

# Draft report of the Productivity Commission on the National Education Evidence Base: Response from Deakin University School of Education

---

## *Contributors:*

Head of School Professor Christine Ure

Professor Jillian Blackmore

Associate Professor Andrea Gallant

Dr Radhika Gorur

Professor Amanda Keddie

Director of REDI Professor Julianne Moss

Professor Andrea Nolan

Dr Shaun Rawolle

Dr Julie Rowlands

Dr Andrew Skourdoumbis

Dr Matthew Thomas

Professor Russell Tytler

Dr Scott Webster

Deakin School of Education has established national standing for research that is above world standard, with an ERA 4 and a QS ranking of five in Australia and 31 internationally. The School is one of the three largest schools of Education in Australia. It offers a diverse range of pre-service teaching specialisations at both undergraduate and postgraduate levels, across the fields of early childhood, primary, secondary, vocational and adult education. The School's strong post graduate programs and its partnerships with schools, systems, governments and other agencies supports the development of innovation and research across diverse education contexts.

The School, through its Strategic Research Centre Research for Educational Impact (REDI), maintains an applied research agenda of high national standing that is committed to social justice and equity. Four programs of research supported through REDI are; Curriculum, Assessment and Digital Learning (CAPDL); Children, Young People and their Communities (CYPC), Science, Technology, Environment and Mathematics Education (STEME) and Governance and Policy (G&P).

The development of a national evidence base for Education is needed to improve mapping and evaluation of the impact educational innovations. A national evidence base will strengthen Australia's capacity to evaluate both international and national influences on the diverse education contexts in which Australian education is delivered.

The outcome of the consultation on this draft report has significant consequences for education systems, educators, education researchers and research funding in Australia, and deserves considered analysis and critique. In this submission we first give an overview of our response to the key points made in the draft report, we then comment on particular recommendations, drawing on expertise from within the School.

We commend the National Education Evidence Base draft report for recognition in the report of a number of major challenges and issues:

- That top down pressures for accountability can be counterproductive and that this has been recognized in other systems.
- That there is a need to focus on bottom up practice as a way of investigating the impact of policies, programs and practices in the complex environment of schools and classrooms
- That there are gaps in existing data collections, particularly in the need to generate evidence of the impact of current policies and practices for different schooling contexts
- That currently there is too much of an administrative burden on schools
- That there is a need to improve the translation of research into practice and that teachers should not be viewed simply as deliverers of policy but should be involved in researching practice.

However, in responding to the argument within the draft report for the generation of evidence and its use across the system:

- i. We raise the question of whether the proposed data base will deliver what is claimed, particularly given the complex nature of the education system, of classrooms, and the diversity of the workforce (e.g. qualification and skill levels vary significantly within and between the early childhood education and care sector and schools). There is a need for a more nuanced approach to translating evidence into grounded practice.
- ii. We question the assertion that there has been a considerable increase in expenditure upon education which has not improved outcomes. We argue that substantial increases in expenditure have not gone to those schools that need it most. The Gonski report fully addressed this issue. Further, the point has been made elsewhere that much of the expenditure increase in education has targeted improved test outcomes but has not focused on social outcomes. However, test outcomes in Australia to date suggest that doing more of the same isn't working e.g. literacy and numeracy classroom instruction has reached saturation point and other pedagogies are required informed by longitudinal qualitative research. Examples of how to improve student outcomes (beyond test scores) by focusing on social equality can be found in Finland and Poland. In summary, the argument for the need more evidence is at least partly questionable, because it risks sidestepping the need for new strategies to address what is now recognized as growing educational inequality in Australia (OECD, 2015).
- iii. We note within the report that there is frequent slippage between the terms 'data', 'evidence' and 'evaluation'; yet these are very different things, that each take different forms, have different purposes and utilise different modes of collection. Closer scrutiny of which forms of data can provide evidence for which purposes is required. The report needs to acknowledge these differences.
- iv. We assert that the nature of evidence required to inform improvements in early childhood education and care (ECEC) and school-based practice is more complex than acknowledged in the draft report. The models being promoted for linking data and evidence privilege quantitative large scale data at population level, with the effect that issues of translation of policy and research findings into the environments of classrooms and schools, and recognition of context, are ignored or silenced. There is a need to expand understandings of evidence beyond quantitative data sets. There is a further risk that a medical model of epidemiological research will simply be applied to the education sector and that this would fail to address the social, emotional, physical and cognitive nature of learning as well as contextual factors that inform learning.
- v. We note that there is currently little within the draft report that acknowledges the different forms that evaluation takes. Evaluation can be responsive (e.g. action research), formative (throughout the program) and summative (end of the program), each approach producing different forms of quantitative and

qualitative data, but all of value to practitioners and researchers because they address the full range of learning outcomes:

- *attainment* as measured by one-off standardised test scores (e.g. NAPLAN PISA), classroom assessments and teacher observations;
- *pedagogical* effects as indicated by improved engagement in learning (proxies such as time on task, self-management);
- *social* in terms of improved student/teacher, teacher/teacher and student interactions, evidence of increased levels of student interpersonal competencies and team work;
- *affective* as indicated by a sense of belonging and inclusion, self-esteem and self-confidence;
- *wellbeing*: physical comfort and health, sense of safety; and
- *behavioural* changes related to retention, vandalism, absenteeism, suspensions, expulsions, disciplinary incidents, violence, disruption in class, lateness, racial incidents, smoking. (Blackmore et al. 2010, p. 6).

- vi. We further argue that the restricted view of evidence outlined in the draft report has implications for what is valued in education research, and what is privileged as worthy of study in the complex environments of education systems. It implies that there is only one worthy model of research and evaluation (randomised controlled trials) yet there are multiple forms of research that can improve classroom and school practice. Education research is diminished by the implied reduction of classroom practices and outcomes to a set of data points, particularly if that data is narrowly focused, restricted and decontextualized.
- vii. We suggest that the evaluative research framework proposed implies that meta-analyses of randomised controlled trials (RCT) research will be privileged in education research yet promotion of such research in the US has not led to significant improvement in international ranking in PISA and TIMSS. The No Child Left Behind policies with their advocacy of RCT as the gold standard of research has led to contradictory outcomes. Nor does the use of statistical modeling necessarily produce robust data (Goldstein 1997, 2015). Data sets are often incomplete and prone to limitations and contamination due to workforce issues amongst other factors.
- viii. We argue that the model of data imposition implied within the draft report is not based on mutual accountability. A more thorough and rigorous treatment of the mutuality is required, including between 'top down' and 'bottom up' capabilities that go beyond a simple evaluative relationship.
- ix. We further argue that judgment of effectiveness is reliant on the completeness of the data base in terms of its relevance to the aims of a program and its capacity to address the multiplicity of student/child outcomes. Currently much of the data referred to within the draft report relates to relatively simplistic models of learning and high stakes tests based on narrow outcomes and as such these have a distorting effect on practice. There is now substantive Australian and international research that indicates that monitoring outcomes, performance benchmarking and competition between schools can be counterproductive to improving performance. There are alternative models for collecting systemic data, other than universal national testing at all levels, to inform policy and evaluate improvement, including evaluations focusing on themes or a sample of different populations.
- x. We assert that effecting improvements in an education system requires the identification of the most effective programs, policies and practices. The reality is that schools experience multiple shifts in policy and programs, which change often in response to funding cycles, making it difficult to generate reliable data. These problems of constant flux and competing program demands in a complex system are not sufficiently recognised in the models proposed.
- xi. We further assert that there are significant dangers in building policy for education that is based on the presumption that only numbers count as evidence. The generation of numbers can necessitate the reduction of meaningful behaviours and outcomes to simplistic categories in the interests of measurement, and the re-translation of these categories into advice and support for teachers/educators can become meaningless. There is a crucial role for more complex forms of evidence, for instance in representing the

types of capacities that are expected for 21<sup>st</sup> century thinking and learning, and for supporting teacher/educator practice to develop these higher order capabilities in students/children. There is a further need to develop teachers/educators as researchers who can generate, communicate and interpret more complex forms of evidence-based practice than is implied in this report as part of the envisaged data base, or as part of critical 'bottom up' capability.

- xii. We argue that the report has significant implications for the funding of education research in Australia. There is little reference within the draft report to the role and approach of the Australian Research Council, which is the primary funder of educational research (Category 1 grants). The draft report also has significant implications as to research training in education and the types of research that will be recognised, for example in scholarships.

## Response to particular recommendations and arguments in the report

### Draft Recommendation 7.1

The Australian, state and territory governments should ensure that an online metadata repository for education data collections is created. The approach used by the Australian Institute of Health and Welfare could serve as a model.

### DRAFT RECOMMENDATION 7.2

The Australian, state and territory governments should pursue a national policy effort to develop a high-quality and relevant Australian evidence base about what works best to improve school and early childhood education outcomes. In particular, five activities need to be supported:

- development of research priorities
- commissioning of high-quality education research
- adoption of rigorous research quality control processes
- dissemination of high-quality evidence
- development of researcher capacity.

The draft report argues, quite rightly, that as far as possible decisions around education policy and practice should be based on quality evidence and that standards of rigor must be applied to education research with a view to providing such evidence. However, the arguments mounted are not critically informed and contain a number of unexamined presumptions about, for instance, the nature of educational experiments, the possibility of rigorous measures of worthwhile outcomes, the translation of research findings, meta analyses, and the nature of controls. This response draws attention to some of these presumptions and the negative consequences that would flow from their use as a means of controlling research practice.

The argument within the draft report that ‘gold standard’ should be adopted for funded education research borrows from policy in the US and UK that is at least a decade old, and which is currently subject to a significant rethinking arising from restrictions on educational research traditions. These more current lessons from the US and the UK should inform development of a suitable standard for education research.

## Can we adequately measure rich educational outcomes?

Australian policy formulations relating to educational purposes are generally wide ranging and deal with longer-term development of knowledge, capabilities and dispositions. The Melbourne Declaration (MCEETYA, 2008) is one such example, encompassing principles relating to successful learners; and confident and creative individuals. In contrast, curriculum outcomes described by ACARA (N.D) are shorter term and more explicit. These more contained outcomes are necessary for defining curriculum topics and activities, but are much more prosaic and restricted than the broader purposes of education. The outcomes identified within system measures such as NAPLAN or TIMSS are themselves restricted accounts of the broader knowledge and capability sets described by curriculum documents. Almost invariably the requirements of reliability in controlled trials force such a reduction in measures of outcome.

Requirements for standards of proof based on an epidemiological model of research risk putting in place outcome measures that are ‘low hanging fruit’ in terms of worthwhile educational purposes and outcomes. The demand for rigor in this circumstance, depending on reduction of complex educational purposes to numbers, amounts to an abandonment of rigor in conceptualizing the complex purposes of a mature education system such as Australia’s. Even within the medical research literature questions are raised about the focus of attention on methodological questions such that the nature of measures are ‘black boxed’ in ways that reduce evidential validity (McClimans, 2013).

## **Is the proposed gold standard the only version of 'rigor'?**

The methodological basis of the gold standard movement has been questioned on a number of grounds (Lawrenz & Huffman, 2006). First, it has been pointed out that science itself builds knowledge by far more varied methods than clinical trials, or experimental methods generally, including through observation and descriptive methods, modeling, and theory development drawing on a wide range of evidence. Second, there exists an enormous body of methodological writing establishing standards of rigor in a wider variety of research designs. For instance interpretivist studies and longitudinal designs (Tytler, 2009) or case study methodologies all have established practices for establishing validity or trustworthiness. Third, an exclusive focus on experimental methods privileges one set of values only, around reductive nomothetic views involving the generation of general laws built on reproducible facts and outcomes, rather than a concern with understanding how individuals interpret and act in the world (Cohen, Manion, & Morrison, 2000). Further, limiting evaluation to "scientific" experimental approaches can privilege a regularity view of causation with a forced emphasis on the determination of patterns of relationships, rather than seriously considering how or why these patterns occur (Maxwell, 2004). Fourth, educational innovation builds on theoretical advances and a range of research designs to generate, validate and scale up new practices, with experimental designs relevant to only part of this process. Even with evaluation of innovation, there are strong arguments for mixed method designs rather than a reliance on experimental methods only (Lawrenz & Huffman, 2006). Finally, it must be acknowledged that at every level, from individual learning to teacher actions and classroom environments and beyond, education systems are complex, and difficult to reduce to categories simply and reliably described and measured in a manner assumed by gold standard advocacy.

## **Do RCTs deliver what is claimed?**

The randomised control trial (RCT) is an import from the health sciences, and in particular from epidemiological studies. The draft report argues that the RCT should be standard for education research. However, in a complex environment such as education where outcomes are not capable of objective definition and human behavior amongst students/children and teachers/educators is necessarily more complex than those of pathogens and body systems, there are dangers attendant on over-reliance and naïve faith in RCT methodology. These dangers relate to a) the definition of both intervention and outcome, b) the nature of the control, c) translating results into policy and practice, and d) discounting research that seeds innovation or offers critical appraisal. In each case, the problems arise from applying numbers to complex human behaviours.

### **A problem of definition**

Particularly when making judgments about new approaches to teaching and learning it is difficult to define the core nature of the innovation, and to ensure adherence to the tenets. For instance, in studies of the effectiveness of inquiry approaches to teaching science, there has been significant variation in what is meant by 'inquiry' and considerable analysis of different approaches to inquiry. This is not in itself a problem, but examination of meta analyses have demonstrated that what can be taken to be the same core innovation can in fact be very different things in different places. Additionally, theory-developed approaches to teaching often contain multiple inter-linked elements such that a problem arises in experimental studies which ascribe causal significance to a particular element. Further, with multiple elements present in the innovation, it is difficult to define what we might mean by a 'control'.

In order to develop a quantifiable measure required by RCT methodology it is always necessary to pare back definitions (e.g. of problems or treatments) to that which can be measured numerically. This is problematic for matters which are highly complex and contextual, such as 'problem solving capability' or 'collaborative reasoning'. Even for conceptual understanding of core ideas such as 'particle theory' the measures in place in normative tests tend to privilege recall or demonstration of understanding in stock situations, rather than flexible understandings.

Thus, as argued above, reliance on experimental methodologies tends to focus attention on low hanging fruit of measurable constructs, such that the purported methodological rigor undermines rigor in pursuing high level conceptual and other outcomes.

### A problem of control

Similar to the problem with defining the treatment in an RCT, controls in education tend to be simply groups, classes or schools proceeding 'as usual'. This is not a 'control' in the scientific sense. The notion of 'control' is ceded to 'normal' which inevitably consists of a wide variety of practices in a wide range of circumstances. The RCT is best known within experimental epidemiological study. The notion of random control is a misnomer, and the RCT bears little relation to scientific studies involving variable control from which the RCT derives its imputed status. In many experimental studies in education, both the treatment and control classes are under-described, and it is difficult to ascertain what is being compared. For example, analysis of meta analyses of research exploring the effectiveness of inquiry treatment has shown that in many of these studies the 'control' group was under described, and to the extent these were described, showed considerable variability—so much so that in one often quoted study showing that inquiry learning was ineffective, the control group in fact employed a guided inquiry strategy while the treatment group was an extreme example where the teacher played no part in helping students.

### The problem of translating results

As the draft report points out, translating education research results into education policy and practice outcomes is problematic. One of the problems in undertaking research and developing policy based on quantifiable measures is that these mask what are highly complex classroom practices, and real reform requires significant changes in teacher beliefs and understandings about learning. In order to develop professional learning approaches to support such change what is often needed are rich case studies and narrative descriptions of teacher-student/ educator/child interactions. These allow teachers with different backgrounds working in varying contexts to translate the research. For example, it is one thing to identify 'feedback' as a significant factor in supporting learning, but feedback to students will always sit within a particular pedagogy, and the reductive extraction of one component such of itself provides little real conviction or support for teachers.

The report identifies the need for a 'bottom up' approach alongside the 'top down' imposition of data, but what has meaning at the school and classroom level must acknowledge the complexities of context and history, which cannot be reduced to simple prescription.

### Discounting seeding and other research

The argument for epidemiological (RCT) research rests mainly on the need for verification that particular innovations can be effectively scaled up. The draft report rightly points out that scaling up from small innovations can be problematic, because of differences in context and also because innovations often work because of the particular attention and support that is possible at this level, that cannot be produced at larger scale. However, to imagine that the education system can be improved only on the basis of support for RCT-based scaling studies ignores the need to support research that explores and refines new innovations. An exclusive focus on RCT studies would take out of the system a rich variety of exploratory studies that are the seedbeds of innovation. There are also many highly valuable strands of research in education that focus on settings other than classrooms, such as policy analysis, or fundamental research into learning, that need support.

### A restricted view of meta-analyses

The draft report rightly identifies meta analyses as important policy tools. However, meta analyses of RCT studies are subject the same limitations as RCT research itself, described above, and their reductive nature can lead to superficiality and lack of contextual relevance if the richness of the educational research oeuvre is ignored. A major

drawback of such work is that so many important studies are placed out of contention by the requirement for RCT trials, that could offer critically important insights into the effects of interventions. As an alternative to meta-analysis of RCT studies, there are other meta-analysis methodologies available that use qualitative and mixed method research studies, and that extract their findings using thematic analysis techniques.

#### Draft RECOMMENDATION 8.1

The Australian, state and territory governments should task the COAG Education Council to provide explicit policy direction through a new Education Agreement, which would build on prior agreements and define the:

- objectives
- nature of the research to be undertaken in the bottom-up evaluation of what works
- evidentiary standards or frameworks to be applied, including assessment of cost effectiveness
- requirement for translation of evidence into guidelines accessible by schools, early childhood education and care services and teachers.

They should also request the Education Council to:

- assign an institution to be responsible and accountable for implementation of the functions set out above and in Draft Recommendation 7.2
- specify the assigned institution's governance arrangements, functions and operations
- including a responsibility for promoting a culture of using the evidence base by policy makers and educators.

#### Information request 8.1

The Commission seeks further information about the strengths and weaknesses of its proposed institutional and governance arrangements.

## **On an Education Agreement**

Education agreements, and the national goals they link to serve as social contracts for education in Australia. In effect, the draft report proposes a major refashioning of this contract. We argue that any future education agreement should be tied to national goals and should also prioritise equity and social justice considerations, in line with both the present national goals, and Australia's commitments in relation to the OECD. There is currently limited mention of these matters in the recommendations of the draft report. Some reservations have also been expressed about whether a national education agreement along the lines proposed in the draft report would be possible given the governance of education in Australia as it stands.

## **On an institution to oversee education evidence**

If policy is to be borrowed from other nations, then one of the considerations for this borrowing ought to be the nature of the problem, and whether the nations that Australia is contemplating borrowing from offer a relevant solution. The draft report makes the case that Australia's education performance has stagnated in the last 10 years in comparisons based on international tests such as TIMMS and PIRLS. It is reasonable then, to look to the policy instruments of countries whose performance has been better than our own over the last 10 years. Although cited almost exclusively throughout the draft report, the United Kingdom and the United States do not appear to fit this condition.

The second substantive point about 'assigning an institution...' raises concerns about 'promoting a culture of using evidence...'. While it is hard to argue against using evidence, the overall sentiment suggests that a new type of enforceable standard/s regime will be put in place that may have the power or powers to reward/punish schools, the ECEC sector and teachers/educators. This is not consistent with medical models of evidence based practice or more recent accounts of evidence informed practice in Australia and elsewhere. If such an institution were to be established it must remain separate from commercial interests. Producer capture would occur if sectional interests



in particular research methodologies were allowed to direct what counts as the core standards of education research, without wider input from the sector, including from universities.

An important issue here to think about here is the notion 'what works'; relatedly, what counts as good schooling and learning; and more generally, the purposes of schooling. While randomized trials might indeed be a good measure in efficiently capturing and comparing academic performance and offering prescriptive and scalable mechanisms through which to raise this performance, they are clearly limited. In particular, they pay little heed to 1) the enormous influence that contextual factors (e.g. situated, material and professional) play on matters of measuring and raising academic performance (thus compromising the reliability and validity of their results to be generalizable to broader populations as they specifically intend to do) and 2) they affirm the spurious and untenable notion that one-size-will-fit-all. Moreover, they are a very narrow measure of what schools are about. The goals of Australian schooling (MCEETYA, 2008) have long mandated that schools and schooling should pursue both private and public goals – private goals being associated with employment credentialing and social mobility and public goals being associated with social and moral learning towards the betterment of the social world. Quality ('gold standard') research whether bottom up or top down should pay heed to context and to accounting and capturing how schools are pursuing both private and public goals.

#### DRAFT FINDING 1.1

Notwithstanding substantial increases in expenditure on education over the past decade, national and international assessments of student achievement in Australia show little improvement and in some areas standards have dropped.

Investment in Australian school and ECEC education must respond to increases in student numbers, staff, their salaries and education facilities. The connection between 'increased spending' and perceived 'improvement' is a tenuous one and should not be reduced simply to improving achievement on national and international assessments of perceived student (under) performance. The focus on basic standards (literacy/numeracy) delimits the work of schools and teachers as there is now an increasing and acknowledged global focus on higher-order thinking. Schools and teachers enable the learning experiences for students that will help develop the resilient, creative and critical thinkers of the future.

The socio-economic profiles (SES) of schools matter. Funding models that provide additional resources to schools with high needs help minimize differences in outcomes (see Perry & McConney 2010). Fundamental to school funding is a basic belief about what schools should be doing. Good schooling and a sound education provisions a base on which the future is founded. Concerns with failing and unproductive schools using international comparisons of school productivity ignore public goals linked to creating active and informed citizens and more democratic and socially cohesive societies. National prosperity and productivity begins in classrooms with quality teachers/educators geared for the demands of a changing and challenging world.

Research has shown that Australian schooling standards have not declined (see Gorur & Wu 2015). There is no 'crisis' in school education, although there is a need to better focus and target resources. Besides reducing school socio-economic differences, foundational aspects involving diversity and choice should not be ignored. Funding objectives should be pursued that broaden the educational opportunities and outcomes of students from socially disadvantaged backgrounds.

#### Draft recommendation 2.1

In supporting the further development of a national education evidence base, governments should be guided by the following principles.

The national education evidence base should:

- meet the varied needs of decision makers at all levels of the education system
- provide high-quality data and evidence to inform decisions
- drive improved student achievement through four interconnected processes — monitoring of performance, evaluation of what works best, dissemination of evidence and application of that evidence by educators and policy makers
- generate benefits in excess of the costs incurred in collecting and processing data and in creating, sharing and using evidence.

Data generation and examination are not ends in themselves but serve as means to improve decisions about education. The interpretation of data is complicated, more so when involving assessment and making comparisons across years, or different sub-groups. Teachers/educators require high level data skills to effectively plan for and respond to student needs. This is most challenging for the ECEC workforce with varying levels of qualification ranging from a Certificate III to a degree or master's.

The emphasis in education on 'what works' or 'best practice' is a misnomer as context matters and varies (see Wrigley, Lingard & Thomson 2012).

There are also problematic assumptions about what constitutes 'high-quality' data and 'evidence'. While there might be principles of practice that generate productive ways of working, these are inextricably shaped by context. Education is complex, and superficial treatment of issues connected to learning is unjust. Schools/ECEC settings and teachers/educators generate and use data regularly and while there is room for teachers/educators to better engage with data, this can be achieved through quality professional learning and development.

Contemporary student populations are diverse and teachers/educators draw on data-informed decision-making processes to facilitate student/child learning. There is scope in Australia to invest in the professionalism of teachers and then to trust them to do their job. This process has begun in the ECEC sector but needs further support. An hierarchical approach to data generation and management will only further remove teachers/educators from their broad sphere of operation. Teachers/educators are professionals in their own right and as such engage with and make professional judgements and decisions about data constantly, particularly regarding pedagogy and assessment.

#### Draft recommendation 3.1

In assessing whether to improve the quality of existing education data, governments should examine whether:

- there is a need to improve the quality of the data so it is fit for purpose
- data quality improvements are feasible given the context of data collection
- other options are available
- the benefits of improving data quality exceed the costs.

There are serious limitations in using cost-benefit analysis in complex fields such as education, health and environment policies. We can learn from the objections to the use – let alone the sole use – of the CBA approach from environmental policy debates, where critics have argued that:

- CBA necessarily requires ascribing monetary value to outcomes– and this is very difficult to do in education, where student well-being, social growth, ethical and moral development and a range of other important outcomes are impossible to assess, let alone quantify.
- Cost benefit analysis requires comparison – the relative benefit of x over y – and this in turn requires commensuration – i.e., different aspects of education need to be comparable. This is highly problematic in education. Even such basics as definitions of 'student', 'teacher' and 'primary school' are difficult to stabilise.

Given this, using universal ‘best practice’ models in policy or in classroom or administrative practice is extremely fraught.

- On the whole, costs are more easily calculated but benefits are much harder to measure and quantify. This is because the ‘productivity’ of students is often manifested over their lifetimes and life trajectories are unpredictable. Moreover, the most valuable, enduring benefits or ‘outcomes’ only become visible over time. ‘Costs’ and ‘benefits’ occur at different times in education and so using CBA as the single lens or approach is highly problematic – it will skew the analysis towards easily measured and largely irrelevant outcomes. An example from the critique of CBA in environmental policy illustrates this well: ‘CBA would encourage officials to adopt a program that saves one life today over a program that would save 600 million lives in 300 years’ (Hwang, 2016, p. 77, citing Clowney, 2006, p. 123).
- CBA assumes money to be a neutral measure – but it is clearly not neutral. When, on what, and who is funded makes a big difference. Funding buildings or providing heating in school buildings may increase comfort but may not produce measurable outcomes on NAPLAN or PISA. Similarly, funding a wealthy school which may already be operating at its optimal level may provide no measurable benefit. In environmental policy, it is argued that calculating CBA using the sum of costs and benefits tends to favour policies which are iniquitous. Such calculations disregard who bears the cost and who gets the benefits (Ackerman and Heinzerling, 2002).

More detailed studies (for example, Gorur, 2011b) point to the highly complex, inter-related factors that influence both decision-making and how policies actually get translated in practice. Conversation with teachers/educators and principals will show that:

- a. ‘Interventions’ such as the introduction of any ‘best practice’ model are variously understood and inevitably adjusted and translated, and often ‘domesticated’ to fit into existing circumstances – and so a ‘cause and effect’ relationship from an intervention to its benefits on a large scale can be misleading.
- b. Not infrequently, some of the practice guidelines that are encouraged (such as, for example, not spending too many hours preparing for NAPLAN) may be ignored depending on the circumstances of the school – again confounding such analyses.
- c. If resources are not allocated to support practices, the practices may not yield results and may be prematurely and unwisely abandoned in such CBA analyses.

In health, the QALY measures based on cost-benefit analysis in the UK are now being abandoned.

If CBA is used to make policy decisions in education, it would require a much more nuanced approach than a national-level estimation of costs versus a national level estimation of ‘benefits’ based on NAPLAN and PISA results.

### **The leap from ‘within school differences’ to classroom practice and ‘teaching practices’**

*“Research has found that only a small share (typically about 20 per cent) of variation in individual student outcomes is explained by differences between schools. The majority (about 80 per cent) is explained by differences between students within schools. Furthermore, there is a substantial body of evidence suggesting that teachers have the greatest impact on student performance outside of students’ own characteristics, and that directing attention to higher quality teaching can have large positive effects on outcomes across the board. All of this suggests that looking within the classroom, particularly at teaching practices, can be more effective at providing insights into how to improve education outcomes across schools and students.” (p. 7)*

In the case of the statement above, regarding the between-school and within-school differences, these data are very complex and do not translate automatically to a conclusion about variability in ‘teacher quality’. Schools are not smooth and regular units. Often, particularly in secondary schools, students self-select or are placed in particular

subjects or levels so that students doing advanced mathematics may be performing differently to students taking foundation mathematics, in the same grade level. In other cases, students in Year 7 may have more indifferent results than students in Year 8, because Year 7 is a transition year. The disengagement of students in Year 9 and the consequent focus on providing a range of special programs to engage them attests to how the same students might display variations in their motivation, their application and their maturity – all accounting for ‘within-school variation’.

Hattie (2003) has attributed variation in student performance to various factors: Students (‘what students bring to the table’ p. 1) – 50%; Home (expectations, encouragement etc.) – 5 – 10%; Schools, Principals and Peer Effects – 10-20%; and Teachers – 30%. This has been one of the most influential pieces of research and has framed discussion on ‘teacher quality’ for over a decade. Hattie asserts that instead of focusing on system-level changes, curriculum changes etc. we should focus on what happens between teachers and students in the classroom. A major issue with this argument is that it assumes that what may be statistically separated through factorial or regression analysis – ‘home effects’, ‘school effects’, ‘peer effects’ etc. are actually separable and operate independently of each other in the classroom. This is not correct. All these factors come together in complexly entangled ways in the classroom. For example, a student being bullied by another (school effect), or hungry because of parental neglect (home effect), or bored because he is forced to study a compulsory subject (policy effect – not included in Hattie’s effects) will find it very difficult to focus on and excel in the classroom or in NAPLAN, irrespective of the ‘quality’ and expertise of the teacher.

Moreover, this research cited in the draft report is quite old now, and it was not conducted in Australia. Since that time, the demographic profiles of our schools have changed, our policies on inclusion and special learning needs have changed, our funding models have changed, and our teacher education programs have changed.

#### INFORMATION REQUEST 3.1

The Commission seeks comment on whether the Australian Early Development Census could be used to monitor progress against Australia’s early learning objectives.

#### DRAFT RECOMMENDATION 3.2

The Australian Government should request and sufficiently fund the agencies that conduct the Longitudinal Study of Australian Children and the Longitudinal Study of Indigenous Children to establish new cohorts of children at regular intervals.

#### Draft recommendation 3.3

Australian, state and territory governments should support greater use of value-added measures of education outcomes.

In Australia, significant attention has been paid to the collection of larger population-based data sets and smaller sample data collections in the early childhood sector over the past 10 years, however, there are limitations to these which result in gaps in Australian specific data. The international evidence base is much stronger in relation to ECEC.

To ensure quality ECEC service provision that will strongly influence child outcomes, data is required that focuses on current policy initiatives and on looking towards the future. Data needs to build on what is currently available such as the AEDC (Australian Early Development Census), NAPLAN (National Assessment Program – Literacy and Numeracy), LSAC (Longitudinal Study of Australian Children), and LSIC (Longitudinal Study of Indigenous Children). Administrative data needs to be incorporated, ensuring that what is collected reflects what it is supposed to be collected e.g. kindergarten attendance data currently collects enrolment data and does not reflect participation. No current data set collects everything that is required to enable an informed assessment of quality in ECEC. For example, as noted in the 2006 OECD report, large data gaps appear in statistics addressing young children, especially children under the age of three. Due to the ‘patchwork’ nature of ECEC services (Elliott, 2006) – different service types, functions, mismatches of funding and regulations, different jurisdictions, different indicators and methods of collecting data on young children are utilised.

The Australian Early Development Census (AEDC) provides a snapshot of how children in local areas have developed by the time they start school. The data is collected by teachers of children in their first year of full-time schooling every three years and is presented at the population level. While there are clear links to significant ECEC documents such as the National Quality Standard (NQS) and the Early Years Learning Framework (EYLF), research (Janus & Offord, 2000; Kershaw, Irwan, Trafford & Hertzman, 2005) has highlighted the problematic nature of such measurement instruments due to their heavy dependence on teachers' understandings of the discourses of early childhood development that underpin these instruments. The Longitudinal Study of Australian Children and the Longitudinal Study of Indigenous Children are significant sample data sets and should continue to be supported. Up until recently there has been an over reliance on international research due to a dearth of Australian data of this type.

In the ECEC sector attention needs to focus on creating cross-disciplinary data sets (education, health, welfare). A systematic procedure is required (OECD, 2016) to ensure consistent and comparable information on ECEC. What has also been identified in the research to date is that to have an impact on health, wellbeing, social competence and education at a population level, communities need access to outcome measures they can use to assess how policies, programs, community assets and social settings for children and families can combine to have an impact on children's development, health and wellbeing.

### Value-Added Measures

The aim of 'fit for purpose' is problematic as purposes change over time. Critically engaging with research is important and Australian social science researchers have a strong interest in the development of innovation and building research capacity. The call for greater use of value-added measures of education outcomes is questionable because it reflects simplistic conceptions of growth in learning (see Reckase 2004). Value-added research in education poses challenges as randomness and the learning conditions of students is not assured (see Darling-Hammond 2015). Furthermore, supporting greater use of value-added measures of education outcomes in isolation from other education research lessens confidence in the authority of value-added and its stated aims/objectives (see Raudenbush 2015).

Value-added measures, developed and used in some states in the US, have been widely challenged by teachers and teacher unions in regard to their ability to assess teacher quality. There is also disagreement about the validity of these measures among researchers. Among the arguments advanced are the following:

- Standardised tests are not always aligned to the curriculum taught – so teachers are being judged on what they have not taught, and, conversely, students are not tested on what they have been taught
- There is considerable 'noise' and error in such tests, and this makes associating them with teacher quality unfair and invalid
- The tests do not reflect a range of other aspects of a teacher's job.

Other challenges to the test have been on the grounds that gauging teacher quality based on standardised tests has narrowed the school curriculum; there has been wide-spread malpractice in student testing because of the high stakes associated with the tests for teachers and schools; and the practice of evaluating teacher quality in this way, through measures devised outside the profession, devalues the professionalism of teachers.

Equally controversial has been the use of these measures to reward or punish teachers. In the US, in particular, one outcome has been the 'naming and shaming' of teachers whose quality was deemed to be poor. In some cases, a teacher deemed outstanding in one year has found herself below standard in the next. The 'fight over teacher quality' has now moved to courts. New York teacher Sheri Lederman filed a suit against state education officials challenging the VAM model, which, she says "actually punishes excellence in education through a statistical black box which no rational educator or fact finder could see as fair, accurate or reliable". The court ruled in her favour, with the judge declaring these assessment measures to be 'arbitrary' and 'capricious'. The full story, with arguments against VAM, can be found here: <https://www.washingtonpost.com/news/answer-sheet/wp/2016/05/10/judge-calls-evaluation-of-n-y-teacher-arbitrary-and-capricious-in-case-against-new-u-s-secretary-of-education/>

The American Educational Research Association took the extraordinary step of issuing a 5-page statement against the use of VAM to assess teachers (see <http://edr.sagepub.com/content/early/2015/11/10/0013189X15618385.full.pdf+html>) . Drawing on state-of-the-art research in testing, statistics and methodology, it argues that there are "potentially serious negative consequences in the context of evaluation that can result from the use of VAM based on incomplete or flawed data, as well as from the misinterpretation or misuse of the VAM results". It points to the use of teacher observation and peer assistance and review models "that provide formative and summative assessments of teaching and honor teachers' due process rights" as better alternatives. There is a Special Issue of Educational Researcher (March 2015) which focuses on the dangers of using VAM for teacher evaluation and high-stakes decisions.

The American Statistical Association has also issued a statement warning that VAM measures should be used with great care and caution, because, among other things, "Most VAM studies find that teachers account for about 1% to 14% of the variability in test scores, and that the majority of opportunities for quality improvement are found in the system-level conditions. Ranking teachers by their VAM scores can have unintended consequences that reduce quality". This statement can be found here: <https://www.amstat.org/asa/files/pdfs/POL-ASAVAM-Statement.pdf>

Given that VAM has been so controversial both with regard to its validity, and with regard to the effects of its use, and given the weight of the evidence that denies its usefulness, it is surprising that the draft report should recommend its greater use.

#### Information request 4.1

The Commission seeks further information on:

- the costs and benefits of moving toward a national student identifier (compared to jurisdictional systems)
- the feasibility of using the unique student identifier system used in the vocational education and training sector to deliver more comprehensive student coverage
- the costs and benefits of children in the early childhood education and care sector being covered by the same identifier as school students.

#### Draft Recommendation 4.1

Agencies responsible for collecting education data should review and adjust their procedures to reduce the administration costs and the compliance burden on respondents, including by:

- to the greatest extent possible, collecting sample, rather than census data
- removing duplication in data collection and processing
- avoiding frequent changes to reporting requirements, but when changes are necessary, allowing sufficient time for respondents to comply with the new requirements.

The draft report as rendered invites the opportunity to move towards development and implementation of a national student identifier. The strengths of this proposal include a minimisation of duplication, greater mobility for

working families and students, greater access for the individual to their records from anywhere and increased capacity to provide proof of qualifications to potential employers. Being student centered, a USI may assist in mapping student records to an individual over a lifetime and authenticating student records. Whilst initial costs would be incurred, there are long term benefits of ECEC students being covered by the same identifier as school students. In the VET sector, a national student identifier may also aid in building a longitudinal skills database accessible by policy makers and researchers to examine patterns of distribution and attainment. Whilst not without potential pitfalls, we have identified benefits of such a system, noting that such systems are already in place in both Victoria and Queensland, but have limited value when students and their families cross borders. Also we would note, for example, that at a State level projects have been undertaken by the Victorian Government to look at whether or not the Victorian Student Number (VSN) could be implemented into the ECEC sector. The final outcome was that it was far too problematic due to the variance in service types accessed by young children and their families, the different jurisdictions that the services operate under, as well the ethical considerations that would need to be taken into account.

There are significant potential benefits but also potential risks, including data security and privacy. These would need to be managed very carefully.

## References

- Australian Curriculum, Assessment and Reporting Authority (ACARA) (N.D.). Science—Foundation to Year 12. Accessed 24 Dec 2015 at [http://www.acara.edu.au/curriculum/learning\\_areas/science.html](http://www.acara.edu.au/curriculum/learning_areas/science.html)
- Australian Privacy Principles: <https://www.oaic.gov.au/resources/individuals/privacy-fact-sheets/general/privacy-fact-sheet-17-australian-privacy-principles.pdf>
- Australia's 'top five' ambitions, *Discourse: Studies in the Cultural Politics of Education*, 36:5, 647-664
- Ball, K., Haggerty, K. D., & Lyon, D. (2014). *Routledge Handbook of Surveillance Studies*: Routledge.
- Black, K. (2009). *Business Statistics: Contemporary Decision Making*: John Wiley & Sons.
- Claire, S., J., B., Deborah, N., & Paul, W. (2014). *OECD Reviews of Evaluation and Assessment in Education OECD Reviews of Evaluation and Assessment in Education: Slovak Republic 2014*: OECD Publishing.
- Cohen, L., Manion, E., & Morrison, K. (2000). *Research Methods in Education*. (5th ed.) London: Routledge Falmer.
- Darling-Hammond, L. (2015). Can Value Added Add Value to Teacher Evaluation? *Educational Researcher*, 44(2): 132-137.
- Elliott, A. (2006). *Early Childhood Education: Pathways to Quality and Equity for all Children*. Melbourne: ACER.
- Gasskov, V. (2006). *Vocational Education and Training Institutions: A Management Handbook and CD-ROM*: International Labour Office.
- Goldstein H & Barber M. (1997). Why we agree to differ over data. *Times Educational Supplement*, 14.
- Goldstein H & Thomas S. (1995). School Effectiveness and 'Value-added' Analysis. *Forum*, 37(2), 36-38
- Goldstein H. (1995). *Interpreting International Comparisons of Student Achievement*. Educational Studies and Documents 63, Unesco Publishing (Paris).
- Goldstein H. (1996). 'Relegate the leagues: Data from performance tables is crude and often misleading.' *New Economy*, 199-203.
- Goldstein H. (1997a). Educational Research and Evidence-Based Practice: a debate. *Research Intelligence*, No 59, February, 18-19.
- Goldstein H. (1997b). Value added tables: the less-than-holy grail. *Managing Schools Today*, 6(6): 18-19.
- Goldstein, H. (2008). Evidence and education policy – some reflections and allegations. *Cambridge J. of Education*, 38: 393-400.
- Goldstein, H. (2015) Jumping to the wrong conclusions *Significance* October.
- Gorur, R. & Wu, M.(2015). Leaning too far? PISA, policy and Leaning too far? PISA, policy and Australia's 'top five' ambitions. *Discourse: studies in the cultural politics of education* 36(5): 647-664.
- Harrison, J. G., Wyndgate Group, Ltd, & California Postsecondary Education, Commission. (1988). *A Comprehensive Student Information System*: California Postsecondary Education Commission.
- Janus, M. & Offord, D. (2000). Readiness to learn at school. *Canadian Journal of Policy Research*, 1(2): 71-75.
- Kagan, L., & Kauerz, K. (2015). *Early Childhood Systems: Transforming Early Learning*: Teachers College Press.
- Kershaw, P., Irwan, L., Trafford, K. & Hertzman, C. (2005). *The British Columbia Atlas of Child Development*, 40. Vancouver, BC: Human Early Learning Partnership.
- Kowalski, T., & Lasley, T. J. (2010). *Handbook of Data-Based Decision Making in Education*: Taylor & Francis.



- Lamb, S, Jackson, J, Walstab, A & Huo, S (2015). Educational opportunity in Australia 2015: Who succeeds and who misses out, Centre for International Research on Education Systems, Victoria University, for the Mitchell Institute, Melbourne: Mitchell Institute.[www.mitchellinstitute.org.au](http://www.mitchellinstitute.org.au) and [www.vu.edu.au/centre-for-international-research-on-education-systems-cires](http://www.vu.edu.au/centre-for-international-research-on-education-systems-cires)
- Lawrenz, F., & Huffman, D. (2006). Methodological pluralism: The gold standard of STEM evaluation. *New directions for evaluation*, no. 109. Wiley.
- Lingard, B. Thompson, G. and Sellar, S. (2016). *National Testing in Schools. An Australian assessment.* Routledge.
- Majumdar, R. (2007). *Marketing Research - Text, Applications And Case Studies: New Age International (P) Limited.*
- Maxwell, J. (2004). Causal Explanation, Qualitative Research and Scientific Inquiry in Education. *Educational Researcher*, 33(2), 3-11.
- McCahill, M., & Finn, R. L. (2014). *Surveillance, Capital and Resistance: Theorizing the Surveillance Subject: Taylor & Francis.*
- McClimans, L. (2013). The role of measurement in establishing interest. *Journal of Medicine and Philosophy*, doi:10.1093/jmp/jht041
- Ministerial Council on Education, Employment, Training and Youth Affairs (MCEETYA) (2008). *Melbourne Declaration on Educational Goals for Young Australians.* Accessed Dec 24 2015 at: [http://www.curriculum.edu.au/verve/\\_resources/National\\_Declaration\\_on\\_the\\_Educational\\_Goals\\_for\\_Young\\_Australians.pdf](http://www.curriculum.edu.au/verve/_resources/National_Declaration_on_the_Educational_Goals_for_Young_Australians.pdf)
- Monahan, T. (2006). *Surveillance and Security: Technological Politics and Power in Everyday Life: Routledge.*
- Monahan, T., Torres, R. D., Kupchik, A., Bracy, N., Apple, M., Hirschfield, P., Casella, R. Gillomm, J. Hope, A. Lewis, T. (2009). *Schools Under Surveillance: Cultures of Control in Public Education: Rutgers University Press.*
- Organisation for Economic Cooperation and Development (OECD). (2006). *Starting Strong II: Early Childhood Education and Care.* Paris: OECD Publishing.
- Perry, L., & McConney, A. (2010). School socio-economic composition and student outcomes in Australia: Implications for educational policy, *Australian Journal of Education*, 54(1): 72-85.
- Raudenbush, Stephen W. (2015). Value Added: A Case Study in the Mismatch Between Education Research and Policy, *Educational Researcher*, 44(2): 138-141.
- Reckase, M. (2004). The Real World is More Complicated than We Would Like, *Journal of Educational and Behavioral Statistics*, 29(1): 117-120.
- Schneider, M. K., & Burriss, C. C. (2015). *Common Core Dilemma—Who Owns Our Schools? : Teachers College Press, Teachers College, Columbia University.*
- Springer, M. G. (2009). *Performance Incentives: Their Growing Impact on American K-12 Education: Brookings Institution Press.*
- Taylor, E., & Rooney, T. (2016). *Surveillance Futures: Social and Ethical Implications of New Technologies for Children and Young People: Taylor & Francis.*
- Tytler, R. (2009a) Longitudinal Studies into Science Learning—Methodological Issues. In M. C. Shelley II, L. D. Yore, & B. Hand (Eds.), *Quality research in literacy and science education: International perspectives and gold standards (83-106).* Dordrecht, The Netherlands, Springer.
- Victorian Information Privacy Principles:  
[https://www.cpdp.vic.gov.au/images/content/pdf/CPDP\\_Information\\_sheet\\_-\\_Short\\_guide\\_to\\_the\\_Information\\_Privacy\\_Principles\\_September\\_2015.pdf](https://www.cpdp.vic.gov.au/images/content/pdf/CPDP_Information_sheet_-_Short_guide_to_the_Information_Privacy_Principles_September_2015.pdf)

Wrigley, T., Lingard, B. & Thomson, P. (2012). Pedagogies of transformation: keeping hope alive in troubled times, *Critical Studies in Education*, 53(1): 95-108.