



# **ESTIMATING TEACHER EFFECTIVENESS?**

Roy Martin and John McCollow

May 2007

Authorised and printed on the internet by Pat Byrne, Federal President,  
Australian Education Union, Ground floor, 120 Clarendon Street, Southbank VIC 3006

**Pat Byrne**  
Federal President

**Susan Hopgood**  
Federal Secretary

**Australian Education Union**  
Ground Floor  
120 Clarendon Street  
Southbank VIC 3006

Telephone: 61 3 9693 1800  
Facsimile: 61 3 9693 1805  
E-mail: [aeu@aeufederal.org.au](mailto:aeu@aeufederal.org.au)

Recently a study conducted by Andrew Leigh of the Australian National University entitled “Estimating Teacher Effectiveness from Two-Year Changes in Students’ Test Scores” received a good deal of media attention. Leigh’s study was widely reported as providing evidence that supported proposals for “performance pay” for teachers. As will be shown, however, the outcomes of the Leigh’s research are unremarkable and fall well short of evidence that could be used to support the case for performance pay. In fact, if anything, they provide some evidence to the contrary.

## **What did Leigh’s study entail?**

Leigh used data from the Year 3, 5 and 7 literacy and numeracy tests in Queensland to track individual student performances at each year level. To eliminate the effects of socio-economic or home factors, Leigh uses each student’s Year 3 result as the base and focuses on the extent to which performance on the subsequent tests improves relative to the “average” performance of the student’s age cohort. Leigh is able to match students with the teachers that they had at each year level. He then assigns the change in relative performance (or lack thereof) to the effects of teaching, specifically to the classroom instruction provided by the teacher the student had at the time of the test. He is then able to calculate the correlation between improved test performance and various factors associated with the teacher (e.g. gender, age, years of experience, possession of a post-graduate qualification, Departmental S-rating).

## **What did Leigh find?**

There was significant variation in the “effectiveness” of teachers as measured by the improvements in test scores of their students. There was also a positive correlation between “effectiveness” as measured by student literacy results and student numeracy results. Of the factors for which correlation was calculated, teacher “experience has the strongest effect” (p.1), though “most of the differences between teachers are due to factors not captured in the payroll database” (p.4).

## **Limitations and problems with the Leigh research**

### ***Defining and measuring teacher effectiveness***

Leigh defines teacher effectiveness in terms of the performance of their students on literacy and numeracy tests. This is a relatively limited and crude measure. Furthermore, because the tests were administered at two year intervals in August, students will have experienced classroom instruction under at least three teachers in the intervening period between the tests. Leigh assigns the improvement either entirely to the instruction of the teacher at the time of the relevant test or “splits” it 50/50 between that teacher and the teacher in the previous year.

All student test results are attributed to teacher instruction – no other factors are considered (e.g. personal, family, school, community circumstances). While Leigh attempts to control for socio-economic and family factors by using each student’s own

Year 3 result as the base, this only works if one assumes these factors have no effect on learning *rates* – a highly contestable assumption.

Finally, because teacher effectiveness is measured in terms of how their students perform relative to the average performance of their year-level cohort, any improvements in student scores that align with overall improvements for the cohort are not picked up as “value-added”.

### ***Overstating the Research Results***

Leigh’s research finds that not all teachers are equally effective in terms of the results achieved by their students on literacy and numeracy tests. This can hardly be considered a startlingly new insight. In fact, as Leigh himself admits on his web log:

*For what it’s worth, I don’t think this dispersion is wider than what one would find among plumbers, dentists, architects or bricklayers. But it does indicate that - at least as measured by test score gains - all teachers are not created equal.*

Similarly, while Leigh’s research purports to show a limited connection between factors identified on a payroll database and teacher effectiveness, he identifies and tests no alternative factors that correlate more highly. The factor identified as having the greatest effect (experience) actually argues against an approach to remuneration based on performance pay.

Leigh is prone to some remarkably simplistic statements. For example, although Leigh’s study did not examine the time taken by teachers to prepare their students for the tests he states that “a teacher at the 90<sup>th</sup> percentile can achieve in half a year what a teacher at the 10<sup>th</sup> percentile can achieve in a full year” (p.11). This sort of hyperbole is clearly aimed for media consumption rather than as a contribution to scholarly debate.

### ***Implications of measuring teacher effectiveness for performance pay using Leigh’s approach***

The principle thing this research “proves” is that if one uses a very crude measure, it is possible to identify those teachers who have most “value added”. However, in order for this to be linked to a system of performance pay it would be necessary to:

- vastly increase the frequency of tests (really, it would be necessary to test on entry and exit from each year, and presumably every time there was a change of teacher);
- increase the range of tests to anything that was valued in teaching and learning (or at least to recognised subjects).

(Of course, the problems for secondary schools would be even greater).

Besides being incredibly intrusive and time consuming, this would distort the process of schooling to an incredible degree, demanding high stakes testing, high curriculum definition and low teacher trust to become the whole basis of schooling. Evidence cited by curriculum and assessment experts such as Allan Luke (2007) and Nichols &

Berliner (2007) strongly indicates that this would actually be counter productive. Experience in the USA and UK shows the deleterious effects on schools and teaching of an over-reliance on standardised tests. Some of the most successful countries (Finland and Canada) in terms of educational outcomes have low stakes testing, low definition curriculum and high levels of teacher trust.

## **Conclusion**

To express scepticism about this particular piece of research is not to argue for the status quo. Improving teacher effectiveness and recognising teacher professionalism are both worthy and compatible goals. It is important to explore how they might be best achieved and research must play a central role in this endeavour. The effort is best served by approaches that engage directly with the complexity of teaching and learning. It is rather more poorly served by approaches that promote quick fixes that owe more to a market-based ideology, political expediency and self-promotion in the media than to scholarly research.

*Roy Martin is AEU Federal Research Officer. John McCollow is QUTU Research Officer.*

## *References*

Leigh, A. (2007) "Estimating Teacher Effectiveness from Two-Year Changes in Student's Test Scores", Research School of Social Sciences, ANU, <http://econrsss.anu.edu.au/~aleigh/>

Luke, A. (2007) "Does Increased Accountability through High Stakes Testing Improve Student Outcomes?", presentation to ACSA Forum, Brisbane, 30 March.

Nichols, S.L. and Berliner, D.C. (2007), *Collateral Damage: How High-Stakes Testing Corrupts America's Schools*, Harvard Education Press, Cambridge, MA.

Watt, C. (2007) "Estimating Teacher Effectiveness", IEUA, Melbourne.