BCES Conference Books, 2021, Volume 19. Sofia: Bulgarian Comparative Education Society
ISSN 2534-8426 (online), ISBN 978-619-7326-11-6 (online)

Boyan Zahariev & Ilko Yordanov

An Assessment of Teacher Performance in "Teach for Bulgaria" Based on Value-added from Test Scores

Abstract

The paper presents results from the evaluation of the Teach for Bulgaria (TFB) program, which is part of the Teach for All global network. TFB activities have relevance for a variety of fast-track pathways to the teaching profession. The evaluation is based on a quasi-experimental assessment of teacher performance through student value-added scores. Value added is estimated using the full datasets from successive standardized state examinations in Bulgaria after grades 4, 7 and 12. We found that TFB had most significant impact in mathematics and natural sciences which tends to be stronger in smaller schools and schools with disadvantaged students. The teacher recruitment system designed by TBF was also quite good in predicting teacher performance in mathematics and natural sciences but was not predictive of value added in humanities and language teaching.

Keywords: value added, program evaluation, teacher effectiveness

Context

The Teach for Bulgaria (TFB) program in Bulgaria started in 2010 with support from the America for Bulgaria Foundation, in order to adapt the Teach for All model and address the achievement gap among Bulgarian students.

Our analysis was conducted in 2018 but covers the first five years of TFB implementation from 2011 to 2015. In this evaluation, we use the value-added measure of student progress – from one point in time to another, based on comparable tests – to analyze the impact of the TFB program on student achievements.

The officially stated mission of TFB is "to provide every child in Bulgaria with equal access to quality education, regardless of their region, type of school, ethnic or socio-economic background" (America for Bulgaria Foundation, 2016, p. 5). TFB also declares its ambition "to raise the achievements of socio-economically disadvantaged, academically underperforming students by recruiting and selecting high-achieving graduates and highly-skilled professionals, training and supporting them to initially teach for two years in schools serving vulnerable communities and to become long-term leaders of change in our education system and society" (America for Bulgaria Foundation, 2016, p. 5).

To achieve its mission, TFB has launched a program which has adapted the Teach for All model to the Bulgarian context and has so far recruited, trained, placed and supported approximately 400 graduates and young-to-mid-career professionals in working for two years as full-time teachers in over 120 schools that predominantly serve underprivileged students. As of December 2018, over 260

participants (from 5 teaching cohorts) have already completed the program and obtained alumni status.

The Teach for All (TFA) model consists of 5 strands of interventions. The first two, which we believe represent the most distinctive feature of the program, focus on the recruitment of teachers from diverse backgrounds and their placement within schools in disadvantaged communities. The ambition of TFA is to "identify future leaders". It then provides teachers with support, which includes training and coaching. The goal is to foster classroom leadership and students' development. TFA partners in different countries, Bulgaria included, often use alternative pathways for recruitment, placement and training of teachers that differ significantly from the norms within each educational system.

Value-added models to assess teacher effectiveness

Value-added models to assess teacher productivity have become very popular in recent years – broadly used and vehemently criticized, mainly over the reliability of their basic assumptions (Sass, Semykina & Harris, 2014).

In the value-added models, the term "value added" refers to the measurement of progress between successive tests, which relies on regressing test scores on previous ones to derive expected scores and then look at the upward or downward deviation of a specific student's scores (the value added). Previous test scores appear in this model as a fixed effect; the plausible assumption is that they are correlated with subsequent test scores.

Additional independent variables of interest are included as fixed or random effects, either to measure contributing units of interest (usually teachers) on scores or to be used as controls, i.e., to calculate the coefficients of these units of interest, while also accounting the contribution of other random factors. Usually such controls are either stable or contemporaneous characteristics of students, teachers or elements of the environment – classes, schools, locations. It is assumed that previous test scores capture all previous (non-contemporaneous) factors affecting student performance: this is an assumption which is sometimes, albeit rarely, challenged (Todd & Wolpin, 2004). We base our model on this assumption, in line with most of existing research using value added. Information about previous characteristics of students and their family or school environment is very limited in availability.

Value added has been applied in the context of the Teach for All program. Value-added scores have been calculated in a variety of educational settings, including India (Azam & Kingdon, 2015), the UK (Slater, Davies & Burgess, 2012), Australia (Leigh, 2010) and Ecuador (Caridad Araujo et al., 2016), illustrating its applicability in the multiple contexts in which Teach for All operates. In Bulgaria, a similar value-added specification was used to assess the America for Bulgaria Foundation's School of the Future Program (Zahariev & Yordanov, 2016). Before that, the feasibility of applying value-added methods using Bulgarian national student assessment data was determined in an experiment piloted by the World Bank (Danchev et al., 2013).

Teacher effectiveness as measure by student value added

Teach for Bulgaria develops and implements procedures to identify future leaders, with diverse academic, professional and personal backgrounds, who will be placed as teachers for at least two years within schools in disadvantaged communities. The program deepens participants' understanding of the school system in Bulgaria and provides them with training to gain the skills and knowledge necessary to improve educational quality. Based on developed partnerships with the schools and other stakeholders, and aided by training and ongoing coaching, TFB teachers work in the selected schools and strive to foster classroom leadership and students' development.

The evaluation we carried out is quasi-experimental. Our assessment is based on a comparison between the achievements of students taught by a TFB teacher and their peers from comparable groups of schools built using propensity scores. A similar approach has been used, for example, by Chacón and Peña (2017), who also used a quasi-experimental design with difference-in-difference as a matching procedure to assess the impact of TFA's Mexican partner on students' socioemotional skills.

Considering the goals which TFB declares to pursue, namely closing gaps in education that arise from socio-economic disadvantage, it appears very logical to assess how the program has performed in terms of closing the gaps in test scores at state examinations. State examinations are the most visible sign of educational gaps, and they are increasingly being used in parallel by the Bulgarian government to guide state education policies and monitor their success or failure. In Bulgaria, state examinations in the form of standardized tests are administered after grades 4, 7 and 12. Standardized tests after grade 4 are relatively easy and have the purpose of guaranteeing that every student has covered an essential minimum requirement before proceeding to lower secondary education. Standardized tests after grade 7 are relatively difficult and are used in a process of competitive access to preferred schools and programs within upper secondary education. State examinations after grade 12 (matriculation exams) are required to obtain a diploma for secondary education and access higher education.

To assess the performance of each teacher, we have only considered examinations in fields related to the subject taught by the teacher. We have used a comparison group of teachers working in schools similar to the ones where TFB teachers were allocated. In this first period of implementation, the typical TFB school was similar to most typical Bulgarian schools. TFB teachers in the reference period taught up to about 5,000 students; the pool of students from which we sampled in the comparison group included about 200,000 students. We included schools in the comparison group based on propensity scores. The matching variables included school-level and territorial indicators related to the size of the school and the staff, the level of urbanization and the socio-economic development of the surrounding communities.

We have found evidence that the positive effects of TFB largely prevail. It shows that TFB teachers' students perform significantly better in the natural sciences and mathematics than in the social sciences and Bulgarian language and literature (BLL). All the effects we found in the natural sciences and mathematics were positive – some significantly so. This was the case for these subjects between

grades 7 and 12. A positive treatment effect also emerged in foreign languages between grades 7 and 12. Bearing in mind that teachers of disciplines such as mathematics, computer science and foreign languages tend to be in short supply in the Bulgarian educational system, we recommend that TFB continues to invest in recruiting and training such teachers and shares its experience with the Ministry of Education.

In the social sciences, we have identified one largely positive shift between grades 7 and 12, as well as one rather small negative effect which was unstable under repeated tests. For BLL between grades 4 and 7, one out of the pair of examinations showed negative value added; however, this is typical for such a school – so it could not be accounted for by the effect of TFB. Still, one of our main recommendations is to review the process of recruitment and training in the social sciences and Bulgarian language.

The positive effects in both the social sciences and natural sciences were even stronger in small schools with up to 420 students, indicating that TFB may work better in smaller and less urban communities – which also tend to be more disadvantaged in the Bulgarian context.

The TFB system of teacher assessment as a predictor of value added

As a recent overview of research on teacher selection puts it, "making decisions about selecting prospective teachers is, at its heart, making a prediction about future teacher effectiveness" (Klassen & Kim, 2019, p. 34). One of the main challenges in making the value-added analysis practically useful for schools, programs and the education system as a whole is to find good predictors of teacher performance as measured by value added. This is especially useful in the process of recruiting teachers. Knowing about performance after the fact is useful, but being able to select teachers who will deliver is much better. Of course, this does not necessarily mean that there are no better indicators of participant performance than school-level test results. We believe there are. So, the actual question we are asking here is: which teacher selection criteria are likely to increase value added? Whether to prioritize such criteria is a decision which schools, programs or educational authorities need to make.

The fact that observed teacher characteristics usually do not help much in predicting teacher productivity is a common challenge in the measurement of value added (McCaffrey et al., 2009; Azam & Kingdon, 2015). Studies that find any link between observable teacher characteristics and student performance tend to emphasize characteristics, such as teaching experience (Leigh, 2010), which cannot be reproduced in the framework of the TFB program or most programs seeking alternative pathways to teaching careers.

During the evaluated period of the program, the set of selection criteria for TFB teachers included a combination of personal characteristics such as: leadership potential and experience, academic achievements, ability to influence and motivate, organization and planning skills, attitudes towards communities, analytical thinking and perseverance.

To get some idea of the predictive power of value-added scores, we just checked for their linear correlation with the TFB system of internal teacher assessment used in the process of recruitment, training and assignment. The TFB system evolved

with the program's deployment. We have used data from an early version of the system which was used during the period covered by the current evaluation: 2010-2015. The most important indicator in the system is the overall score, which is meant to provide a weighted summary of general teacher effectiveness.

The overall score used by TFB is an especially good predictor of value added in mathematics. This predictor works best for upper secondary education, explaining 80% of variations in value-added scores in mathematics between 2010 (grade 7) and 2015 (grade 12).

In general, the scoring system used by TFB would select more teachers that are likely to demonstrate higher value added in mathematics. At the same time, the TFB scoring system selects teachers that are likely to perform worse in terms of value added in the social sciences and Bulgarian language and literature.

We do not assume that value added is the best ultimate assessment of teacher performance. Other measures of performance, like the TFB system for teacher assessment, can do a better job according to concrete tasks. We have merely tried to find out whether it is possible to predict via other prior assessment tools the value added by teachers when value added is considered important. In many cases, this is true for various stakeholders like school management, educational authorities, parents and students, among others.

Conclusion

The most notable conclusion is that TFB teachers are far more likely to have a positive effect in mathematics and the natural sciences than in the social sciences, reading and writing. Teach for America teachers do as well or better than comparison teachers at raising student achievement in math and science, and show no differences in reading (Chiang, Clark & McConnell, 2017).

The nature of BLL and humanities training presupposes a longer period of pedagogical interaction to achieve learning outcomes – especially when teaching is offered to students with a lower level of linguistic skills in the official language. Bulgarian language knowledge is a key prerequisite and a basis for the understanding of humanities-related subjects. Language is learned mainly through a gradual process of building experience, including through the child's own social experience. This social experience may be quite limited for children from vulnerable groups, and minority children do predominantly learn through their mother tongue.

Our results show much higher retention rates for students of TFB teachers. One possible interpretation of the increased retention of students is that TFB teachers may have increased the interest of students towards learning and the motivation to stay at school, even when their work did not have a direct cognitive impact measurable by standardized tests.

The typical TFB teacher is younger than the average mainstream teacher, has demonstrated high academic achievement and is able to pass a difficult selection procedure which puts both cognitive and non-cognitive skills to the test. However, they probably would need more in-depth training to understand the needs of the groups of students they work with and to create a "common language" with them. Teachers need time to translate "high" science into a pedagogy of knowledge for middle and lower level students applying student-friendly and sensitive teaching.

Presumably, most of the TFB teachers at the beginning of their work have only theoretical knowledge about children from vulnerable communities, and their social experience is quite different from that of their students. The deepening social divisions over the last three decades have undoubtedly influenced the mutual knowledge and ability to communicate effectively between students from vulnerable communities and those who are supposed to work with them. Therefore, part of the intervention time, which usually lasts for 1-2 years, is necessary to create a positive teacher-student relationship and, in this period of mutual adaptation, the effect on educational outcomes is more limited. The adjustment time needed for teaching Bulgarian language and the humanities could be even longer than for mathematics.

TFB teachers, almost without exception, either lack previous teaching experience or their pedagogical practice is too limited. In BLL and the humanities, they often need to experiment with different classroom approaches, and there is far greater diversity in these approaches than there is in mathematical teaching methods. Young teachers have to apply the principle of trial and error in the weeks and months to come and over and over again. And, in this respect, more experienced teachers have one more advantage – they have already gone down that path.

References

America for Bulgaria Foundation (2016): *Request for Proposal for Evaluation of Teach for Bulgaria*. Sofia: America for Bulgaria Foundation.

Azam, M. & Kingdon, G. G. (2015): Assessing teacher quality in India. *Journal of Development Economics*, 117, 74-83.

Caridad Araujo, M., Carneiro, P., Cruz-Aguayo, Y. & Schady, N. (2016): *Teacher Quality and Learning Outcomes in Kindergarten*. USA: Inter-American Development Bank.

Chacón, A. & Peña, P. A. (2017): *The impact of Enseña por México on student socioemotional skills*. Mexico: Microanalitica.

Chiang, H. S., Clark, M. & McConnell, S. (2017): Supplying Disadvantaged Schools with Effective Teachers: Experimental Evidence on Secondary Math Teachers from Teach for America. *Journal of Policy Analysis and Management*, 36(1), 97-125.

Danchev, P., Bankov, K., Stoimenova, V. & Atanassov, D. (2013): *Pilot implementation of statistical models for estimation of the value-added of Bulgarian schools using national student assessment data. Key results and findings*. Washington, D.C.: World Bank: Europe and Central Asia Region Human Development Unit.

Klassen, R. M. & Kim, L. (2019): Selecting Teachers and Prospective Teachers: A Meta Analysis. *Educational Research Review*, 26, 32-51.

Leigh, A. (2010): Estimating teacher effectiveness from two-year changes in students' test scores. *Economics of Education Review*, 29, 480-488.

McCaffrey, D. F., Sass, T. R., Lockwood, J. R. & Mihaly, K. (2009): The Intertemporal Variability of Teacher Effect Estimates. *Education Finance and Policy*, 4(4), 572-606.

Sass, T. R., Semykina, A. & Harris, D. N. (2014): Value-added models and the measurement of teacher productivity. *Economics of Education Review*, 38(C), 9-23.

Slater, H., Davies, N. M. & Burgess, S. (2012): Do teachers matter? Measuring the variation in teacher effectiveness in England. *Oxford Bulletin of Economics and Statistics*, 74(5), 629-645.

Todd, P. E. & Wolpin, K. I. (2004): *The Production of Cognitive Achievement in Children: Home, School and Racial Test Score Gaps (04-019), Technical report*. USA: Penn Institute for Economic Research, Department of Economics, University of Pennsylvania.

Zahariev, B. & Yordanov, I. (2016): *Impact Assessment of the School for the Future Program*. Sofia: America for Bulgaria Foundation and Open Society Institute – Sofia.

Dr. Boyan Zahariev, Open Society Institute – Sofia, Bulgaria

Mr. Ilko Yordanov, Open Society Institute – Sofia, Bulgaria