

## **SHORT-TERM AND LONG-TERM CONSEQUENCES OF SYSTEMS OF VALUE-ADDED MODELLING: DISCUSSION (BEN JENSEN)**

*Workshop for the Committee on Value-added Methodology for Instructional Improvement, Program Evaluation and Accountability*

1. The purpose of this brief paper is to provide comments on the two papers prepared for the session on short-term and long-term consequences for using systems of value-added modelling. I would start by taking the opportunity to thank those responsible for creating the opportunity to comment on two interesting papers that highlight a number of key issues relevant to policymakers and analysts. In providing comments the focus has been on identifying commonalities and differences across the two papers and how these impact upon consideration of short-term and long-term consequences.
2. The first paper by Ashish Jha focused on experiences in the U.S. health sector and highlights a number of pertinent issues. It is interesting to consider these in the light of their applicability to education given the similarities of the sectors but also the differences in their evaluative structures. It should also be noted that the paper focuses upon the public reporting of data rather than of value-added modelling. The second paper, by J. Douglas Willms, identified key issues in value-added modelling and how these may impact both measures of school performance and the interventions stemming from those actions. It is clear that the papers agree on most aspects of the issues they discussed. In particular, the importance of the adjustment mechanism in value-added modelling (and the difficulties that may arise) and the benefits of a broader array of information being used in an evaluative framework. It is also clear that the authors do not have strong disagreements on the key issues.
3. Both papers emphasise the difficulties of adjusting for various background factors and those outside of the control of the unit of evaluation (e.g. the school, teacher, hospital, doctor). There are difficulties in value-added modelling in picking the correct level (or variables) of adjustment for external factors. Willms cites several instances where this is a problem in education and Jha discusses a number of concerns in the health sector, particularly those stemming from changes in the adjustment mechanism that impact upon results. Jha emphasises an issue that is also common to education of the incentive this creates for selection bias. Doctors or hospitals refusing to treat patients who are less likely to survive or be successful in an operation are commensurate with the actions of schools refusing to educate low-performing students. Emphasising the difficulties of adjustment in value-added modelling, Willms discusses that adding-value has a fundamentally different meaning for educators in low SES urban public schools. Educating students at the lower end of the performance spectrum is substantially different to educating those at the higher end and that this is not reflected in value-added scores. In addition, he suggests the possibility for distinguishing between Type A (controlling for prior ability and SES but not school composition) and Type B effects (controlling for prior ability, SES, and mean SES of the school) as Type B effects better isolate the value-added by the school. But he correctly states that in doing so may 'over-adjust' and unfairly reduce the measured school value-added.
4. As a reader of these articles I share the concerns of the problems for accurate adjustment mechanisms. There is little doubt that these issues can cause problems of selection bias and unfairly discriminate against schools serving poorer or more disadvantaged communities. However, I believe that there are two further issues that need to be considered. First, is the issue of the accuracy of adjustment

methods in value-added modelling compared to other methods of adjustment. Generally, I believe that value-added modelling deals with this issue better than other techniques and this should be recognised in decisions concerning whether or not to implement such a system (see the forthcoming OECD report for a discussion of this issue<sup>1</sup>). Second, problems of accurate adjustment and the potential selection issues it brings emphasise the need for both the correct interpretation of results and also an effective system upon which value-added modelling is based. It is possible to place a greater emphasis on the gains made with disadvantaged students and to implement systems that negate particular selection problems. For example, in a system that uses value-added schools to reward specific schools or teachers, greater rewards can be provided for gains made with students from poorer backgrounds. Moreover, penalties can be given to schools that refuse or ‘encourage’ poor performing students to leave the school or to be absent on the day of the test (it has proven effective in England to penalise schools with students missing on the day of the test).

5. The second main issue that these papers share is the need to emphasise broader measures of performance. While it is not always clear how these measures would be viewed alongside value-added measures, their importance is emphasised in both papers. Willms provide examples of programmes that focus on process measures that complement value-added indicators. Jha provided some interesting examples of process indicators in the health system which appeared to have positive results. We are left to consider the applicability to education systems. Most systems would gather information on school processes through some form of school evaluation<sup>2</sup> and there are some differences to the process data collected in the health sector that could impact upon the importance of the use of value-added measures. Some of the processes in the health sector (e.g. timely prescription of medicine) lends itself to a quantitative measure much more than many education processes that often focus on an evaluative judgement of the performance of teachers (or of specific actions by teachers). This difference can be important because it affects both the need for quantitative measures (e.g. value-added measures) but also how they are used. The paper correctly asserts that in an evaluative framework information is required on both outcomes and processes as both are fallible. This is also true of education but given the qualitative nature of measures of education processes, a greater need exists for the introduction of quantitative outcome measures. This need derives from two sources. First, the qualitative process information that has dominated evaluative structures in most education systems (and has dominated much education research) can only impact actions, foster improvements, and promote accountability and choice to a limited extent without a link to outcomes. Improvements are restricted if we cannot properly evaluate the current situation or show how change improves outcomes. The second point is strongly related to the first but emphasises the quality of both qualitative and quantitative measures. Evaluative judgement based upon qualitative measures and quantitative outcome data are both prone to bias and miss-measurement. However, it can be difficult to have faith in evaluative qualitative judgement if they are not linked to outcomes as the question arises of the extent to which they are ‘evidence-based’. The quality of such judgements can only improve if the impacts upon outcomes are measured. Clearly, we can only learn from outcome measures if we analyse them alongside process measures but I believe the benefits can be greater for utilising quantitative outcome measures if an existing evaluative structure has a greater reliance upon qualitative measures.

6. Two additional issues that I should draw attention to is first the issue raised by Jha that health information was not used by patients in a manner that increased effectiveness. This has been supported by evidence in education that the main driver in parents’ choice of schools for their children remains location

---

<sup>1</sup> OECD (2008) *Measuring Improvements in Learning Outcomes: Best Practices to Assess the Value-added of Schools*, Paris: OECD.

<sup>2</sup> Note: Amongst OECD countries, school self-evaluations are required in 14 countries (with data available), normally on an annual basis. School inspections by an external body are also required in 14 countries but tend to be required once every three years or so (see OECD (2008) *Education at a Glance (Indicator D5)*, Paris: OECD.

(“it’s the local school”)<sup>3</sup>. The second is the contention by Willms in his concluding argument that value-added scores should be measured relative to a fixed-standard rather than other schools. I believe great care should be taken before advocating this approach. By focusing on a standard an incentive is created to focus on specific students of specific abilities. This can quickly lead to sub-optimal outcomes if that focus is at the expense of other areas of education. It also does not avoid the problem of different schools facing different circumstances if a single standard is applied to all schools.

7. To conclude, there are some other points of interest that could be mentioned in terms of short and long-term consequences. These are listed below:

1. *Policy objectives:* The forthcoming OECD report emphasised that the first step in implementing a system of value-added modelling was to identify the policy objectives and the use of value-added information. Virtually all consequences stem from this decision. Whether the focus is on school accountability or school improvement (particularly internal school improvement) greatly affect the likelihood of achieving long-term benefits. Of great importance is their use in terms of school or teacher accountability (an issue emphasised by Willms), school improvement, and the publication of school value-added scores.
2. *Institutional response:* This very much depends upon the current context. The introduction of school performance measures can illicit negative reactions from numerous stakeholders. However, in a system where there is already a focus on raw test scores, then there are numerous examples of stakeholder support for the introduction of a system of value-added modelling.
3. *Resourcing:* Clearly a system of value-added modelling can be a resource intensive exercise for a Government to undertake. Much of the costs rest in the student assessments so if these have to be developed then further costs arise.
4. *Sub-optimal incentives:* A great danger with all performance measures, not just value-added, is the creation of sub-optimal incentives. Schools focusing resources to improve their value-added score at the expense of other education objectives clearly negate the effectiveness of the entire education system. Detailing the likely incentives created from the system and then monitoring commensurate actions are therefore needed.
5. *Evaluation:* Does the system make a difference? Clearly this can only be evaluated in the longer term as it becomes clearer the impact upon student learning but there are various short-term aspects that can be evaluated. For example, to what extent is value-added information being used within schools to evaluate and develop their programs? Are families using the data for school choice? Has the system of school accountability improved with value-added measures?

8. I would like to finish by thanking the authors for their interesting papers. I believe they can benefit this project on the benefits of value-added methodology for instructional improvement, program evaluation, and accountability. To this end, I believe the considerations of value-added methodology should be made relative to existing practice to identify what advances can be made.

Ben Jensen

---

<sup>3</sup> OECD (2006) *Demand Sensitive Schooling? Evidence and Issues*. Paris: OECD