

THE ASSESSMENT OF TEACHER QUALITY

AN INVESTIGATION INTO CURRENT ISSUES IN
EVALUATING AND REWARDING TEACHERS

SUMMARY DOCUMENT

EDUCATION POLICY RESPONSE GROUP
INSTITUTE OF EDUCATION
MASSEY UNIVERSITY
SEPTEMBER 2013

Ivan Snook
John O'Neill
Stuart Birks
John Church
Peter Rawlins

THE ENGINE
OF THE NEW
NEW ZEALAND



THE ASSESSMENT OF TEACHER QUALITY

An investigation into current issues in evaluating and rewarding teachers

EDUCATION POLICY RESPONSE GROUP

Institute of Education

Massey University

Ivan Snook

John O'Neill

Stuart Birks

John Church

Peter Rawlins

September 2013

CONTENTS

	Page
SUMMARY	1
The Treasury policy agenda (Section 1).....	1
Effective teaching: an international perspective (Section 2).....	2
International studies of student achievement (Section 3).....	3
Teacher effectiveness (Section 4).....	5
Value added measurement (Section 5).....	6
High stakes teacher assessment (Section 6).....	7
Performance Pay Schemes (Section 7).....	8
Conclusion (Section 8).....	9
Members of the Educational Policy Response Group for this Project.....	13
Acknowledgment	13
About the Education Policy Response Group.....	13

SUMMARY

1. In the past thirty years a movement for educational reform has spread throughout the world. Now known popularly as the Global Educational Reform Movement (GERM) it is promoted by international development agencies and private enterprises seeking to influence educational policy. This movement advocates competition, choice, business-style management, curriculum standardization, “value added” assessment of teacher effectiveness, and performance pay.
2. When we look at recent developments in New Zealand education and policy documents emanating from the Treasury it seems clear that many elements of GERM are on the ascendancy here. Viewed in this way, many particular innovations (such as national standards, charter schools, the extension of integrated status to elite private schools, and the proposals for schools in Christchurch) seem to fit neatly into it.
3. The Treasury policy agenda would depart significantly from the social democratic philosophy underlying the reforms of *Tomorrow's Schools*. It seeks to initiate a regime in which educational research (however controversial, subjective and unreliable) is to be interpreted by state officials and used by state agencies to determine in close detail what happens daily in all state classrooms and, hence, influence very significantly the educational experience of every child. It is very important, therefore, that these proposals and the claims which underpin them should be subjected to critical scrutiny.

The Treasury policy agenda (Section 1)

4. Treasury's education policy agenda and advice to government are based on conventional economic theories of business process and human capital. In these it is assumed that variations in the quality of teachers and teaching can be removed over time through the use of performance incentives based on value added student achievement data. This, in turn, will resolve the problem of structural education inequalities.

5. Treasury's analysis is simplistic and largely ignores or discounts other plausible contributory causes to educational inequalities. For example: the education system's existing policy settings may not support teachers to prepare students for assessments; some parents may spend considerably more than others in supporting their children's education outside school; increases in government funding may not keep pace with demands made on the school system itself; or inequalities in achievement may be a function of social and economic inequalities outside the school and outside the teacher's control.
6. It is also assumed that value added assessment measures and performance-related pay incentives will: (1) be valid, reliable and practicable; (2) be motivational to teachers; and (3) produce no unintended consequences. Such assumptions may be wrong and need to be subjected to critical educational scrutiny before such a radical departure from current policy settings is countenanced for New Zealand children and teachers.

Effective teaching: an international perspective (Section 2)

7. A 2005 OECD study of 25 educational systems around the world produced three main findings: (1) the largest variations in student learning are attributable to what students bring to school (ability, attitude, background, and so on); (2) "teacher quality" is the single most important school variable; (3) Most research on teacher quality is narrowly focused on test scores and readily measurable teacher characteristics. Characteristics that are harder to measure, but may be just as vital to student learning (such as clarity, enthusiasm, creativity, warmth and the ability to create effective learning environments and relations), are typically not measured.
8. Despite the overwhelming findings about the effects of home background, politicians and the popular media persist in minimising it and insisting that great improvement in learning can be gained by concentrating on schools and teachers in isolation from home background and parental income.
9. Numerous studies demonstrate that the socio-economic background and prior experience of students are the main influences on learning. Their effects need to be carefully taken into account if they are not to be overlooked in the creation of school-

based “solutions” which cannot work and which merely alienate teachers whose morale and cooperation are essential.

10. Despite the OECD’s claim of the critical importance of teachers, it is important to remember that, in another way, the student is of most importance: what he or she brings to school in the way of personal ability, cultural norms, values, skills, and background experience plays a central role in what and how he or she learns.

International studies of student achievement (Section 3)

11. Treasury (and Government speakers) frequently refer to international studies of school achievement to indicate that all is not well with the New Zealand education system. In particular they claim that: (1) New Zealand has a wider distribution of student achievement than similarly achieving countries, in other words “The long tail of underachievement”; and (2) that due to this wide distribution of achievement, New Zealand’s education system is considered only “fair to good”.
12. We examine in detail the findings of PISA (which are most often quoted by Treasury and Ministry speakers) and also look at TIMSS, PIRLS and our own New Zealand based monitoring study NEMP. Overall, the results do not support the Treasury’s claim that New Zealand has a “long tail of underachievement” in relation to similarly achieving countries.
13. On PISA New Zealand has both a higher than average proportion of students in the higher proficiency levels and a lower than average proportion in the lower levels on tests of reading, mathematics, and science literacy. This indicates that, although New Zealand might have a wide spread of achievement, this spread is situated towards the upper end of the achievement spectrum. High achieving students are doing better and lower achieving students are generally performing on a par with comparable countries.
14. On the other hand, the TIMSS data tell us that, compared to other similar countries/economies, our performance in mathematics and science is rather poor at Year 5. However, students improve as they get older, in both their average performance, and in the proportion of students meeting international benchmarks. The

PIRLS data tell us that there is a relatively wide spread of achievement in reading at Yr 5 with higher proportions of lower achieving students than the international median. According to PISA, however, by age 15, our lower end performance is similar to comparable countries and large proportions of our students reach higher benchmarks.

15. Treasury also claims that New Zealand teachers are less successful than others in counteracting socio-economic disadvantage but fails to acknowledge that in the late 2000s, New Zealand was one of the most unequal countries in terms of income distribution (Gini index), ranking 28th out of 31 OECD countries. Between the mid 1980s and mid 2000s, income inequality rose in 15 of the 19 OECD countries for which data are available. New Zealand (together with Sweden) had the highest cumulative rate of change in this period. In the same period, there were significant increases in the numbers of people living in poverty (below 50% of median income) in New Zealand.
16. Predictably, PISA found that the impact of economic, social and cultural status on students' performance in New Zealand is the highest among the OECD countries and the likelihood of disadvantaged students performing at levels similar to those of their advantaged peers is much lower in New Zealand than in countries such as Australia, Canada, and Japan.
17. However, despite this, the proportion of “resilient” students—those who come from a disadvantaged socio-economic background and perform much higher than would be predicted by their background—is 9.2% of all New Zealand students, statistically higher than the OECD average of 7.7%. Moreover, if we take this as a proportion of those who are considered disadvantaged then approximately 36% of disadvantaged students can be considered “resilient”, higher than the OECD average of 31%.
18. The interpretation placed on these data by the Treasury is that New Zealand teachers have been less successful than others in overcoming social disadvantage. This is surely to misunderstand the PISA data which, on the contrary, suggest that despite higher than average levels of income inequality and child poverty in New Zealand society, higher than average numbers of students overcome the odds and achieve.

19. Although some international comparison studies might identify areas of content or skills where students are relatively stronger or weaker nationally, this could equally well be related to the structure and content of the curriculum or preferred pedagogy. In many instances these are national policy mandates rather than teacher or school decisions. Accordingly, in order to provide a balanced picture, these studies should be considered in context, in their entirety, along with other national measures of student achievement, in light of the international research literature, and in consultation with expert educational groups.
20. The Treasury claims that the third McKinsey Report shows that as a result of our wide distribution of scores, the New Zealand system is rated only “fair to good”. Inspection of the report reveals that although it investigated 20 educational systems from around the world in some depth, New Zealand was not among them. Indeed, New Zealand is mentioned only twice in the report, once on a graph depicting various countries’ achievement levels in international assessments against public expenditure per student, and once in the list of OECD countries that act as a control group in developing a “Universal Scale”. Even if we accept New Zealand’s position on this Universal Scale, the claim that it is our wide distribution of achievement scores that has placed us in a particular position on the scale is not supported by the methodology used to create it.
21. On our analysis, the Treasury’s principal claims about New Zealand’s performance in international assessments are false: (1) There is no “long tail of underachievement” in any general sense; (2) New Zealand teachers are not deficient in remedying socio-economic disadvantage; and (3) The claim that New Zealand education is only “fair to good” because of its “tail” of underachievement, is unsubstantiated by the report which is used to support the claim.

Teacher effectiveness (Section 4)

22. While officials and politicians favour the term “teacher quality”, there is no agreed definition of what this means in practice, other than raising student achievement scores. Research studies of teacher effectiveness focus on what the teacher does and what the student learns. All of the studies we reviewed have limitations and therefore suggest caution on the part of policymakers who wish to link periodic assessments of

teacher effectiveness with teacher registration, employment, career progression and/or remuneration.

23. These studies generally have been limited to measures of improvement in reading (and sometimes mathematics) and have largely ignored all other schooling outcomes such as improvements in the other six curriculum areas, improvements in social skills, improvements in self-directed ability to learn and improvements in the unique talents of individual children.
24. All of these studies leave a major part of the variation in student achievement unexplained. This is because most of the measures of achievement used are not very accurate and because many of the factors affecting achievement (e.g. home background) have been measured with even less accuracy.

Value added measurement (Section 5)

25. In recognition that teachers cannot be held accountable for effects outside their control (e.g. home background, previous knowledge, classroom composition, natural ability and mental engagement) researchers have turned to trying to assess what “value” a teacher has “added” to the student’s achievement over the time they were in his or her classroom. This is a particular form of achievement based assessment known as value added measurement (VAM). VAM is problematic because: (1) it takes little account of differential student abilities; (2) it is difficult to control for the widely different learning experienced by students outside school; (3) the variables which are the most easily rated are not necessarily the best for assessing the overall quality of learning; and (4) there are many other explanations for improved achievement scores.
26. VAM scores have been shown to be very unstable, with dramatic fluctuations from year to year, and high rates of measurement error. Consequently, the scores gained by teachers in the first year of assessment, cannot predict even broadly their scores in the second and subsequent years. VAM schemes fail partly because so many factors influence student learning. Among these are: (1) other teachers of the child; (2) the quality of school resources; (3) the variety of out of school learning experiences at

home and in the community; and (4) the ‘learning loss’ that occurs over the long vacation. These factors do not influence all students equally.

27. There are also negative consequences that flow from reliance on VAM. The main ones are: (1) it leads to narrowing of the curriculum; (2) teachers are discouraged from working with special needs students and children from low income homes, and schools (anxious for their reputation) resort to all kinds of strategies to exclude such children; and (3) teachers are discouraged from co-operating with other teachers on the grounds that they might thereby lose some tactical advantage. At a time when team teaching and shared professional knowledge are regarded as educationally very important, VAM works against the cooperative spirit.
28. Econometric studies suggest important connections between teacher qualities and student achievement on tests but they face serious statistical and analytical problems and are unable to provide any analysis of what constitutes a good teacher or any advice to policy makers regarding teacher recruitment, training or reward. Based on documented error rates in such approaches, we estimate that if there were 100,000 teachers in a system and three years of student achievement data were used, at least 5,000 teachers would be dismissed unfairly and 5,000 underperforming teachers would be retained. If the administrators used data accumulated over 10 years, “only” about 2,500 teachers would be wrongly categorised. The suggestion, therefore, that teachers be dismissed on the basis of student achievement is simply bizarre

High stakes teacher assessment (Section 6)

29. Two specific teacher performance evaluation schemes were examined, the Teacher Evaluation System (TES) in Cincinnati, USA, and the Contextual Value Added (CVA) scheme in the UK. The notion of “performance” in each is quite distinct. In the TES, the teacher is evaluated on what she does (her performance, like an actor). In the CVA, the teacher is evaluated in terms of “how successful she is” (her performance in raising the “achievement” of her students). This dichotomy is marked by different approaches to the measurement of “performance”.

30. The Cincinnati TES does not involve any measurement of student learning and hence is not a “value added” tool. It is a “standards” based procedure which is intended to examine and refine the activities of teachers through multiple structured classroom observations by trained evaluators who are experienced classroom teachers. Comprehensive evaluations are undertaken at approximately five yearly intervals for experienced teachers and more frequently for newer teachers. In addition, a scheme is in place to assist new and experienced teachers who experience difficulties. Two studies we reviewed have claimed to identify increases in student achievement as a result of the scheme’s introduction. However, the reported student achievement gains have been modest, inconsistent across learning areas, and could equally well be explained by other variables not considered by the study authors.

31. The English CVA sets out to show the progress children have made while attending a particular school. Unlike statistics that merely report actual exam performance, the Contextual Value Added system attempts to take into account the circumstances of children attending the school that are beyond the school’s control. This is done by comparing a child’s performance with that of children with a similar prior performance and similar circumstances. Numerous variables are included in the model in an attempt to “control” for factors beyond the school or teacher’s control. Conservative politicians have claimed that the statistical methodology discriminates against socially advantaged schools, while liberal/progressive politicians have claimed that it underestimates the challenges faced by disadvantaged schools. Researchers have claimed that the CVA suffers from serious reliability and measurement errors to the extent that the results are only marginally more reliable than unadjusted achievement test results.

Performance Pay Schemes (Section 7)

32. One major review of performance pay concluded that: (1) judgments about teacher performance for “high stakes” purposes such as registration, reward, career advancement and promotion must be based on several sources of evidence; (2) performance standards are multidimensional (e.g. knowledge, skills, dispositions), and therefore multiple forms as well as multiple sources of evidence are required; and (3) valid and reliable assessment results require independent, trained assessors together

with evidence about the context in which judgments about performance are being made.

33. Performance pay schemes that are linked with student achievement adopt two quite distinct approaches: (1) teacher performance is based on student performance on standardised tests; or (2) teacher performance is based on evidence about what students are doing in classrooms as a result of conditions for learning established by teachers.
34. In our view teachers may be held responsible only for exercising their professional judgment to create supportive conditions for students to learn: they cannot be held responsible for what, or how much, students learn since this is beyond their control.
35. Rewards can only motivate additional effort. They cannot teach new skills. Rewards will motivate additional effort only if the recipient judges that they are worth the extra effort. Rewards can have perverse effects. For example they can function to reduce intrinsic motivation. Given that much teacher effort is intrinsically motivated, great care must be taken to avoid any kind of extrinsic reward system that would damage this intrinsic motivation.

Conclusion (Section 8)

36. We conclude by noting that Treasury's policy advice for education is based on a relatively simple economic model of the effects of investing in human capital: variations in teacher quality strongly influence variations in student achievement. To secure the required improvements in teacher quality, incentives and sanctions should be applied according to gains or losses in expected student achievement.
37. We have already argued that teachers are only one of many factors that influence student achievement and that Value Added Measures are entirely unreliable. Therefore they cannot ethically be used to reward or sanction teachers. Nevertheless, we strongly support the goal of improving school effectiveness.
38. We agree with the McKinsey and Company's common sense observation that at system level: "High performing school systems consistently do three things well. They

get the right people to become teachers (the quality of an education system cannot exceed the quality of its teachers). They develop these people into effective instructors (the only way to improve education is to improve instruction). They put in place systems and targeted support to ensure that every child is able to benefit from excellent instruction (the only way for the system to reach the highest performance is to raise the standard of every student”.

39. We are also aware of the OECD’s observation that teaching profession is in long-term decline, yet there is increased awareness of how crucial teachers can be in the achievement of students and in the progress of society. It is essential, therefore, that we avoid policies (such as constant surveillance and performance management) which alienate teachers, deter students (especially the most gifted) from a teaching career, encourage unprofessional behaviour (such as cheating on tests) and lead to rapid turnover of teachers in a system that needs stability.
40. The Treasury agenda for education is a dangerously narrow and simplistic interpretation of what the evidence says about assessing and improving the quality of teaching. Bluntly, their solution is the problem in our view.
41. Instead of Treasury’s business process model and economic theories, we urge consideration of a range of policy options that are based on educational theories and research evidence that recognises the complexity of teaching and learning and the social contexts in which they occur both in and out of school.
42. Broadly, at the system level we argue for policy settings that place higher trust in teachers and in the power of teacher collegiality to improve teaching; for targeted student engagement and retention policies in the middle and senior secondary years, particularly in low decile schools; and abandonment of the current, harmful obsession with National Standards and NCEA Level 2 targets.
43. We observe that various policy and funding decisions over the past twenty years have led to a situation in which many candidates for initial teacher education in the primary school sector may not have the knowledge and skills needed to deliver the mathematics and science curricula. If this is true, simply measuring student achievement and

sanctioning teachers will do nothing to address underachievement. Indeed, in this context, proposals to reduce teacher education programmes from three or four years to one year by making primary ITE postgraduate entry only are of considerable concern.

44. We offer six alternative policies that, if implemented carefully with sufficient resources, have the potential to make schools more effective than they are currently.
45. Teachers cannot change the things they do not know how to do. For a teacher to become more effective, that teacher must acquire new skills. This requires the teacher to discover that a more effective teaching method exists, engage in active rehearsal of the new skill with feedback and to continue practising until they have achieved mastery.
46. Priority should be given to the acquisition by teachers of the new diagnostic, teaching and evaluation skills that have been shown through research to be more effective in fostering student learning than those which are currently being used.
47. Policies are also needed to accelerate the production of teaching materials that have been field-tested and shown to be effective in developing the learning outcomes contained in the New Zealand curriculum.
48. Policies are needed to upgrade the management, appraisal and professional development skills of all staff involved in education at system level and at school level.
49. All those involved in the system also need to develop clearer understandings of what counts as sound educational research and reliable research evidence, and how research may be used to improve teaching and learning.
50. Finally, education policies must encourage collaboration by capitalizing on the intrinsic motivation of school leaders and teachers. Motivation cannot be mandated and trust cannot be legislated for. Yet both are integral to securing system-wide improvements in schooling over the longer term.

We hope that we have pointed the way toward a better approach to that currently being advocated. We argue that alternative policies are required to enhance and upgrade the status of the teaching profession, to attract and retain high achieving candidates and to allow them to flourish as autonomous professionals. Although we do not claim to have all the answers, we submit that the evidence we have presented strongly suggests that the agenda we put forward is more defensible than that of the New Zealand Treasury and we welcome informed debate on it.

Members of the Educational Policy Response Group for this Project

Ivan Snook is Emeritus Professor of Education, Massey University

John O'Neill is Professor of Teacher Education, Massey University Institute of Education

Stuart Birks is Senior Lecturer, Director of the Centre for Public Policy Evaluation, Massey University

John Church is Adjunct Senior Fellow, Department of Psychology, University of Canterbury

Peter Rawlins is Senior Lecturer, Massey University Institute of Education

Acknowledgment

The authors acknowledge the significant contribution of Emeritus Professor Terry Crooks to successive drafts of this paper. His detailed comments have improved the document greatly. He is, of course, in no way responsible for the opinions expressed or for any mistakes in citation or interpretation which remain.

About the Education Policy Response Group

Massey University's Education Policy Response Group (EPRG) has a history dating back to the early 1980s. It is an ad hoc, eclectic grouping of policy scholars and other interested staff and postgraduate students at Massey University. Its purpose is to meet when necessary to prepare considered responses to major government education policy initiatives and other topical educational debates. The EPRG's first venture was an analysis of the proposed Core Curriculum issued by then Minister of Education, Hon Merv Wellington, in 1983. Since then it has issued reports on many topics such as the curriculum, teacher education and student assessment. Its most recent reports were a critical appraisal (Snook et al, 2009) of Professor John Hattie's 2009 book, *Visible Learning: A Synthesis of Over 800 Meta-analyses Relating to Achievement* and an examination (Snook et al, 2012) of the Charter Schools proposal: *Charter Schools for New Zealand: An investigation designed to further the debate in New Zealand on education policy in general and on charter schooling in particular*.

