

Australian Council for Educational Research (ACER)

**ACEReSearch**

---

[ACER eNews Archive](#)

[ACER Archives](#)

---

6-2007

## ACER eNews 06 June 2007

ACER

Follow this and additional works at: [https://research.acer.edu.au/enews\\_archive](https://research.acer.edu.au/enews_archive)



Part of the [Education Commons](#)

---

### Recommended Citation

ACER, "ACER eNews 06 June 2007" (2007).  
[https://research.acer.edu.au/enews\\_archive/72](https://research.acer.edu.au/enews_archive/72)

This Article is brought to you by the ACER Archives at ACEReSearch. It has been accepted for inclusion in ACER eNews Archive by an authorized administrator of ACEReSearch. For more information, please contact [repository@acer.edu.au](mailto:repository@acer.edu.au).

Published June 2007

In this issue:

- [How not to measure teacher performance](#)
- [Year 12 aptitude test on the horizon](#)
- [Survey seeks to measure student engagement](#)
- [ACER UPDATE](#)

## How not to measure teacher performance

Most parents recognize that good teachers are worth their weight in gold. There is little debate about the need to place more value on teachers' work, for example, by providing substantial pay rises to teachers as they attain higher standards of performance. This is unlikely to happen, however, unless we become better at evaluating teacher performance in ways that are valid, reliable and fair.

A recent research study by Andrew Leigh ("Study reveals teacher skill discrepancy," *The Age*, May 21, 2007) has been widely reported as demonstrating that the best teachers can be readily identified, and have been shown to be twice as effective as the worst teachers.

*The Australian* reported that the study "has successfully linked teacher performance with student results, bolstering the Federal Government's efforts to introduce performance-based pay" (Justine Ferrari, *The Australian*, May 21). Dr Leigh stated in the conclusion to his paper that he "has shown how to estimate a measure of teacher performance by using panel data with two test scores per student." (p.17)

Readers could be forgiven for thinking that the research opens the way for the "best" teachers to be identified and rewarded on the basis of their students' test performance, as they were at the end of the nineteenth century. It does nothing of the sort.

**Published June 2007**

Dr Leigh examined the test scores of three cohorts of Queensland government school students in literacy and numeracy. Each cohort included around 30,000 students – about three-quarters of the students who actually took the tests in government schools. The focus of his study has been reported as gains in achievement over two years, but this was not so. Dr Leigh examined the changes in relative positions of classes of students within the overall state results from year 3 to year 5 (two cohorts) and from year 5 to year 7 (one cohort).

Not surprisingly, he found that some classes improved their position within the state results, while others went in the opposite direction. This, of course, was inevitable. For every class that gained in its relative position, another had to go down. This is the nature of relative data.

But what grabbed the headlines was the assertion that the classes that improved their relative position must have had the “good” teachers, while those whose relative position declined must have had the “bad” teachers. If you follow the same logic, you would conclude that Leigh Matthews was not a successful coach of the Brisbane Lions in 2002 and 2003 because he failed to improve the position of his team on the AFL ladder. They finished first in 2001, and no better than that in 2002 and 2003. Many a football fan from the southern states would have been delighted had the Lions’ management used the same reasoning and decided it was time to replace their coach.

Just as the Lions’ failure to improve their ladder position over two years did not show that their coach was a failure, Dr Leigh’s research provides no basis for the identification of effective and ineffective teachers.

First, the research was based on relative measures (where students stood in a statewide ranking). Inevitably, some students’ scores were bound to increase and others to decrease. If the literacy and numeracy tests were replaced by a coin-tossing test (“Toss this coin ten times and count the number of heads”) some students’ scores would increase, some would decrease, and some would stay the same. Some classes would go up in the rankings, some would go down, and very few would stay the same. Such is the nature of data.

**Published June 2007**

Second, we need to look at the nature of the data that were used by Dr Leigh. Students were tested in August of one year, and then retested in August two years later. In the time between tests they would have had up to three teachers: one from August to December of the first year; one from January to December of the second year, and another from January to August of the third year. If things go well, who gets the credit? If things go poorly, who gets the blame?

Dr Leigh suggested two approaches to dealing with this (insurmountable) difficulty: to ignore the intervening year altogether, or to create an assumed test score in the intervening year, which lies at the midpoint of the other two tests. He chose the second option "to maximize sample size." So students were assigned test scores for a year in which they had not taken a test, and their teachers were judged to be effective or ineffective on the basis of how well their students were assumed to have performed on this non-existent test.

Neither of the approaches considered by Dr Leigh addresses the real problem – that the data contain no valid basis for linking students' achievement growth to the performance of a single teacher. Statistical analysis, no matter how complex, cannot overcome this.

Students learn (and sometimes fail to learn) because of a multitude of factors. A good teacher is vitally important. So are supportive parents, a learning culture in the school, a community that values and rewards learning, and a school that is provided with the resources that it needs to perform its role well. And school learning comes more easily to some than to others. Research that ignores these factors fails to recognize the subtlety of schooling.

Of course some teachers are more effective than others. There is plenty of rigorous research which shows that students' achievement growth varies from class to class. It also varies significantly from family to family, neighbourhood to neighbourhood, and of course, mostly, from student to student. The impact of each of these factors cannot be assessed with the precision necessary to isolate the effect of the teacher, and to suggest otherwise is to mislead the public.

Published June 2007

Even if Dr Leigh had made better use of the Queensland student achievement data, by developing value-added measures of student achievement, the basic problem would not have gone away. Such measures are designed to discriminate between students, not individual teachers. After reviewing the literature on the use of value-added modelling (VAM) in estimating teacher effects, McCaffrey and his colleagues (2004) concluded that

. . . VAM-based rankings of teachers are highly unstable, and ... only large differences in estimated impact are likely to be detectable given the effects of sampling error and other sources of uncertainty. Interpretations of differences among teachers based on VAM estimates should be made with extreme caution. (p.113)

The reliability of value-added estimates depends, in part, on the quality of the student achievement measures that underpin them, and the margins of error in most existing measures need to be understood. While there have been significant advances in our ability to measure educational growth, we are a long way from measures with anything like the reliability of, say, measures of growth in children's weight or height.

In addition, measures available so far are limited to reading and numeracy. For other areas in the primary and secondary school curriculum there are no measures to which value-added modelling could be applied in judging teacher performance. Valid evaluations of performance need to be based on evidence that covers the full range of a teacher's responsibilities.

Nobody should be tempted to believe that the approach used in Dr Leigh's paper can be translated into a viable and legally defensible system for assessing the performance of individual teachers, as claimed.

Such a system, which we believe is overdue, would need to be based instead on a range of direct evidence of a teacher's capacity to provide quality conditions for his or her students' learning across the curriculum. These conditions must be consistent with current research and with profession-defined standards - as in any profession. You have to look directly at what students are doing and learning in classrooms to find the valid evidence of "performance" that is needed.

Published June 2007

The danger with Dr Leigh's paper, which was released on his own website without going through the standard process of peer review, is that it promises much more than it can deliver. It will be interpreted by some as evidence that there is a simple solution to the challenge of linking teachers' pay to performance.

The past 100 years is littered with merit pay schemes that failed, mainly because proponents did not do the hard work of developing standards that cover the full scope of what effective teachers know and do, nor the hard work of developing valid measures of teacher performance against those standards.

A defensible teacher evaluation scheme must be based on a clear understanding about what it reasonable to hold a teacher accountable for. The appropriate basis for gathering evidence about a teacher's performance is a set of professional standards that describe the full scope of what a teacher is expected to know and be able to do.

Glenn Rowley and Lawrence Ingvarson are Principal Research Fellows at the Australian Council for Educational Research.

*This article was originally published in the opinion pages of the Education supplement of The Age (Reference: 'Teacher study fails the test,' opinion by Glenn Rowley and Lawrence Ingvarson, Education, The Age, June 4 2007, page 16).*

## Further information

### Reference

McCaffrey, D., Lockwood, J.R., Koretz, D.M., & Hamilton, L.S. (2004). *Evaluating Value-Added Models for Teacher Accountability*. Santa Monica, CA: Rand Corporation.

**Published June 2007**

## Research report

Lawrence Ingvarson, Elizabeth Kleinhenz and Jenny Wilkinson recently prepared a review of research on performance pay for the Australian Government Department of Education, Science and Training - Research on Performance Pay for Teachers. The report is available on the [DEST website](#) and also the ACER website.

Published June 2007

## Year 12 aptitude test on the horizon

In its recent budget the Australian Government announced its intention to pilot a Year 12 aptitude test to assist in university entry. The test is not intended as a replacement for existing tertiary entrance ranks (UAI/ENTER/TER), but as an additional piece of evidence that universities will be able to use to identify students with the capacity for higher education study.

The proposed test will further broaden the evidence already being used by Australian universities as part of their student selection processes. It joins tests used by a number of law, business, dental and medical schools, including the Undergraduate Medicine and Health Sciences Admissions Test (UMAT), the Graduate Australian Medical School Admissions Test (GAMSAT), and the Special Tertiary Admissions Test (STAT) used by many universities to select mature-age applicants who do not have recent Year 12 results.

So what is an aptitude test? Unlike traditional IQ tests which include questions on vocabulary, patterns and abstract content, modern aptitude tests measure candidates' abilities to understand information provided and to apply that information to realistic problems. The focus is on academic reasoning: the kinds of thinking that underpin higher education studies and that are essential across a range of academic disciplines.

Two Australian universities, Monash and ANU, recently trialled an aptitude test to select students from a range of social backgrounds. This test, known as *uniTEST*, was developed by the Australian Council for Educational Research (ACER) and Cambridge Assessment to assist Australian and UK universities to identify students with the verbal and quantitative reasoning skills required for university study. By providing applicants with a 'second chance', the test is designed to make university more accessible to students whose backgrounds and circumstances may have placed them at a disadvantage in regular selection processes.

Perhaps predictably, concerns sometimes are expressed that aptitude testing will lead to a lowering of standards on campus. But research shows that a combination of Year 12 results and aptitude test scores provides a good indication of likely success at university.

Published June 2007

Aptitude tests also provide an additional piece of evidence for differentiating among high-scoring students who currently can be compared only on the basis of their tertiary entrance ranks. Evidence from *uniTEST* trials in the UK suggests that the test may be helpful in comparing students who have very similar secondary school results.

Finally, a common Year 12 aptitude test may provide more direct comparisons of inter-state applicants and assist in the statistical process used to compare tertiary entrance ranks across Australia.

For all these reasons, the introduction of a Year 12 aptitude test may be a welcome addition to the student selection landscape.

footnote: ACER will be piloting [uniTEST](#) in a number of Australian universities during 2007.

Article by Louise Reynolds, Geoff Masters and Deirdre Jackson.

Published June 2007

## ACER UPDATE

### ACER to conduct PISA 2009

ACER is pleased to announce that it has again been selected by the OECD to conduct the major international components of the Programme for International Student Assessment (PISA) for 2009. This will be the fourth cycle of the PISA study, which assesses the reading, mathematical and scientific literacy of 15-year-old students in more than 50 countries. ACER will also develop a computer-delivered assessment. ACER will lead an international consortium of research and educational institutions and eminent individuals to deliver the International PISA project. The testing will take place during 2009 with publication of the results due in late 2010. For further information on ACER's role in PISA 2009 please visit ACER's PISA web page.

### Australia-India Teacher Exchange Program

ACER's International Institute hosted 10 New Delhi teachers participating in the 13-day Australia-India Teacher Exchange Program in late May. ACER conducted a two-day professional learning program for the teachers, which included hands-on activities, practical exercises, and lively discussions focused on the teaching of mathematics and science.

This professional learning program, funded by Australian Education International (AEI) through the Australian High Commission in New Delhi, was arranged to coincide with the selective exchange program managed by the Asia Education Foundation. It was the first time ACER had provided a professional learning program for teachers in the program, who had to undergo a rigorous selection process to participate. This round of participants came from a variety of schools in and around New Delhi.

Published June 2007

## Research Developments latest issue now available

Issue 17 of ACER's news magazine *Research Developments*, is now [available online and in print](#).

This issue features articles on Civics and Citizenship education, performance pay for teachers, the proposed national core curriculum for Year 12 and the movement of non-metropolitan Australian youth towards the big cities.

## Copyright © Australian Council for Educational Research 2015

All rights reserved. Except under the conditions described in the Copyright Act 1968 of Australia and subsequent amendments, no part of this electronic publication may be reproduced, stored in retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying or otherwise, without written permission. Please address any requests to reproduce information to [communications@acer.edu.au](mailto:communications@acer.edu.au)

[Subscribe](#)

Australian Council for Educational Research  
Private Bag 55, Camberwell, Victoria Australia 3124  
Tel: + 61 3 9277 5555  
Fax: + 61 3 9277 5500  
Web: [www.acer.edu.au](http://www.acer.edu.au)

