ARTICLE

Testing the feasibility of a value-added model of school quality in a low-income country

Lee Crawfurd1 | Phil Elks2

1University of Sussex and Tony Blair Institute for Global Change
2DFID

Abstract

Value-added models (VAMS) are commonly used in high-income countries for measuring the quality of teachers and schools, on the grounds that they are a fairer reflection of true quality than simple average test scores, as they account for differences in student intake. Not accounting for student's prior test scores can give a misleading impression of school quality. In this article, we adapt the current VAM of secondary school quality to the Ugandan context, and test its robustness. Using official test score data from Uganda, we test the robustness of the model to a range of different empirical specifications, including sensitivity to the inclusion of controls for student socioeconomic status. We find that the model is robust to a variety of different specifications and control variables. The VAM is low cost and has the potential to provide a clearer signal to parents, teachers, schools, and policy-makers about how much learning is actually happening in different schools. This approach could be carried out at low cost in a wide range of low-income countries that have similar testing regimes.

KEYWORDS
accountability, education, school quality, testing, Uganda, value-added

1 | INTRODUCTION

Value-added models (VAMs) are commonly used in high-income countries for measuring the quality of teachers and schools, on the grounds that they are a fairer reflection of true quality than simple...
average test scores, as they account for differences in student intake. Not accounting for students’ prior test scores can give a misleading impression of school quality.

In this article we demonstrate the feasibility of using existing official examinations data to generate an improved model of secondary school quality in a low-income country. We adapt the current UK VAM of secondary school quality to the Ugandan context, and list 25 other low- and middle-income countries for which this approach is feasible with existing data. The model is robust to a variety of different specifications and control variables, is low cost, and has the potential to provide a clearer signal to parents, teachers, schools and policy-makers about how much learning is actually happening in different schools.

VAMs have sometimes been controversial where they have required new and high frequency high-stakes exams with punitive consequences for individual teachers, for example in many states in the US. VAMs have been much less controversial where they have been built using already existing examinations data, and have been developed primarily to improve on existing published measurements of school quality, by adjusting for student intake, such as in the UK. Just looking at average test scores flatters schools that receive the best-prepared students, and penalizes schools that receive underprepared students but still manage to help them make progress.

The motivation behind generating VAMs for policy is that by better measuring performance, it is possible to put in place better incentives, motivation and support to improve performance. Teaching and learning are fundamentally concepts related to the change or growth in knowledge and skills, not their absolute level, and it is appropriate for teachers and schools to be judged on this change that they effect rather than primarily on their ability just to select high potential students.

As VAMs are based on test scores, they do not capture everything that parents and other stakeholders might care about in relation to school quality. However, looking at the change or growth in test scores is still an improvement on looking only at the level of test scores. Furthermore, many of the criticisms of VAMs, chiefly around bias and noise in estimates, actually apply to any approach to measurement of teacher or school quality, be it based on lesson observation, student evaluations or peer evaluations.

A common first policy step in publishing information on school effectiveness has been to publish average test scores. However, rankings of schools based on test scores are often little different to rankings based purely on socioeconomic status (SES)—suggesting that such a measure of school quality might be driven more by selection of pupils than the causal impact of schools. In Chile, for example, 80% of the variation in school test score rank can be explained just by SES (Mizala, Romaguera, & Urquiola, 2007). A simple first step to take account of this selection is to make an adjustment for the intake of students before making any ranking of schools—commonly by applying a variant of VAMs.

1.1 How better information can improve school performance

A key assumption underlying this approach is that generating higher quality measures of school performance (based on value-added rather than just test scores) can help improve school performance. Why might this be so? There are three main channels; first, better information can help parents hold schools accountable, either through voice or exit (choice). Second, it might help governments hold schools accountable and provide better targeted support. Third, aside from increased accountability, it might also help headteachers and teachers themselves understand better how to improve.

A systematic framework for thinking about the role of information in school quality is provided by Pritchett (2015), in an elaboration and extension of the accountability framework developed by the World Bank (2003). The 2004 World Development Report proposes a framework for understanding

---

1See Table A1 in the Online Supplementary Material
service delivery in which accountability is key, and can come either from the direct relationship between consumers and providers in a market setting (the short or direct route), or the indirect route from citizens, through the political process to the state, and from the state down to service providers (the long or indirect route). Pritchett's elaboration of this framework disaggregates each relationship of accountability into four components; delegation (the specification of the task), resources (required to carry out the task), information (to judge if the task has been completed or not), and motivation (some way in which the well-being of the agent is contingent on their performance). To achieve good learning outcomes, the hypothesis is that each of these relationships of accountability must be coherent for learning. This means that, first, within each relationship, each of the four elements must be coherent with each other. For example, within the relationship between the state and schools, learning must be delegated to schools as an objective, the state must provide schools with adequate resources to achieve learning, the state must be able to know if learning has been achieved or not, and there must be some way in which schools are motivated by ensuring learning happens. Secondly, each individual component of the accountability relationships must be coherent. For example, for the delegation component, what is being delegated to each actor must be consistent and focused on learning.

Applying this framework to a typical developing country, it seems clear that the lack of good quality information about school performance could be an important constraint to better performance, weakening all of the critical relationships of accountability. Measuring school quality using value-added may also shift the delegation component of the relationships, shifting the focus of what is delegated from targets that are focused on access and inputs, to targets that also include some focus on learning outcomes. What is also apparent from this framing of constraints is that relieving a single constraint may or may not be sufficient to improve performance depending on whether there are other more binding constraints. For example, better information alone may not lead to improved performance if teachers are still not motivated to improve their performance.

1.2 Evidence on where information has improved performance in practice

In the US, experience from a range of states and studies suggests that top-down school accountability systems based on test scores and value-added estimates have increased test scores (Figlio & Loeb, 2011), though uncertainty remains about impacts on teachers and long-term impacts on students. Similar positive impacts on test scores have been found in the UK (Burgess, Wilson, & Worth, 2013) and the Netherlands (Koning & Wiel, 2012).

In developing countries, there have been a number of experiments focused on the bottom-up client power route of accountability, by providing information to parents about school quality, with the hypothesis that better information will strengthen the client power relationship of accountability. In Uganda, two studies have estimated the effect of publishing information in newspapers about school budgets, finding positive effects on both enrolment and learning outcomes (Reinikka & Svensson, 2011; Björkman, 2007). Similarly, Björkman and Svensson (2009) found that encouraging communities to be more involved in holding health providers to account improved health outcomes in Uganda.

Studies in India that relied on active parent involvement in school management have had more mixed results (Banerjee, Banerji, Duflo, Glennerster, & Khemani, 2010; Pandey, Goyal, & Sundararaman, 2008). By contrast an experiment in Pakistan providing information to parents saw test scores increase by 0.10 standard deviations and fees reduced by 18% (Andrabi, Das, & Khwaja, 2017). In another experiment in Uganda, Barr Mugisha, Serneels, and Zeitlin (2012) found that providing information to community school management committees was only effective when they also addressed the local collective action problem. Cerdan-Infantes and Filmer (2015) compare different modes of information dissemination to parents in Indonesia, finding that “low-intensity” modes, such
as a simple letter or pamphlet, have no impact, but higher intensity modes, such as physical meetings or personalized SMS messages, did improve parents’ knowledge. In the US, Hastings, Van Weelden, and Weinstein (2007) find that providing clearer information on school quality to low-income parents leads to their choosing better quality schools for their children. Lieberman, Posner, and Tsai (2014) find null results from providing feedback to parents on their children's results, and explain this in the context of a causal chain required for information to lead to action, suggesting that the chain broke down in this case because the information was not entirely new, and was presented in a way that made performance not seem as bad as it was.

Some authors have highlighted the heterogeneity in response by parents to information in their school choice decisions, depending on their social class, cultural capital and values (Gewirtz, Ball, & Bowe, 1997). Mizala and Urquiola (2013) study the effects of being identified as a high-quality school in Chile (based on an index including both raw test scores and value-added measures, as well as other factors), finding no impact after two or four years on enrolment, tuition fees or socioeconomic composition. De Hoyos, Garcia-Moreno, and Patrinos (2015) have looked at the effect of giving diagnostic feedback to schools in Mexico. They provided detailed breakdowns of standardized test results to schools, finding substantial (0.12 standard deviations) impacts on test scores.

1.3  |  Weaknesses and criticisms of value-added models

A key concern in the use of value-added estimates for the ranking of schools is the reliability of the estimates for individual schools, which may be of particular concern for schools with a small number of students (and therefore a small sample size). Kane and Staiger (2002) study this problem in the US where VAMs have been used extensively in teacher and school accountability systems, highlighting several issues with the application of value-added measures. First, where measures are used to target incentives at the very top and bottom of the distribution, care needs to be taken where small schools are found in these groups, which may be more likely to occur through sampling variation. Second, annual test scores can be unreliable for identifying best practice or fastest improvement at the school level, and care needs to be taken in separating signal from noise. Kane and Staiger (2002) find that a simple averaging of test scores across years substantially reduces the sampling variation in estimated school effects. Third, more complex contextual VAMs which adjust for student poverty and other factors have been criticized for a range of technical and policy concerns (Gorard, 2010). These criticisms include the greater propensity for non-random missing data the more data is required, the lack of transparency to the public of a complex regression model, and the incentive effect of lowering targets and therefore expectations of pupils with high levels of poverty. The primary model we propose in this paper is a simple model without contextual factors, and so unaffected by these concerns. Other relevant technical concerns raised by Gorard (2010) include data quality issues (which can be addressed through building in quality assurance checks) and the comparability of different subjects (which does warrant consideration). Other relevant policy concerns include the risk that high-stakes tests for schools encourage teaching to the test (which is only a concern if the test is low quality, and perhaps a second order concern relative to the current lack of focus on results in many low-income contexts), and that value-added measures are always relative, meaning that some schools will always perform high or low no matter the absolute level of the system (this point is correct, but does not invalidate the value of a relative measure for distinguishing between schools).
1.4 | VAMs are possible in many developing countries using existing data

Drawing on a list compiled by Cheng and Gale (2014), we have identified 25 low- and middle-income countries\(^2\) that already have in place standardized national examinations at both primary and secondary level—the data necessary to generate a VAM for secondary school performance, based on the difference between actual secondary exam performance and performance predicted by primary scores (see Appendix in the Online Supplementary Material). This is likely to be an underestimate of the number of countries for which this approach is feasible, and the cost is low, given that the actual testing is already taking place, and all that is required is the linking of individual student’s primary and secondary test scores in a single database.

1.4.1 | Uganda context

In the remainder of this article we demonstrate with the case of Uganda the feasibility of applying in a developing country a standard VAM of secondary school quality, of the kind used for accountability purposes in the UK.

At present in Uganda, schools are judged primarily on final student exam scores, without accounting for the ability of children when they entered the school. However, students’ primary exam scores explain nearly half (48%) of the variation between students in secondary exam scores. Therefore, only looking at final secondary exam scores penalizes schools whose students start with weak results but make strong progress.

Current results reported by the newspapers in Uganda rank schools by the percentage of candidates in each overall division (aggregated across all subjects), and the percentage of candidates within each subject who receive a distinction, credit or pass. One of the key challenges in developing a VAM is reporting results which can be as easily understood by parents as such simple measures.

1.4.2 | Data

The data used here come from two sources; first, we have obtained the complete national results for all students who sat the 2015 Uganda Certificate of Education (UCE) exam, linked to their individual Primary Leaving Exam (PLE) score. Second, in order to test the stability of school value-added estimates over time, we collected similar data from 2014, plus contextual information about the school for a nationally representative sample of 335 secondary schools from 36 of Uganda’s 112 districts (plus results for 2013 from a further sub-sample of these schools).

The 2014 school survey was carried out with a sample of schools stratified across Uganda’s four regions and across school type (public and private). A total of 36 districts (of 112 districts in Uganda) were randomly sampled; 10 from each of the Central, Western and Eastern regions, and six from the less populated Northern region. Within each district 10 schools were randomly sampled, of which four were government schools and six private schools.

In each school, three instruments were implemented; first, secondary school leaving UCE scores from each student who sat exams in 2014, including aggregate score, division, and English, Maths and Science results. In addition, the corresponding PLE scores were collected for each student, along with their gender and the year they enrolled in the school. Second, a short questionnaire was carried out with the school headteacher to collect some contextual information about school resources, management and teachers. Third, a student SES survey was carried out of students currently in the fourth grade of

\(^2\)See Table A2 in the Online Supplementary Material.
secondary school (S4). This data is not linked to the individual UCE test score results, and instead just gives an indication of school-average SES. In addition, for 153 of the schools, data was collected on 2013 UCE scores and corresponding PLE scores in order to test the stability of the model over time.

2 | ESTIMATING SCHOOL QUALITY

The VAM we use is based on the UK “Progress 8” Model (Department for Education, 2015, 2012; Paterson, 2013; Burgess & Thomson, 2013). This approach is designed to balance statistical robustness with a relatively simple approach that is easily understandable by school leaders and teachers.

The model is estimated in two steps: first, individual student-level value-added is calculated. This is the difference between a student’s predicted test score (predicted based on their prior scores) and their actual test score. The second step is then calculating school-level value added, which is done by taking the average of individual student’s value-added scores (sometimes with an additional adjustment for schools with a small number of students).

Several approaches have been used to estimating a student’s “predicted” or “expected” secondary school test scores, based on their primary school test scores. The most common approach uses regression-based statistical modelling, in which a student’s “value-added” is the residual in a regression of their final test score on their prior test score. In the UK a simpler approach has been taken which provides very similar results to the more complex regression approach, but by its simplicity is more easily understandable and communicated to non-technical audiences. The simple approach predicts each student’s secondary school exam score by taking the national average (mean) of the secondary exam score of all students with the same primary exam score (Burgess & Thomson, 2013).

Following Burgess & Thomson (2013), here we provide a comparison of the statistical properties of this simple averaging approach with the more complex regression-based approaches using the Ugandan data. In the standard ordinary least squares (OLS) regression models, student test scores are estimated as a function of prior test scores and any other variables of interest, using statistical analysis software such as Stata. A key advantage of the simple averaging model is that it can be calculated and updated using simpler software such as Microsoft Excel. The more sophisticated regression approach allows for the inclusion of additional control variables more easily, and so this approach is used in this article to test the importance of different approaches and specifications.

2.1 | Estimating student value-added

The simplest approach to averaging UCE scores is to take the average for each individual PLE score aggregate.3 Below, we first compare student estimates and student ranks from the simple model with a bivariate OLS regression model, and a multivariate OLS regression model using individual subjects and interactions, finding high correlations of above 0.98.4

The simple model is equivalent to an OLS piecewise regression model, which for convenience of comparison provides r-squared and root mean square error statistics. We next present a comparison of

---

3 There are a number of desirable features of regression residuals or student value-added estimates, such that normal tests of statistical significance are applicable (Burgess & Thomson, 2013). We discuss these in Appendix 2 in the Online Supplementary Material.

4 For details see Table A3 in the Online Supplementary Material.
diagnostic statistics across a range of more complex statistical models, including those controlling for student SES, finding only a very marginal improvement in fit over the simplest model.5

2.1.1 How much are student estimates biased by not controlling for socioeconomic status?

A key concern in the development of school quality measures based on test scores is that these scores may be influenced by student family background more than school-related factors. Using a simple VAM based only on prior test scores may omit other important determinants of test score gains. Although controlling for prior test scores in the value-added framework removes any time-invariant unobserved differences between students, SES may be related not just to prior test scores, but to the ability to learn during secondary school as well.

To some extent this concern may be lesser in lower income countries—though income inequality is higher in Uganda (Gini index of 44.6 in 2012, compared to 38 in the UK in 2010, according to the most recent available statistics for each from the World Bank World Development Indicators), as there is much lower secondary school enrolment (27%), income inequality within the school system is likely to be lower.

A related concern here is the degree of school segregation. The OECD (2010) calculates an “index of social inclusion” which expresses the degree to which socioeconomic inequality within schools reflects socioeconomic inequality in the country as a whole. OECD countries with above average levels of school inclusion also have good average test scores and high levels of academic inclusion. The index value for Uganda is 61.5,6 (where 100 is perfect inclusion and 0 is perfect segregation), substantially lower than the OECD average of 74.8. Though lower inclusion may have negative educational consequences, it is helpful for this research, in which only school-average SES was collected, as it implies that school-average SES will be predictive of an individual student’s SES, and so controlling for school-average SES will be informative.

Table 1 below shows that school-average SES is indeed related to student prior test scores (this table reverts to use the OLS regression approach for ease of adding additional explanatory variables and assessing the improvement in model fit). SES is also related to student secondary test scores. However, after controlling for prior test scores and looking at student value-added, the effect of SES falls by around two thirds.

For value-added, including SES does increase the regression model fit (r-squared) marginally, from 0.46 (column 5 Table 1 above) to 0.475 (column 6). Adding individual PLE subjects and the square of the PLE aggregate (to capture nonlinear effects), raises the r-squared to 0.49. Overall this improvement in fit seems relatively minor for the loss of simplicity in explanation. Further, the coefficient of PLE scores does not change substantially after adding school SES and location, suggesting that the predictive power of PLE is independent of these variables. The impact of including these variables on the ranking of schools is explored in the next section. On average the model without controls for SES performs very similarly to the model with controls—the value-added

---

5For details see Table A4 in the Online Supplementary Material. Not reported here, we have also tested a range of weighting of UCE and PLE test scores in the model—finding that differential weighting of different subjects at either level has little effect on the overall fit of the model.

6The index of social inclusion is calculated as 100*(1-rho), where rho is the proportion of variance in SES that is between schools (as a proportion of all variance between students). The student-level index of SES used in this article for Uganda is not exactly comparable to that used by the OECD, but it is comprised of very similar components—household assets and parental education.
estimates from the two models have a correlation of 0.986 (Table A4 in the Online Supplementary Material).

It is also more difficult to try and control for these factors in the simple model without using regression analysis. One way of doing this would be to show the predicted (mean) UCE score for each PLE score broken down by SES groups (here using the proxy of the school level SES for all the students within each school). On average, students from a school in the top third (highest school-average SES) score 3.55 points better than students in the bottom third with the same PLE score (Figure A1 in the Online Supplementary Material). For comparison, the difference between cut-off points for divisions 1, 2 and 3, is 13 points, and the standard deviation of UCE scores is 12 points. This, then, is not a negligible difference, but also not huge, and would significantly complicate the calculation of these statistics by schools and other stakeholders.

There are several disadvantages to proposing a model incorporating SES for policy use at all: new data about student background would need to be collected, if this was not done robustly then schools might also have an incentive to manipulate results. This could also, in effect, reduce expectations for schools with more disadvantaged students and be seen to embed inequalities.

If exam scores are strongly correlated with SES, there is a risk that VAMs of school quality will be biased. It is clear from Table 1 above that student test scores are correlated with school SES, even after controlling for PLE scores, gender and school location. The question remains what impact this has upon schools’ scoring—which is explored in the following section.

### 2.1.2 Calculating school value-added

The second step in calculating school value-added is taking a simple average of student value-added scores by school. We first show that there is a very high correlation between school value-added scores and ranks calculated using the simple averaging method, using a Bayesian shrinkage estimator, and using a regression-based approach (Table A6 in the Online Supplementary Material).

---

**Table 1** Socioeconomic Status and Student Test Scores

|                   | PLE   | PLE   | UCE   | UCE   | VA    | VA    |
|-------------------|-------|-------|-------|-------|-------|-------|
| Socioeconomic Status | 0.302*** | 0.324*** | 0.298*** | 0.303*** | 0.085*** | 0.107*** |
|                   | (0.035) | (0.055) | (0.039) | (0.065) | (0.020) | (0.036) |
| Controls          | No    | Yes   | No    | Yes   | No    | Yes   |
| N (Students)      | 58,648 | 58,615 | 57,895 | 57,862 | 57,895 | 57,862 |
| N (Schools)       | 343   | 342   | 343   | 342   | 343   | 342   |
| r2                | 0.101 | 0.256 | 0.098 | 0.253 | 0.475 | 0.517 |

|                   |       |       |       |       |       |       |
|                   |       |       |       |       |       |       |

Standard errors clustered at school level in parentheses. Controls include student gender and school location, school size, and school type

*** p<0.01, ** p<0.05, * p<0.1

---

It is common to apply a “shrinkage factor” which adjusts value-added estimates for schools based on the number of students they have, in order to remove some of the sampling noise for those schools with particularly small numbers of students (Department for Education, 2012). In practice, this adjustment seems to have little impact on most schools in our sample, in line with results using simulated data on teachers (Guarino, Maxfield, Reckase, Thompson, & Wooldridge, 2015). We therefore treat this as a robustness check, but prefer the simple (non-shrunk) average of student value-added estimate for school value-added.
A key question is the impact of moving from a UCE-based ranking to a VA-based ranking. As Table 2 below shows, the majority (71.4%) of schools in the top quartile based on UCE are also in the top quartile based on VA, but this leaves a significant number of schools that are placed differently.

### 2.2 Confidence intervals and uncertainty

Here we present estimates of school value-added with 95% confidence intervals (CI) to demonstrate the degree of uncertainty attached to estimates of school quality. CIs are calculated in a standard manner as a function of the variation in estimates, adjusted for the number of students attending each school (schools with more students have more precise estimates of value-added and therefore smaller CIs).

High/Low CI = School VA +/- (1.96 x (Standard deviation of School VA / square root (number of students)))

110 schools are statistically significantly above average value-added, and 74 schools are statistically significantly below average. The remaining schools cannot be statistically distinguished from zero (see Figure 1). This finding, that it is possible to distinguish around half of schools from average, is in line with other value-added analysis of school performance.

| TABLE 2 | School Rank Quartile Based on UCE Aggregate vs. VA |
|---------|-----------------------------------------------|
| 4 Quartiles of VAM rank | 4 Quartiles of UCE rank |
| | 1 (Top) | 2 | 3 | 4 (Bottom) | Total |
| 1 (Top) | 72.6 | 27.7 | 0.0 | 0.0 | 100.0 |
| 2 | 19.0 | 44.6 | 34.5 | 1.2 | 100.0 |
| 3 | 7.1 | 18.1 | 44.0 | 31.3 | 100.0 |
| 4 (Bottom) | 1.2 | 9.6 | 21.4 | 67.5 | 100.0 |
| Total | 100.0 | 100.0 | 100.0 | 100.0 |

**FIGURE 1** Distribution of School Value-Added [Colour figure can be viewed at wileyonlinelibrary.com]
2.3 How much are school estimates biased by not controlling for socioeconomic status?

The previous section showed that SES is correlated with individual student UCE test scores, but that adding this SES variable into the model alongside prior PLE scores only increases the explanatory power of the model by a small amount. At the school level, there is also a correlation between school-average SES and value-added (0.30). However, the relationship between SES and test scores is weaker when accounting for prior test scores using the VAM than it is for raw average UCE test scores. To see this, note that in Table 3 the coefficient of SES on UCE scores is 2.45 without PLE scores and 0.44 with PLE (implying that a 1 point increase in SES is associated with 0.44 point increase in value-added, compared to a 2.45 increase in test scores).

By controlling for prior student attainment, we are also controlling for prior student SES, so it is unsurprising that the correlation is weaker with value-added than raw test scores. The remaining correlation between school-average SES and value-added does not necessarily imply any bias in the model, but does suggest that there may be some school-level factors that are correlated with school quality that this school-level measure picks up. A simple regression of school value-added on school socioeconomic status suggests that SES alone can explain 9% of the variation in VA (Table 3). Though the VAM does not entirely eradicate the influence of SES on estimated school quality, it does represent a substantial improvement on just UCE test scores.

In terms of individual schools, the average difference between school VA scores with and without SES is small - 1.1 points (using a regression approach). The correlation between school rank in models with and without SES is 0.90.

Looking at rank quartiles of schools, 82% of those in the top quartile after controlling for SES are in the top quartile without controlling for SES, while 18% slip down into the next quartile. For those in the bottom quartile, the majority (75.6%) remain in the bottom quartile if we do not control for SES, and 25% fall into higher categories (Table 4).

2.4 Stability over time

There is a high correlation of 0.94 between school value-added estimates between 2015 and 2014. Some variation over time will be expected, due in part to real changes in teaching quality, partly to different student cohorts, and partly to some random variation.

The stability of school value-added scores is lower than the stability in raw average test scores, consistent with experience in the US. Kane and Staiger (2002) document the substantially higher persistence of test score levels (77%) than value-added (41%). This is because test score levels combine

| TABLE 3 | Correlation Between School-Average Test Scores and SES |
|----------|-----------------------------------------------|
|          | UCE   | UCE VA          |
| PLE      |       | 1.806***        |
|          |       | (0.0746)        |
| SES      | 2.453*** | 0.437***  |
|          | (0.223) | (0.157)         |
| Observations | 331  | 331  |
| R-squared | 0.270 | 0.738 | 0.084 |

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Electronic copy available at: https://ssrn.com/abstract=3594727
a noisy measure of school quality, plus a measure of student intake, including SES and other home factors, which is highly persistent. The value-added measure is the same noisy measure of school quality, after stripping out the highly persistent measure of school intake, resulting in a lower overall persistence across time. Stability of PLE scores in Uganda is 65%. This may be problematic for the use of the VAM for ranking schools. Kane and Staiger (2001) find substantial improvements in the future predictive power of school value-added estimates after averaging performance across two to three years.

Considering a similar issue, the UK Department for Education decided that VA scores in one individual year will still be used as the headline indicator of school performance, and so will drive decisions around which schools to inspect as potential underperformers. A concern was that three-year rolling averages might mask a decline in standards in a school and slow down necessary intervention. However, three-year rolling averages will still be made available in performance tables, to facilitate more nuanced discussions and judgments about school performance.

Some difference in estimated school quality is to be expected, due to actual real changes in school quality and due to estimation error. The magnitude of stability in teacher value-added has been shown to vary substantially depending on different modelling assumptions, and the stability found here is very much within the normal range (Koedel, Mihaly, & Rockoff, 2015). A key point here is that, even though there is some variability in estimated quality from year to year, each single year’s estimate of quality has been shown to be highly correlated with long-run permanent quality.

The stability in estimates for individual schools can be expressed by the transition table between school rank in 2013 and 2014. While a majority of schools in the top (56%) and bottom (60%) quartiles in 2013 remain in the top/bottom in 2014, this leaves a large proportion of schools that move categories (Table 5).

### TABLE 4  Impact of Controlling for Socioeconomic Status on School Categorization

| VAM Rank Quartile | VAM Rank Quartile (Controlling for SES) |
|-------------------|-----------------------------------------|
|                   | 1 (Top) 2 3 4 (Bottom) Total            |
| 1 (Top)           | 81.9 18.1 1.2 0.0 100.0                 |
| 2                 | 18.1 57.8 15.7 7.3 100.0                 |
| 3                 | 0.0 24.1 59.0 17.1 100.0                 |
| 4 (Bottom)        | 0.0 0.0 24.1 75.6 100.0                 |
| Total             | 100.0 100.0 100.0 100.0 100.0            |

### TABLE 5  Change in School Rank Quartile Between 2013 and 2014

| VA Rank (2013) | VA Rank (2014) |
|----------------|----------------|
|                | 1 (Top) 2 3 4 (Bottom) |
| 1 (Top)        | 56.4 18.4 15.8 2.7 |
| 2              | 33.3 39.5 21.1 13.5 |
| 3              | 7.7 28.9 28.9 24.3 |
| 4 (Bottom)     | 2.6 13.2 34.2 59.5 |
| Total          | 100.0 100.0 100.0 100.0 |
Further robustness checks

Late joiners

One possible confounding factor is the influence of some students who join a school in the last year before UCE tests just to take the test, anecdotally a common occurrence, for example where students are felt not to be on course for success at a previous school and so are encouraged to move on, or where parents save to pay higher fees for one year of schooling just before exams. It can be argued that it is unfair to judge schools on the performance of these students, since they have only taught the student for a short period of time. Overall 8% of students only joined their school in 2014, and 6% in 2013. Excluding these late joiners has very little impact on the overall fit of the model at student level. There is no statistically significant difference between the value-added scores for individual students who joined their school in different years.

Dropouts

Another important issue is dropouts. The 2014 Ark school survey found that 19% of students enrolled in S3 do not take S4. These S3 to S4 dropout rates are highest for government non-Universal Secondary Education (USE) schools (23%) and lowest for private non-USE schools (17%), with USE schools falling in between (18% in government and 20% in private). There is no systematic correlation with school-average SES.\(^8\)

Dropout is a concern for the estimation of VAMs, as some schools may be more likely than others to encourage potentially low value-added students to drop out before taking the UCE, artificially inflating their value-added scores. This is particularly a problem if low-performing schools have higher rates of dropouts, and if dropouts would have scored poorly even after controlling for their PLE scores. Unfortunately, we do not have data on the PLE scores of children who have dropped out, so cannot test if they are systematically different. Exploring the role of dropouts in school value-added, perhaps by capturing PLE data for students who then drop out, should be a priority for future data collection and analysis. This concern is assuaged to some extent by the focus on value-added rather than just UCE scores. Simply encouraging students with poor expected UCE scores to drop out might artificially inflate average UCE scores, but would not inflate value-added unless these students were expected to have poor UCE scores even conditional on their (poor) PLE scores.

Of some reassurance is that there does not appear to be any overall systematic relationship between the percentage of S3 students who dropped out or did not take UCE for another reason, and school VA. However, this does not rule out that some schools might still engage in some strategic behaviour by trying to select for high potential S4 students. Of particular concern might be the minority of schools for which dropouts are more than 50% of S3 students.

\(^8\)This dropout rate is similar to that implied by national secondary enrolment rates by grade. While there is some dropout between S1 and S3, by far the largest numbers drop out between S3 and S4 (see Appendix 1).

### Table 6 Pros and Cons of Alternative VAM Models

|                         | Model fit (R-squared) | Simplicity | Cost |
|-------------------------|-----------------------|------------|------|
| Simplest Model          | Good                  | Best       | Low  |
| Regression Model with admin data | Better               | Worse      | Low  |
| Regression Model with survey data | Best                | Worse      | High |
2.5.3 | Missing data

Some schools were missing records for some students for their PLE scores. To investigate the importance of this we compare mean UCE scores for students with and without PLE scores. There is a very slight difference between UCE aggregate scores between the two groups, but no statistically significant difference in UCE divisions (Table A9 in the Online Supplementary Material). In addition, some schools initially sampled were unable to be surveyed as they no longer existed or refused to participate, and so replacement schools were sampled. There is no statistically significant difference between test scores or value-added of students in “replacement” schools versus schools which were in the original sample (Table A10 in the Online Supplementary Material).

3 | CONCLUSION

This article has demonstrated the feasibility of applying standard VAMs of secondary school quality in a developing country context using existing official test score data. Such models have the potential to provide a clearer signal to parents, teachers, schools and policy-makers about how much learning is actually happening in different schools, and at relatively low cost. The model applied here provides an incentive for schools to teach all students—above average performance made by any student counts equally, not just those at the top end of the distribution. Linking student test scores also provides a high-quality public good for research purposes, allowing for low-cost quasi-experimental research (such as Crawfurd (2017)).

This approach could be used in a wide range of other developing countries where students are already tested at the end of primary school and the end of secondary school—we count at least 25 such countries. We are not here proposing any new standardized examinations with high stakes for students, just the use of data produced by existing examinations, addressing any possible concerns of increasing pressure on students. Where high-stakes accountability systems are used in high-income countries for schools, a range of concerns are commonly raised which should be considered in any expansion of the use of VAMs in developing countries. Care needs to be taken to ensure that strategic behaviour by schools does not lead to adverse unexpected consequences, such as pupil exclusions, cheating, teacher stress or excessively teaching to the test.

A number of technical issues would benefit from further investigation. First, the stability of estimates across time is not perfect—averaging data from several years might be a preferable option to relying on school VA estimates from a single year, which is vulnerable to sampling variation. Second, better tracking of students who register at a school but drop out before taking the UCE exam would allow for further investigation of whether this issue biases school quality estimates (and if so, how important this bias is). Third, the reliability of the tests themselves and the consistency with which they are sat and graded across schools warrants further investigation. Fourth, the issue of whether introducing this performance measure could drive strategic behaviour from schools. For example, at present students absent for the UCE test have been dropped from the analysis. Their inclusion or exclusion does not seem to substantially affect school level value-added estimates, but further analysis may be warranted if this aspect of the model was to be taken forward in a “live” setting. Similarly, this analysis suggests that the number of students dropping out in the last year before UCE is not correlated with school VA scores, but care would need to be taken to ensure that schools would not perceive a benefit from particular students leaving the school in advance of the exams and act on this perception. Fifth, there is a similar decision to be made on the treatment of results on specific categories of schools. For example, one school in the sample for this study was found to be cheating in the exams, and so all UCE
results for the cohort were annulled. This school has been omitted from this analysis, which does not affect the ranking of other schools, but a decision may be needed on how this school should be ranked relative to other schools. Similarly, a decision would also need to be taken on how to present the results of very small schools.

Sixth, in terms of choice of modelling approach to estimating predicted scores for individual students, there is a trade-off to be made between statistical model fit, simplicity in terms of explanation to policy-makers and stakeholders, and cost. The simple model is the easiest and most transparent to explain, is low cost and achieves a good model fit, only slightly worse than a regression-based approach, which achieves a slightly better statistical fit at the cost of a relatively opaque (to non-statisticians) approach—OLS. The best fit model is to fit a regression model using contextual data, for instance on student background, but, as well as being complex, such a model would have substantial financial costs in terms of new survey work (beyond collecting existing test scores), which seems like an excessive cost for a relatively small improvement in precision.

ACKNOWLEDGEMENTS

The authors would like to thank the Economic and Social Research Council for funding (Award Reference ES/J500173/1).

The data and questionnaire used in this article have been submitted to the UK Data Archive which can be found at https://discover.ukdataservice.ac.uk/

REFERENCES

Andrabi, T., Das, J., & Khwaja, A. I. (2017). Report cards: The impact of providing test-score information on educational markets. *American Economic Review*, 107, 1535–1563. https://doi.org/10.1257/aer.20140774
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
Barr, A., Mugisha, F., Sernerks, P., & Zeitlin, A. (2012). *Information and collective action in community-based monitoring of schools: Field and lab experimental evidence from Uganda*. Retrieved from https://www.tilburguniversity.edu/upload/397359d2-09a0-40d5-9502-8c3592aaae40_zeitlin.pdf
Björkman, M. (2007). *Does money matter for student performance? Evidence from a grant program in Uganda* (IGIER Working Paper No. 326). Retrieved from the Innocenzo Gasparini Institute for Economic Research website: http://www.igier.unibocconi.it/folder.php?vedi=3957&tnb=albero&id_folder=4878
Björkman, M., & Svensson, J. (2009). Power to the people: Evidence from a randomized field experiment on community-based monitoring in Uganda. *The Quarterly Journal of Economics*, 124, 735–769. https://doi.org/10.1162/qjec.2009.124.2.735
Burgess, S., & Thomson, D. (2013). *Key Stage 4 accountability: Progress measure and intervention trigger*. Retrieved from http://www.bris.ac.uk/media-library/sites/cubec/migrated/documents/report11.pdf
Burgess, S., Wilson, D., & Worth, J. (2013). A natural experiment in school accountability: The impact of school performance information on pupil progress. *Journal of Public Economics*, 106, 57–67. https://doi.org/10.1016/j.jpubeco.2013.06.005
Cerdan-Infantes, P., & Filmer, D. P. (2015). *Information, knowledge and behavior: Evaluating alternative methods of delivering school information to parents* (Policy Research Working Paper No. 7233). Retrieved from World Bank website: http://documents.worldbank.org/curated/en/488221467999136216/Information-knowledge-and-behavior-evaluating-alternative-methods-of-delivering-school-information-to-parents
Cheng, X., & Gale, C. (2014). List of national learning assessments [Spreadsheet]. Retrieved from the Education Policy and Data Center website: https://www.epdc.org/education-data-research/list-national-learning-assessments

Electronic copy available at: https://ssrn.com/abstract=3594727
Crawfurd, L. (2017). School management and public–private partnerships in Uganda. *Journal of African Economies*, 26, 539–560. https://doi.org/10.1093/jae/ejx021

de Hoyos, R., Garcia-Moreno, V. A., & Patrinos, H. A. (2015). *The impact of an accountability intervention with diagnostic feedback* (Policy Research Working Paper No. 7393). Retrieved from World Bank website: http://documents.worldbank.org/curated/en/297561468188928817/The-impact-of-an-accountability-intervention-with-diagnostic-feedback-evidence-from-Mexico

Department for Education. (2012). *A guide to value added key stage 2 to 4 in 2012 school and college performance tables & RAISE Online*. London: Department of Education.

Department for Education. (2015). *Progress 8 measure in 2016 and 2017: Guide for maintained secondary schools, academies and free schools*. Retrieved from https://dera.ioe.ac.uk/22181/

Figlio, D., & Loeb, S. (2011). School accountability, In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the economics of education* (Vol. 3, pp. 383–421). Amsterdam: Elsevier.

Gewirtz, S., Ball, S. J., & Bowe, R. (1997). *Markets, choice and equity in education*. London: Open University Press.

Gorard, S. (2010). Serious doubts about school effectiveness. *British Educational Research Journal*, 36, 745–766. https://doi.org/10.1080/01411920903144251

Guarino, C. M., Maxfield, M., Reckase, M. D., Thompson, P. N., & Wooldridge, J. M. (2015). An evaluation of empirical Bayes’s estimation of value-added teacher performance measures. *Journal of Educational and Behavioral Statistics*, 40, 190–222. https://doi.org/10.3102/1076998615574771

Hastings, J. S., Van Weelden, R., & Weinstein, J. (2007). *Preferences, information, and parental choice behavior in public school choice* (NBER Working Paper No. 12995). Retrieved from the National Bureau of Economic Research website: https://www.nber.org/papers/w12995

Kane, T. J., & Staiger, D. O. (2001). *Improving school accountability measures* (Working Paper No. 8156). Retrieved from the National Bureau of Economic Research website: https://www.nber.org/papers/w8156

Kane, T., & Staiger, D. O. (2002). The promise and pitfalls of using imprecise school accountability measures. *Journal of Economic Perspectives*, 16(4), 91–114. https://doi.org/10.1257/0895300202950093

Koedel, C., Mihaly, K., & Rockoff, J. E. (2015). Value-added modeling: A review. *Economics of Education Review*, 47, 180–195. https://doi.org/10.1016/j.econedurev.2015.01.006

Koning, P., & van der Wiel, K. (2012). School responsiveness to quality rankings: An empirical analysis of secondary education in the Netherlands. *De Economist*, 160, 339–355. https://doi.org/10.1007/s10645-012-9194-9

Lieberman, E. S., Posner, D. N., & Tsai, L. L. (2014). Does information lead to more active citizenship? *World Development*, 60, 69–83. https://doi.org/10.1016/j.worlddevel.2014.03.014

Mizala, A., Romaguera, P., & Urquiola, M. (2007). Socioeconomic status or noise? Tradeoffs in the generation of school quality information. *Journal of Development Economics*, 84, 61–75. https://doi.org/10.1016/j.jdeveco.2006.09.003

Mizala, A., & Urquiola, M. (2013). School markets: The impact of information approximating schools’ effectiveness. *Journal of Development Economics*, 103, 313–335. https://doi.org/10.1016/j.jdeveco.2013.03.003

OECD. (2010). *PISA 2009 Results: Overcoming Social Background*. PISA: OECD Publishing.

Pandey, P., Goyal, S., & Sundararaman, V. (2008). *Community participation in public schools: The Impact of information campaigns in three Indian states* (Policy Research Working Paper Series No. 4776). Retrieved from World Bank website: http://documents.worldbank.org/curated/en/735071468051274602/Community-participation-in-public-Schools-the-impact-of-information-campaigns-in-three-indian-states

Paterson, C. (2013). *Measuring what matters: Secondary school accountability indicators that benefit all*. Retrieved from https://www.bl.uk/collection-items/measuring-what-matters-secondary-school-accountability-indicators-that-benefit-all

Pritchett, L. (2015). *Creating education systems coherent for learning outcomes: Making the transition from schooling to learning* (RISE Working Paper No. 15/005). Retrieved from the Research on Improving Systems of Education website: https://www.riseprogramme.org/publications/RISE-working-paper-15005-creating-education-systems-coherent-learning-outcomes

Reinikka, R., & Svensson, J. (2011). The power of information in public services: Evidence from education in Uganda. *Journal of Public Economics*, 95, 956–966. https://doi.org/10.1016/j.jpubeco.2011.02.006

World Bank. (2003). *World development report: Making services work for poor people*. Retrieved from World Bank website: http://documents.worldbank.org/curated/en/527371468166770790/World-Development-Report-2004-Making-services-work-for-poor-people-Overview
SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Crawfurd L, Elks P. Testing the feasibility of a value-added model of school quality in a low-income country. *Dev Policy Rev*. 2019;37:470–485. https://doi.org/10.1111/dpr.12371