8 Malvina Place
Carlton
Victoria 3053
Australia

T. +61 3 8344 3637
W. www.grattan.edu.au
E. info@grattan.edu.au

National Education Evidence Base
Productivity Commission
By email: education.evidence@pc.gov.au
1 June 2016

Dear Commissioner

Re. Grattan submission to the PC Inquiry into the National Education Evidence Base.

The national education evidence base is important and needs substantial improvement. It is a key area where the federal government can add value in school education policy, given state governments hold the main levers for education reform. Yet the use of data and evidence to improve school education outcomes should be considered more broadly and holistically than just the national evidence base. The key points of our submission are summarised below.

Section 1 argues that the scope of the inquiry should be broadened to include a focus on:

1.1 The use of data and evidence by teachers in the classroom, including the monitoring of classroom practices.

1.2 The *use* of national data, not just the *collection* of it. Greater synthesis and promotion of research findings can assist state governments in policy design.

Section 2 comments on two specific topics raised in the *Issues Paper*:

2.1 Education outcomes

- Broaden the use of national metrics beyond literacy and numeracy.
- Measure 'student progress' not just 'achievement'. A new NAPLAN 'years of progress' measure should be adopted.
- Improve the measurement of school performance for policy purposes.

2.2 Data for monitoring system performance

- Improve data collection for longitudinal analysis.
- Better monitor the extent of effective classroom teaching practices in use to inform policy as well as the support and guidance given to schools.

Please do not hesitate to contact Grattan if you have any questions about this submission.

Yours sincerely,

Dr Peter Goss
School Education Program Director

Julie Sonnemann
School Education Fellow

1. Scope of the Productivity Commission Inquiry

This section has two parts:

1.1 Use of data and evidence by teachers and for monitoring classroom practice

1.2 The *use* of national data, not just the *collection* of it.

1.1 Use of data and evidence by teachers and for monitoring classroom practice

The scope of the Inquiry should be broadened to improve the use of data in the hands of teachers in the classroom, and for monitoring of classroom practice, rather than just national evidence for policy makers.

There is great value in using standardised national data to monitor and evaluate student outcomes and inform policy development. For example, Grattan's *Widening gaps* report uses NAPLAN data to compare the learning progress of different groups of students.[1] For example, it shows that, even if they were doing as well in Year 3, students with parents of low education make one to two years less progress by Year 9 than students whose parents have more education. This type of finding has important implications for national and state-level policy development and resourcing.

However, other uses of data and evidence can be equally or more powerful.

Grattan's *Targeted teaching* report describes the large potential for gains from putting data and evidence into the hands of teachers so they can adapt and improve their teaching practice. Three teaching strategies – frequent formative assessment, teacher-student feedback, and teachers evaluating their own impact – lift student learning by between five and 11 months over the course of a year. These strategies all use evidence and data to make teaching more effective, by adaptively responding to individual student learning.[2] High performing systems use similar approaches to build the capability of their teachers to critically evaluate and adapt their own practice.[3]

Effectively policy development also needs to be based on a better understanding of what is happening today. It is one thing to know that the use of a specific teaching approach such as synthetic phonics has great value.[4] It is quite a different thing to know how widely and how well synthetic phonics is being applied in Australian schools today. Better data is needed on the practices that are actually being used in schools.

Grattan therefore recommends that the scope of the inquiry should go beyond the

[1] Goss, *et al.* (2016)
[2] See Goss, *et al.* (2015), Chapter 3
[3] See Jensen, *et al.* (2012), Chapters 10 and 11
[4] See, for example, http://www.fivefromfive.org.au

national education evidence base for use by policy makers, and consider the collection and use of data to improve teaching practice and school practice.

This recommendation is in line with the OECD's 2011 review of evaluation and assessment in Australian school education:

> *The overall evaluation and assessment framework appears as highly sophisticated and well conceptualised, especially at its top level (national and systemic levels). However, there is a less clear articulation of ways for the national agenda to generate improvements in classroom practice through the assessment and evaluation procedures which are closer to the place of learning.*[5]

A proposal for a different framework around feedback

The Productivity Commission's *Issues Paper* proposes a feedback loop as a mechanism for using data to improve education outcomes (see Figure 1).

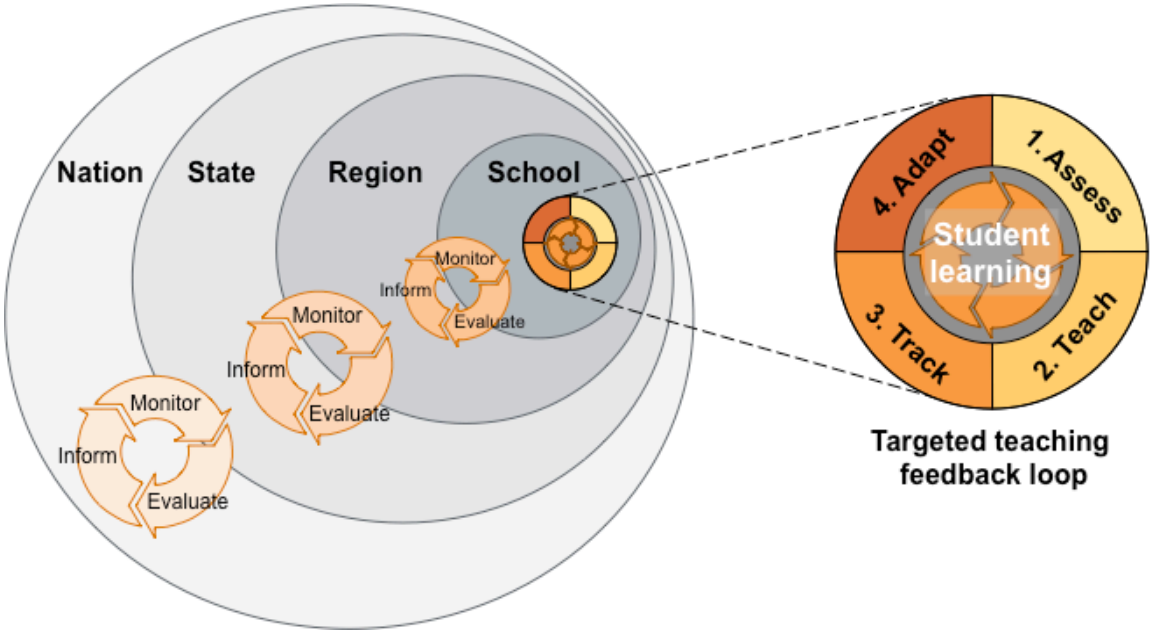
Figure 1: Productivity Commission framework for using data to improve education outcomes

Grattan agrees that feedback loops in education are essential. However, given the

[5] OECD (2011), page 9

complexity of education systems, it is more useful to think about multiple feedback loops and how they link different levels of the system (see Figure 2).

Figure 2: Proposed Grattan model for using data to improve education outcomes



For the sake of illustration, this framework has been drawn with five levels: student, school, region, state and nation. Feedback loops are shown between each pair of levels.[6] In this model,

- The central feedback loop is extracted from Figure 5 of Grattan's *Targeted Teaching* report. Working together, teachers should assess what each student knows now, target the teaching to what each student is ready to learn next, and track each student's progress over time. Teaching should then analyse their own impact, keep what works best and change what does not.[7]

- The second feedback loop is where regional support staff monitor schools, evaluate their impact, and inform practice. Feedback at this level is close enough to the classroom to be able to consider specific teaching and school management practices, while also enabling comparisons across schools and sharing of good practice.

[6] This is most appropriate for government schools in a state that has a regional structure, where regional staff interact more closely with schools than the education department can. The structure of the feedback loops may be slightly different for independent and Catholic schools, but the core concept is valid.
[7] Goss, *et al.* (2015)

- The third feedback loop is where states and territories monitor and evaluate the impact of all the schools in their jurisdiction. The information they gather and analyse allows them to define policy and allocate resources, as well as to identify specific interventions if required.

- The outermost feedback loop is national benchmarking and building the evidence base.[8]

Designing feedback loops using multiple levels has several benefits.

- This approach balances the power of *proximity and immediacy* with the power of *scale and independence*. The inner feedback loops are closer to student learning, and may have a more immediate impact on teacher practice. The outer feedback loops are likely to have greater statistical power and validity.

- Each level of the system has an explicit feedback mechanism for continuously improving its own outcomes, as well as for comparing itself to an 'out-group'.

- Different types of data can be captured for different purposes, allowing for accountability to be applied in nuanced ways. For example, a powerful mechanism for improving teacher practice is to help teachers take responsibility for tracking the progress of all of their students and discussing the data with their peers; but this requires trust. Using that same data to compare the effectiveness of every teacher in a state level could create perverse incentives and reduce trust.[9]

- Defining different data for different levels of feedback loop may also help to reduce collection of data for its own sake, and thereby reduce the reporting burden.

In sum, a multi-level feedback model should enable better decisions while minimising reporting burden and reducing the risk of perverse consequences.

[8] The model could be extended to international benchmarking of educational practices and outcomes, such as the work that the OECD does with the Programme for International Student Assessment (PISA).
[9] See Goss, *et al.* (2015), Box 5, p.39 for a discussion of why high-stakes testing is a high risk approach.

1.2 The use of existing data and research

Broaden the scope to examine ways to improve the use of existing data and research, not just the collection of data.

Australia is relatively well progressed in its national performance system in comparison to other countries, as emphasised in the OECD review in 2011.[10] Yet more can be done in analysing the existing data, in particular around evaluating what works and translating key findings into informed policy decisions. In the context of the PC's framework for using data to improve outcomes (page 9 of the issues paper), a lot of emphasis (and investment) has occurred around collecting and managing data for step 1 (monitoring progress), rather than step 2 (evaluating policies) and then translating that back to step 3 (informing decisions).

More could be done at a national level to analyse school performance data to identify pockets of good practice across the country. For example, the Centre for Education Statistics and Evaluation (CESE) in NSW recently conducted a study that identified high growth schools using a robust value-added methodology (that isolated the contribution that a school makes to growth in student achievement while controlling for contextual factors).[11] While larger state departments are well positioned to undertake this function for government schools, there could be gains from doing it at a national scale and across all three sectors.

Improve the accessibility of the existing research evidence to policy makers

Education policies should be based on a thorough review of the existing research. There are a number of ways to make this easier to do in practice. Some states and territories have departments or bodies that facilitate this information, however there could be gains from developing more niche expertise at scale nationally. The *US What Works Clearing House* is an example of a high quality federal resource. A similar model could add value in Australia.[12]

The New South Wales (NSW) government established a dedicated agency – the Centre for Education Statistics and Evaluation (CESE) – in 2012, which is responsible for supporting decision-making with robust evidence of best practice. CESE also analyses and evaluates NSW education programs and student outcomes. An agency or departmental branch undertaking this type of work at a federal level could be very valuable. While the Australian Institute for Teaching and School Leadership (AITSL) plays a role in this area to some extent, more needs to be done to analyse, synthesise

[10] OECD (2011)
[11] CESE (2015)
[12] See http://ies.ed.gov/ncee/wwc/

and promote the evidence for policy makers.

The federal government could also better support initiatives that promote existing research, such as the new Australian Teaching and Learning Toolkit. This initiative was established with the support of the Victorian Department of Education and Training and Social Ventures Australia. The Toolkit provides a valuable synthesis of the existing research on the impact and cost of a range of individual education policies and interventions. It provides guidance on a number of interventions that have been adopted or debated in Australian education for many years but which lead to only relatively modest increases in learning at best (in some cases they harm learning overall), are relatively expensive compared to other more effective interventions or are not strongly grounded on research evidence.

2. Response to specific issues in PC Issues Paper

This section covers two items:

2.1 Education outcomes

2.2 Data to monitor system performance

2.1 Education outcomes

Issues Paper, p.7 "What education outcomes do you see as relevant?"

Broaden the measurement of outcomes beyond narrow literacy and numeracy skills

Australian schools and systems must keep focused on the outcomes that matter to students, families, society and economic growth. Where possible, we should measure them. This means moving beyond a narrow focus on academic achievement metrics (NAPLAN, ATAR, PISA, etc) to also track broader cognitive capabilities such as problem solving, and personal characteristics such as resilience. To be clear, Grattan is not arguing that NAPLAN should be replaced, but supplemented.

The Melbourne Declaration (2008) defines the goals of school more broadly than academic achievement, but our current assessments do not measure them.[13] The Melbourne Declaration on the Education Goals of Young Australians has two goals: (1) Australian education promotes equity and excellence; and (2) All young Australians become successful learners, confident and creative individuals, and active and informed citizens. This is a key area for national leadership in future, and with time robust metrics should become a part national reporting. Keeping up with advances in assessment tools to test deeper thinking and broad skills (such as cooperative problem solving) as well as knowledge, should be a state and national priority.

Measure 'student progress' not just 'achievement'

The PC *Issues Paper* canvasses an important question '*what does it mean to 'improve' education outcomes?*' One approach might consider improved attainment levels as a success. Another is the extent to which every student is making individual learning progress with time. Grattan has consistently argued that the best way to improve outcomes is to focus on individual learning progress.[14] Student progress measures tell us how much students improve from one year to the next.

[13] The Melbourne Declaration on the Education Goals of Young Australians has two goals: (1) Australian education promotes equity and excellence; and (2) All young Australians become successful learners, confident and creative individuals, and active and informed citizens

[14] Jensen (2010), Goss, *et al.* (2016), Goss, *et al.* (2015)

Measuring student progress enables policymakers to see how students are progressing across the system. This data should influence how priorities are set, and where resources are allocated. Those who are making the least progress, or those who are failing to reach their potential, should be the focus of our policy efforts. This includes high performing schools and students who are making low progress.

At present, it is not easy to compare student progress using NAPLAN data. It would be easy to do so if students gained NAPLAN scores at a steady pace as they moved through school. But they do not. The Australian Curriculum, Assessment and Reporting Authority (ACARA) notes that:

> *… students generally show greater gains in literacy and numeracy in the earlier years than in the later years of schooling, and that students who start with lower NAPLAN scores tend to make greater gains over time than those who start with higher NAPLAN scores.* [15]

The NAPLAN non-linear growth curve makes it hard to compare gaps between different groups of students, or their learning progress. It is especially difficult to compare students of different backgrounds, who are likely to be at very different scores on the curve (in other words, at different stages of their learning), even though they are the same age and in the same year level.

NAPLAN gain scores do not show the full picture of progress. 'Gain scores' are the difference in NAPLAN scale scores between two points in time. They measure student progress in NAPLAN points, and are reported in *ACARA's* National Report on Schooling. But they need to be interpreted very carefully. In particular, gain scores have limitations when policymakers want to compare different groups of students from different starting points (i.e. answer questions of relative progress). In these cases, a face-value interpretation of gain scores can suggest students are catching up when they are actually falling further behind. For an example of this, see Figure 2 in *Widening Gaps*, page 11.

Adopt a new NAPLAN 'years of progress' measure

To address this issue in NAPLAN gain scores, Grattan's recent report recommends a new NAPLAN measure, 'years of progress'. The measure estimates what a year of learning progress looks like on the NAPLAN scale, so that we can better compare relative student performance.

The 'years of progress' measure effectively benchmarks student performance in NAPLAN to the typical student. It allows us to see if students are catching up or falling

[15] ACARA (2016) p.5.

further behind relative to others. For example, instead of saying that a group of Year 5 students are achieving at a NAPLAN score of 540, we can now say they are achieving in Year 5 what the typical student would achieve in Year 7. In other words, they are two years in front of the typical Year 5 student.

Policy makers should adopt Grattan's new 'years of progress' approach to better understand relative student progress and learning gaps'.[16] This metric could be useful in value added analysis (discussed below).

Improve the use of value added analysis

Value-added scores consistently measure student and school performance more accurately than raw scores, because they are better able to isolate the performance of schools from other factors that affect student performance. This creates a fairer system for comparing school performance that is not biased against schools serving more disadvantaged communities.

School value-added scores are calculated by comparing the progress made by each student between assessments, measuring the contribution the school makes to that progress, controlling for students' background. Value-added measures can be used by schooling systems to indicate the contribution that a school makes to student learning, over and above the contribution made by the average school. They are a good starting point for policy makers for more in-depth understanding of 'what works'.

More could be done at a national level to advance the development and application of value-add measures using the national dataset. For example, CESE has developed a set of value-added measures for NSW government schools that adjust for factors outside the control of schools (such as students' SES).[17] However, non-government schools are not covered by this model, and education departments in smaller states and territories may lack the sophisticated capabilities needed for value-added analysis.

[16] Goss, *et al.* (2016), page 3
[17] See CESE (2014)

2.2 Data to monitor system performance

Issues Paper, page 10. "The Commission is seeking input about the kinds of data required to support the three processes in the framework above [monitoring, evaluation and decision making]

As mentioned previously, Australia's national evaluation and performance framework appears relatively well progressed compared to other countries.[18] Any plans for expanding data collection should be weighed up against the benefits of investing elsewhere. Clear distinctions should be made between the kinds of data that should be collected on a consistent, regular basis at a national level, and what additional data can be collected for 'deep dives' in specific research studies and evaluations of policies and programs.

With this in mind, Grattan recommends two key issues to examine in national data collection, and note that these issues apply to schools in all sectors, including non-government schools.[19]

Student contextual data for longitudinal analysis

Australia now has a rich dataset with NAPLAN results at Year 3, 5, 7 and 9, making it possible to track the progress of specific student cohorts over time. However for this student cohort data to be meaningful for longitudinal analysis (including descriptive, correlational and experimental research), it is important to consider some of the following issues:

- Is relevant student contextual data collected in an accurate, regular manner? For example the level of parental education should be collected regularly collected (not once off) to capture changes in circumstances as students move through school.
- Is it possible to capture more information on NAPLAN student progress in Years 2, 4, 6, 8, 10, for example through sample testing NAPLAN assessment? This would help to accurately track student progress over time through school.
- Is it possible to connect NAPLAN data to other longitudinal studies such as LSAY, so that the relationships between student outcomes and further study, work etc can be explored?

Another specific issue is whether there is sufficient data collected on student refugee

[18] OECD (2011), page 9

[19] The OECD (2011) highlighted a need to improve the monitoring of performance in the non-government school sector at state and territory level. Ibid. page 133.

status. Analysing the progress of this specific group of students is a research priority, given their educational needs and the importance of achieving equity in education outcomes.

Information on the extent of effective teaching practices in use across the system

To evaluate the effectiveness of programs and policies, it is useful to have information not only on student outcomes but also teaching effectiveness. This is because of the long causal chain involved in tracing the impact of policies and programs on student learning. Attempting to isolate the impact of policy initiatives on student outcomes is difficult because of other influences outside of the school not within education policy sphere (i.e. home factors). While value added measures can help isolate school impact, they are in still in infant stages of use in Australia.

Collecting information on the extent of effective teaching practices can shed light on whether government and school interventions are working, at least in part. There is strong evidence that improvements in teaching effectiveness can lift student learning.

Yet national data is focussed on student learning outcomes, for example NAPLAN results and attendance rates. Policy makers at state and national levels appear to have little information on teaching practices in use. This may be limiting effective evaluations of policies and programs at scale.

For example, the 2010 Victorian Auditor General's review on the impact of performance and development culture on teaching effectiveness found there was a dearth of information on teaching quality at a state level. This made the evaluation of the policy impact near impossible. It found [it is] *"still not possible to assess and reliably demonstrate whether the quality of teaching has improved. Currently there is insufficient evidence to reliably demonstrate this."* [20]

The cost of collecting data on teaching quality must not be greater than the benefits. The PC should explore cost-effective methods for collating such data. One method that may hold potential is student feedback surveys. There is recent evidence of their reliability as a tool for judging teaching effectiveness at scale.[21]

[20] Auditor-General. (2010)

[21] The Bill and Melinda Gates Foundation (2013) study showed that student surveys are a valid and reliable measure of teacher effectiveness under certain conditions. Reliability is a function of the content of the questions, the consistency of the data collection process, and for survey questions, assurance of confidentiality.

References

ACARA (2016) 'Interpreting NAPLAN results', from http://www.acara.edu.au/verve/_resources/Interpreting_NAPLAN_results_file.pdf

Auditor-General., V. (2010) *Managing teaching performance in government schools*, Victorian Auditor-General from http://www.audit.vic.gov.au

Bill and Melinda Gates Foundation (2013) *Feedback for Better Teaching*, from http://k12education.gatesfoundation.org/wp-content/uploads/2015/05/MET_Feedback-for-Better-Teaching_Principles-Paper.pdf

CESE (2014) *Using value-added measures to identify school contributions to student learning*, Centre for Education Statistics and Evaluation from http://cese.nsw.gov.au

CESE (2015) *High Value-Add Schools: Key Drivers of School Improvement*, Centre for Education Statistics and Evaluation from www.cese.nsw.gov.au

Goss, P., Hunter, J., Romanes, D. and Parsonage, H. (2015) *Targeted teaching: How better use of data can improve student learning*, Grattan Institute

Goss, P., Sonnemann, J., Chisholm, C. and Nelson, L. (2016) *Widening gaps: what NAPLAN tells us about student progress*, Grattan Institute from http://grattan.edu.au/wp-content/uploads/2016/03/937-Widening-gaps.pdf

Jensen, B. (2010) *Measuring What Matters: Student Progress*, Grattan Institute from http://grattan.edu.au/wp-content/uploads/2014/04/016_education_performance_measures_report.pdf

Jensen, B., Hunter, A., Sonneman, J. and Burns, T. (2012) *Catching up: learning from the best school systems in East Asia*, Grattan Institute, accessed 10 May 2013, from http://grattan.edu.au/publications/reports/post/catching-up-learning-from-the-best-school-systems-in-east-asia/

OECD (2011) *OECD Reviews of Evaluation and Assessment in Education, Australia*, from http://www.oecd.org/australia/48519807.pdf