cheaper mean better? The impact of using adjunct instructors on student outcomes." *The Review of Economics and Statistics*, 92 (3), 598-613.

[6] Woessmann, L. (2011). "Cross-country evidence on teacher performance pay." *Economics of Education Review,* 30 (3), 404-418.

[7] Glewwe, P. and K. Muralidharan. (2016). "Improving education outcomes in developing countries: Evidence, knowledge gaps, and policy implications" in The Handbook of the Economics of Education, Volume 5: 653-743, Elsevier, New York, NY.

[8] 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.

[9] Mbiti, I., K. Muralidharan, M. Romero, Y. Schipper, C. Manda, and R. Rajani (2017). "Inputs, incentives, and complementarities in primary education: Experimental evidence from Tanzania." (mimeo).

PHOTO CREDITS:
Cover: "A principal helps a student in class" by GPE/Kelley Lynch, license: CC BY-NC-ND 2.0
Page 3: "A teacher grades homework outside of class" by GPE/Kelley Lynch, license: CC BY-NC-ND 2.0
Page 5: "Students take an exam outdoors" by GPE/Kelley Lynch, license: CC BY-NC-ND 2.0
Page 7: "A principal walks through a classroom" by GPE/Kelley Lynch, license: CC BY-NC-ND 2.0

RESULTS IN EDUCATION FOR ALL CHILDREN (REACH)

REACH is funded by the Government of Norway through NORAD, the Government of the United States of America through USAID, and the Government of Germany through the Federal Ministry for Economic Cooperation and Development.

worldbank.org/reach
reach@worldbank.org

Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung

Norad

USAID
FROM THE AMERICAN PEOPLE

WORLD BANK GROUP
Education