Submission to the Productivity Commission Issues Paper on
National Education Evidence Base from the Learning Sciences Institute Australia

The *Learning Sciences Institute Australia* (LSIA) is an interdisciplinary research institute within Australian Catholic University (ACU). Its focus is generating and disseminating new knowledge on policy and practice to improve the learning outcomes of children and young people. *Its research aligns with the Australian National Priority on population health and wellbeing, with a focus on 21st century learning, student diversity, innovation and engagement* across six areas of research concentration:

- Educational assessment, evaluation and student learning
- Learning, learner diversity and reforming classroom practices
- Science, Technology, Engineering and Mathematics in education
- Early childhood futures
- Educational semiotics in English and literacy pedagogy
- Teacher education, quality and professional practice.

Since its inception in 2014, LSIA has received funding for 25 research projects and is currently seeking funding for a further 11. Funding has been received from the Australian Research Council (10), Government agencies (7), Education bodies (5) and Philanthropic agencies (3). Some projects include analysis of several Australian data sets (LSAY, LSAC, NAPLAN from ACARA, and NAPLAN from QCAA). LSIA currently has 13 professorial and other researchers, 5 professorial fellows, 3 RHD students, 7 administrative and research assistants, 2 visiting scholars, 2 adjunct professors and 5 honorary professors.

LSIA has a substantial and expanding presence in educational research contributing to Australian education policy and practice, experience in using existing national education data bases, including with methodological issues surrounding large-scale data sets, and interest in the further development of a National Education Evidence Base (NEEB), including both contributing to and making use of such a facility.

1. LSIA supports the development of a National Education Evidence Base

The term ‘evidence base’ suggests something more than ‘data base’. ‘Data’ is codified information, and is a form of evidence; however, data require analysis and interpretation to be useful. Evidence can include transformations of the primary (basic or raw) data to produce secondary data, as well as other analyses and interpretations in the form of reports, articles, reviews, commentaries and other publications, all of which contribute to deliberations and decisions on educational policy and practice.

There is value in a *national repository of evidence* that includes secondary data, analyses and interpretations. Currently, there is no site for collection of such documents. Many data analyses and reports are 'fugitive', in the sense that their existence is often not widely known, and access to them is often difficult. Their quality is also often not tested in the arena of open debate. A repository of such documents would be a valuable tool for making them easily available to other analysts and researchers, enabling the validity of data analyses and interpretations to be tested through open discussion, and ensuring that maximum benefit for educational policy and practice is derived from the costs of data collections and analyses.

There is also a need for *coordination of existing primary data sources*, in order to make them more easily accessible for research and analysis. Data are neither easy nor cheap to collect. A repository or a mechanism for easier access to existing databases would be desirable both in terms of the potential for more extensive and more useful data analyses and for greater return on the cost and effort of data collection. Access issues can present a very significant barrier to the use of databases.

Existing educational data collections include data on educational outcomes, student characteristics, parent characteristics, and school characteristics. There is potential to link to other data collections, such as medical/health, Medicare, Census, Centrelink, to answer questions about the relative influences on educational outcomes and to monitor change over time as a result of specific changes in policy and practice.

Coordination and alignment of different databases is itself neither easy nor inexpensive. Further, there are always issues to be resolved in terms of the extent of access allowed, privacy and anonymity, and legal issues requiring a contractual arrangement for usage of the data. There are often issues concerning who owns the data and how permission for access can be obtained. Accordingly, it would be sensible to start by assembling a compendium of existing data, their sources, their characteristics, and access requirements. This might begin with an invitation to contribute, and with dissemination of information about the characteristics and accessibility of existing databases.

It would be unwise to initiate the collection of additional data without consideration of how those data would be used. Data need to be collected for a purpose, which requires a clear intention and design for their analysis. Collecting data 'just in case' requires clever anticipation, but prediction of future data usage is often inaccurate and may not be worth the costs. The focus of data collection is, therefore, analysis for a purpose. Other uses can be serendipitous. In the short term, the need is to deal with issues of identification, linkage and accessibility of existing databases across various data sources with an interest in linking to each other's data.

In such a process, it is difficult to imagine that the outcome would be one single primary database. A single database would be unwieldy and unmanageable. Different databases have different data characteristics, such as who is included, how measures are defined and what coding systems are used. There is a raft of issues that must be dealt with for accessing and using large-scale databases for secondary analysis (Strayhorn, 2009).

A major distinction is that between census data and sample data. Each of these admits a variety of further possibilities.

Census data do not necessarily include everyone in the target population, as there may be exemptions (for example for students with disabilities, or for recent migrants) or omissions (such as people who do not give a response or are not registered or are absent on the day of data collection). There are also significant problems with:

- linking data over time (for example, for tracking educational development or work/study destination) because of movement of residence and/or school
- lack of/problems with unique identification (though new techniques are becoming available for matching on the basis of personal information)
- missing data (through absence from testing or non-response to survey) (though sophisticated techniques are now available for resolving that problem to some extent) (Cumming & Goldstein, 2016).

Sample data will have a unique sampling frame (the population to be sampled and the way in which the sample is selected from that population). Selection could be random, representative or circumstantial, could involve unique stratification (sectioning of the total group to be sampled into subcategories), and could deal with non-participation or non-response in different ways (such as accepting that some data are missing, or using random or matched replacement). These are merely illustrative of the potential differences across sample databases. Such differences mean that specific action is needed to resolve the problems of linkage between different databases. These are likely to need one-to-one resolution, and can be a substantial part of the costs of analysis for any particular research project.

The known problems of linking databases of different kinds make universal linkage unmanageable. Linkage of databases is rarely easy and transparent, and there are always additional costs involved, especially for resolving the technical issues. Yet there are potentially important issues that can be studied by connecting different databases. We conclude that it is important to direct attention to expediting such connections through improved access, transparency and linkage across the range of relevant existing databases.

2. LSIA considers that a National Education Evidence Base should be conceived as a research database

Because of some of the problems already noted, national databases are not amenable to analysis by people lacking skills in research design, statistical analysis and information technology. Primary (raw) data need to be analysed and interpreted to give meaning to the data and such analyses and interpretations are often contested, since they depend on particular research and statistical models with specific theoretical assumptions and procedural techniques. Analyses and interpretations should, therefore, be explained and defended in written reports that provide details of the rationale and reasoning.

It is sometimes the case that the results of data analyses can be summarised in the form of charts and visual displays that allow particular interpretations of the data to be more directly accessed and understood. Such summaries are useful for conveying the findings of data analyses in a form that members of the public may be able to understand and explore for themselves. An example of such summarisation is the MySchool website. Such websites have a role to play in conveying the results of data analyses. However, data displays offer a secondary or interpreted database; they are not the primary or original database. Further, data displays offer particular perspectives on the data, and involve assumptions about what messages ought to be conveyed. The way in which data are displayed can sometimes shape a particular interpretation of the data (Harvard Management Update, 2006; Huff, 1991; Reichmann, 1961; Tufte, 1983).

Public display of some of the conclusions that can be drawn from a primary database (in the case of MySchool from the NAPLAN database and relevant administrative data) only scratches the surface of possibilities for primary database use. Additional analyses can be conducted in a variety of ways if the relevant databases are accessible to researchers.
The effects of public display of interpreted data are not necessarily benign. In the case of MySchool, some caution is needed about the effects — the manner and extent of usage by schools and parents, what conclusions they draw and whether they are accurate, what actions are taken as a result of those conclusions, and whether the outcomes of those actions lead to educational improvements and outcomes.

While MySchool has been beneficial in some respects (encouraging discussion on school improvement in the wider community, creating a successful focus on improvement in literacy and numeracy in some schools, challenging schools to develop strategies for development and improvement of all students), there have been untoward side-effects on some students and schools (through public exposure turning a ‘low stakes’ test into a ‘high stakes’ test, and through ‘disengenuous’ uses of what is actually complex information) (Senate Standing Committee, 2014). Consistent with the experience of accountability (‘high stakes’) testing in other countries, MySchool has had detrimental effects on curriculum breadth, pedagogical practices, teachers’ morale, and student health and wellbeing (Dulfer, Polesel & Rice, 2012).

Some restriction on the kinds of information displayed in MySchool, and alternative ways of providing useful feedback to schools, have been suggested in order to mitigate the untoward negative effects (Senate Standing Committee, 2014), and LSIA supports such moves.

Determining appropriate ways to communicate interpreted data from NAPLAN to schools and the public is an important matter that requires some attention. Communication to schools and the public should be separated. Schools need more detailed information than general public and the best way to communicate database interpretations to schools is to do so directly and individually. However, while large-scale testing can convey some messages to schools about the progress of their students relative to defined benchmarks and relative to other schools and students, analysis of the data takes time. By the time it is available, student learning has moved on, and the information is of limited use for individual student remediation. Other, more local assessments are needed to monitor developing student knowledge and capabilities to inform adjustments in teaching and planning of next steps (Klenowski & Wyatt-Smith, 2014).

Consequently, the provision of information to schools and the public ought *not* be the main purpose of a national education evidence base. The main purpose is providing access to a repository of relevant primary data for research, enabling a variety of investigations that can inform educational policy and practice.

Three particular aims of such research are monitoring outcomes, analysing difference, and tracing effects. Each of these requires deliberate research design and sophisticated analysis of the data. That is, data analysis is not a simple and straightforward activity; naïve exploration of data sets is rarely productive. Rather, data analysis needs to have a purpose and justification, explicit research questions to be answered, choice of appropriate analytical procedures, and interpretation of the data transformations and findings. These are the essential components of research. The inspiration of research can be varied, perhaps arising out of theoretical or policy issues, or perhaps arising out consideration of how data could be analysed further, or an interplay of both. Also, there is the possibility of using existing data or collecting new data that are relevant for the particular research. Relevant research might arise from researcher curiosity or from government priorities, and both possibilities should be encouraged.

3. LSIA agrees with the terminology of objectives, outcomes, measures and determinants but suggests distinguishing between long-term and short-term policy goals

The intention that a national education evidence base should ‘improve educational outcomes’ is more rhetorical than useful. It could be an overarching aim or purpose. However, improvement is a long-term project, and gains may be difficult to realise in the short-term. Evidence requires interpretation through research and analysis. Therefore, an evidence base

would be better conceived as one that will enable useful and comprehensive research and analysis to inform policy and practice. The goal should, therefore, be 'informing better policy and practice'.

There is also no direct path between an evidence base on the one hand and policy and practice on the other. All the steps in translation of evidence to policy and practice are contestable. What counts as evidence to some is not always seen as such by others. Interpretations of the evidence are often seen as valid by some but not by others. Even more importantly, decisions about policy and practice are future oriented. Evidence is about past and/or current outcomes and the effects that may have produced them. Even without change in policy or practice, these outcomes and effects can change. There are always uncertainties about whether what worked in the past or in one circumstance will work in the future or in other circumstances. Research and analysis rarely suggests a single course of action. Decisions about future policy and practice necessarily go beyond any evidence base. Personal, political and institutional understandings, perspectives and values play a role (CERI, 2007; Finnigan & Daly, 2014; Nutley & Davies, 2007).

This suggests that there should be some modesty about aims for a national education evidence base. It can inform, but it cannot assure. Further, research into the ways in which research informs and shapes policy and practice suggests that an evidence base, even one that includes a repository of research and analysis, is insufficient in itself. Attention needs to be given to how relevant messages are disseminated and understood by key stakeholders. This is not a simple matter and requires careful consideration and design. It is helpful in this respect to see the education system as a complex adaptive system (Maroulis et al., 2010; Lemke & Sabelli, 2008) in which various component parts interact in complex and interdependent ways, with dynamic and loose linkages between different governments, departments, agencies, schools, teachers, students, parents and the wider public.
Some issues that make the translation of evidence into policy and practice a complex and unpredictable matter include (see Finnigan & Daly, 2014):

- Some evidence is simple but most evidence is complex
- What counts as evidence is often not universally agreed
- Student performance data are only one type of evidence
- We need to expand the range of what counts as evidence
- Analysis and interpretation of evidence is usually complex
- Different conclusions can be reached from the same evidence base
- Evidence is interpreted and used in a social, political and institutional context
- Interpreting and understanding evidence and research is a learning process
- Interpreting and understanding evidence requires time and effort
- Evidence is often used selectively to support a predetermined position rather than to understand processes, diagnose problems and uncover defensible initiatives

- Making research relevant to policy and practice requires collaboration among stakeholders and better connection between supply and demand
- A 'nuanced conversation of the evidence' is only possible when trust is built between researchers, policymakers and practitioners
- Information technology provides new ways of establishing learning communities and developing collaborative work on use of evidence
- Public dissemination of research is not highly valued in universities, and needs a higher profile and recognition to impact public policy effectively
- Consideration should be given to encouragement and funding of 'research mediators' with special skills in translating research for policy and practice
- Decision-making based on evidence is of interest around the world, but much still has to be learned about how to make this effective.

Improvement in educational outcomes requires more than a rhetorical flourish, or naive expectation that 'data speak for themselves'. The translation of evidence into policy and practice needs effort and resources, and that requires funding. There is little point in establishing a national education evidence base if ways in which it can be used are not addressed and supported.

4. LSIA considers that existing achievement data are inadequate for evaluating educational quality

The *Melbourne Declaration on Educational Goals for Young Australians* (MCEEDYA, 2008) defines learning goals (objectives) for all young Australians and carries substantial weight as an agreement between federal, state and territory education ministers. The learning goals for all young Australians are expressed as follows:

> All young Australians become successful learners, confident and creative individuals, active and informed citizens.

This is further elaborated as including:

a. capacity for self-directed learning; literacy and numeracy skills; creative and productive use of technology; deep and logical thinking; problem solving skills; collaboration in teams; communication skills; making sense of their world.
b. sense of self-worth, self-awareness and personal identity and wellbeing; optimism about their lives and the future; enterprising and creative; personal values and attributes such as honesty, resilience, empathy and respect for others.
c. moral and ethical integrity; appreciation of Australia's social, linguistic and religious diversity and understanding its government, history and culture; understanding and acknowledgement of the value of indigenous cultures; knowledge and skills to contribute to, and benefit from, reconciliation between indigenous and non-indigenous Australians; commitment to national values (such as democracy, equity and justice); intercultural awareness, empathy and

communication (especially with Asian cultures); civic mindedness (especially for sustaining and improving natural and social environments); responsible citizenship (both local and global).

These goals include learner skills, learner identity and learner values — the latter two in more substantial detail than is usual. What is omitted is any mention of specific kinds of knowledge (subject matter); the closest is 'appreciation' of national history and culture, and this features as part of a list of values, not content. The range of goals goes some way to covering the range of so-called 21st century competencies, but needs some extension to cover the full range of such competencies (European Communities, 2007; European Parliament and European Union Council, 2006; Gordon Commission, 2013; OECD, 2005; UNESCO, 2013; Pelligrino & Hilton, 2012; Voogt & Roblin, 2010, 2012).

Almost no evidence has been collected on whether most of these goals are being achieved. For these goals, evidence is available on literacy and numeracy, and the sample testing of civics and citizenship covers a few other goals, but the coverage is not complete and systematic. We simply do not know very much about whether any progress is being made in the direction of most of these goals. Any monitoring or evaluation of progress in achieving these goals would require the collection of a substantial range of data. For a National Education Evidence Base to be worthy of its name, this lack of evidence on many nationally agreed goals needs to be addressed.

The absence of content areas in the list of goals is surprising. However, the National Curriculum defines relevant content areas (subjects) and specific learning outcomes. These learning outcomes are presumably also 'national goals', and also need to be monitored.

We recognise that evidence is available to some extent on some of the subject areas of the national curriculum. The National Assessment Program (NAP) includes sample testing in Science Literacy, Civics and Citizenship and ICT Literacy and provides useful evidence about learning in these areas. The tests of Science Literacy are related to the national curriculum in science. The tests of Civics and Citizenship cover some aspects of the Melbourne Declaration. And the tests of ICT Literacy assess an important area of practical awareness that is important in the modern world. The concentration on Years 6 and 10 is justifiable as they are respectively the end points of primary school and junior secondary school. The three-year cycle for each of these assessments is also an appropriate timeline and use of resources. The inclusion of PISA, TIMSS and PIRLS data within NAP widens the available evidence on Mathematics, Science and Reading; however, none of these is directed explicitly at learning outcomes of the national curriculum and they cover only some interim years of schooling (variously Years 4 and 8, and 15 year olds).

In suggesting an extension of measures of learning goals in both the Melbourne Statement and the National Curriculum, it is not considered that these need to involve a census collection of data. Data are not needed on every student to assess the status of educational outcomes and provide useful feedback to schools. In fact, a sampling approach has many advantages in terms of the coverage of a wider range of learning outcomes, the kinds of measures that can be used, and the costs of obtaining relevant data. The US *National Assessment of Educational Progress* (which produces a *National Report Card*), and has operated for over 50 years, provides a baseline of useful examples and techniques (though the range of their measurements is much less than envisioned here).

In saying this, we acknowledge that data on all these goals could only be collected over the long term, and consideration needs to be given to how this can be done without making unreasonable impositions on schools and students and within reasonable cost-benefit limits. Therefore, there needs to be some modesty in our ambitions: *Think big but start small*. Nevertheless, a start needs to be made and evidence on goals collected progressively over time. Some evidence could be collected by specific commissioned studies leading to the publication of reports that can themselves be part of the National Education Evidence Base. Other evidence ought to be collected more systematically through a variety of measures that are revisited over time to chart progress. In any case, there needs to be a plan for how these data would be analysed and reported.

The narrow focus of NAPLAN on Literacy and Numeracy is also a concern. International research has shown that 'high stakes' tests have a constricting effect of the taught curriculum, as schools and teachers focus their attention on raising their students' scores on the test ('teaching to the test') and devote less attention to other parts of the curriculum. A consequence is that scores on what is tested may rise without improvement in the underlying understandings and skills; further, knowledge and skills in areas of study and learning that are not assessed may atrophy (Koretz, 2008). Similarly, for example, studies have shown that standardised testing in science in England was having a detrimental effect on the learning of science, whereas the effects of the abolition of testing in Wales had been beneficial — resulting in the abolition of such testing in England as well (Collins, Reiss & Stobart, 2007; House of Commons Children, Schools and Families Committee, 2008; SCORE, 2009). We are concerned that NAPLAN, despite some documented benefits, is having similarly constricting and detrimental effects here, and that more public debate is needed on the limitations and appropriate use of such data (Lingard, Thompson and Sellar, 2016; Wyn, Turnbull & Grimshaw, 2014).

NAPLAN measures of literacy and numeracy are not only incomplete and inadequate measures of the educational outcomes and the quality of schools, their use as single measures of literacy and numeracy is probably indefensible. This is a question of validity. How valid are the scores that result from these tests as measures of literacy and numeracy?

This is not a simple issue. Internationally accepted guidelines for testing, such as the AREA/APA/NCME (2014) *Standards for Educational and Psychological Testing*, the validity of the inferences to be drawn from large-scale standardised tests should be thoroughly analysed and publicly defended. These standards list five sources of validity evidence:

- Evidence based on test content
- Evidence based on response processes
- Evidence based on internal structure
- Evidence based on relations to other variables
- Evidence based on consequences of testing.

Rigorous analysis of NAPLAN according to these guidelines may have been undertaken but is not available publicly. In any case, it could not be concluded that NAPLAN exhausts the coverage of these two constructs. They represent only some aspects of literacy and numeracy. NAPLAN continues to be used as if it represents the entirety of what it means to be literate and numerate (which it clearly does not), and as if it is an adequate surrogate for all the desired outcomes of schooling (which it clearly is not). More attention needs to be given to public reporting on the quality and validity of these measures and cautions about over-interpretation. This ought to begin with clarification of what constructs are being assessed and how these align with the curriculum (Delandshere, 2002; Klenowski & Wyatt-Smith, 2014, p. 102).

There is also evidence that some of NAPLAN's consequences are detrimental to student wellbeing (APPA, 2013; Rice, Dulfer, Polesel & O'Hanlon, 2016; Wyn et al., 2014). Such consequences can be felt by students who are most 'at risk' (Cumming, Wyatt-Smith & Colbert, 2016) or Indigenous (Vass & Chal, 2016) or disadvantaged (Ng, Wyatt-Smith & Bartlett, 2016). Recommendations on these matters from the Australian Primary Principals Association (APPA) (2010) need to be heeded.

A recent study by LSIA has examined NAPLAN writing data, which show increasing numbers of students achieving below the national minimum standards in the domain of 'writing', while other aspects of literacy, such as reading comprehension, appear to be stable. We conclude that there needs to be a sharpened focus on writing standards as benchmarks in the national curriculum and increased attention to teaching writing effectively in every Year level. However, we note that there is currently no large-scale data on the teaching of writing in Australian classrooms and no longitudinal data on teachers' pedagogical practices. More concentration on teachers' pedagogical knowledge in developing students' writing skills is needed. This goes beyond simply measuring writing ability. It also goes beyond issues of teacher quality. It requires a multi-pronged approach for enhancing teacher knowledge and skill. Without support for developing pedagogical practice, the huge investment in NAPLAN is relatively impotent. We are currently developing an approach to sampling Year 9 writing skills (not tested in NAPLAN) and to a survey of writing instruction and the use of writing benchmarks. These kinds of studies can complement the existing databases and provide new evidence for a national evidence base on student learning outcomes (Wyatt-Smith & Jackson, 2016).

School-based data, such as grades, are problematic for a national database because of substantial comparability issues. There may be value in collecting a wider range of data from schools on learning outcomes, affective outcomes and values. However, without some form of moderation, it is unlikely that the data would be meaningful. Despite national curriculum guidelines, teachers would be assessing different constructs and applying different standards in assessing their students (Klenowski & Wyatt-Smith, 2014). An alternative is to systematically sample student work for analysis by a team of specially trained assessors. This would be far less intrusive, most able to produce valid data, and highly cost effective. Such studies could be commissioned on a rolling basis to cover key aspects of the national curriculum over time.

Measures of affective outcomes and values (including quality of life and wellbeing) could be obtained by survey sampling using questionnaires. Again, if teachers were required to make judgments about their students, training in the application of such measures would be needed, including moderation processes, to assure validity and comparability of the judgments. On the other hand, survey data could be collected directly from students and could have higher validity than teacher observation. There is a need to attend much more to students' own voices, and surveys on their dispositions and values would be a useful start in this direction.

5. LSIA advises distinguishing between the different costs and benefits of census, survey and institutional data

Census data are the most expensive to collect. They have the advantage of including everyone in the target population and therefore being comprehensive. They also allow for feedback to individual students and schools; though, as previously noted, such feedback is not often useful, or even if useful, not acted on (Klenowski & Wyatt-Smith, 2014). The disadvantage with census data is that there are time constraints on what can be asked of individual students, as well as constraints on the types of tasks (e.g., their complexity and authenticity). Further, much expense has to be devoted to administrative matters, security, and management of absences, illnesses, and unexpected events. Not least is the need to provide accommodations and modifications for students with special needs, such as students with various disabilities and recent migrants (along with the consequential interpretive issues) (Elliott et al., 2014).

A concern with current testing is that many students with special needs are not being included. The number of exemptions from NAPLAN is increasing. Ideally, all students should be included in the testing and the reasons for current exemptions should be researched with a view to universal inclusion (Cumming, 2012; Cumming & Dickson, 2013; Cumming & Maxwell, 2014; Cumming, Dickson & Webster, 2013; Elliott, Davies & Cumming, in press). Sample surveys of achievement allow the possibility of richer and more comprehensive assessment of learning outcomes. Assessments can be more complex and extended to cover

a range of learning outcomes that require more time and access to resources (such as complex problem solving, design and construction, communication and collaboration, information finding and synthesis). All students in a survey do not need to take the same tasks. The use of multiple matrix sampling allows sampling of both tasks and students, extending the range of possibilities for assessment even further, and reducing the amount of time any individual student is required to engage in an assessment. Accommodations and modifications are still needed for students with special needs, and there is a danger that they will be overlooked. But it is easier to include such special cases and make appropriate accommodations and/or modifications in their assessment.

An example of sample survey data collection is the National Assessment of Educational Progress (NAEP) <naep.ed.gov> in the USA. NAEP monitors a range of educational outcomes (subject matter achievement) using representative samples of students at critical stages of schooling (e.g. at years 4, 8 and 12, or at ages 9, 13 and 17) as well as a range of questionnaires of students, teachers and school administrators. NAEP reports the outcomes of its data analyses in the form of *The Nation's Report Card* as well as various reports, technical documentation, example assessment and questions, analytical tools, and analyses of trends. NAEP has had almost 60 years of experience (since 1969) and developed an impressive range of approaches and strategies to data collection, analysis and reporting. This expertise is worth emulating in Australia.

Institutional databases (such as attendance data and school policies) are typically 'dirty data' and require 'cleaning'. Dirty data are mainly missing data and wrong data (such as from coding errors on data entry, inconsistent use of categories and descriptors, misalignment of data columns, slippage from one computer system to another (Kim et al., 2003). Various techniques for data cleaning, data cleansing or data scrubbing are available. All require diligence in checking to see that the data are accurate and compatible. This is often an expensive and time-consuming part of any analysis of the data.

6. LSIA considers that accessibility and linkage are the most important and difficult issues to address

Linking different databases for analysis is not usually a simple matter. Various problems have to be overcome, such as data accessibility, data format, data alignment and data incompleteness (though sophisticated techniques now exist for inputting missing data statistically — Cumming and Goldstein, 2016). Agreement is also needed on ethical issues, such as individual or sector confidentiality. Accessibility currently differs across states and nationally, and for different databases, and needs more transparency and commonality.

Limited research use has been made of existing NAPLAN data at a national, or even state and territory level. Researchers at the LSIA are among the first to obtain full national data sets from the Australian Curriculum, Assessment and Reporting Authority (ACARA) as well as from the Queensland Curriculum and Assessment Authority (QCAA) in 2015. Analyses in 2015 and 2016 by Professor Joy Cumming and Professor Harvey Goldstein at LSIA have identified a number of issues in appropriate use and interpretation of these data.

First, as ownership of NAPLAN datasets reside with states and territories, ACARA has access only to cross-sectional data, of limited use in major educational research in this area, and longitudinal data sets for consecutive years of testing, that is, Year 3 to Year 5, Year 5 to Year 7, and so on. To create a national longitudinal dataset would require negotiation with each state and territory. Further state and territory data sets are not held in common formats. Clearly the longitudinal value of such data is access to national cohort longitudinal datasets.

The second issue that has arisen even in analyses to date is that different datasets tell different stories. The original dataset obtained from the QCAA, while for the full longitudinal cohort, was only for students who had undertaken all assessments, with large numbers of students missing non-randomly. Analyses of these data for Indigenous and non-Indigenous student comparisons yielded interesting outcomes for Year 3 to Year 5 growth. However, analyses of data for Queensland for the same purpose using the ACARA dataset, including all available student cases and with use of appropriate and new imputation technologies to address missing cases, yielded quite different outcomes (Cumming, Goldstein, Harris & Hand, in preparation). It is important therefore for limitations underpinning the datasets to be extremely clear for analysts, and for analysts to be sensitive to the appropriate use of datasets to create evidence that may inform future policy.

Attaching health and other databases to education databases, such as NAPLAN, is challenging because of the problems of matching individual records without a unique identifier. However, it is noted that statistical (probabilistic) matching has the potential to resolve the connection of data on individuals from different databases, such as LSAY, NAPLAN and PISA (Fowler, 2016) and there is further potential to link LSAY, NAPLAN, Medicare, Centrelink and other administrative collections (though not LSAY with LSAC) (Gemici & Nguyen, 2013). Exploring linking of previously unconnected data has considerable possibilities and benefits, but needs active support to realise its potential.

LSIA has developed a project proposal that illustrates the possibilities for linkage across databases. We propose to link NAPLAN individual achievement data with Pharmaceutical Benefits Scheme (PBS) and Medicare data related to chronic diseases. Currently, there are significant gaps in knowledge about the impacts of chronic diseases on children's academic achievements and development. We plan also to link with other datasets, such as school completion, workforce preparedness and care status. Issues such as data confidentiality and access will need to be resolved and their resolution will provide a prototype for additional studies. An ultimate aim of this research is to circumvent the need for intervention at later stages of schooling by early identification and monitoring of students with chronic diseases throughout their schooling. This will require additional resolution of access and usage of multiple databases for real time analysis. The realisation of these possibilities for database linkages will require administrative and funding support. The outcomes could have significant benefits, for individuals, families, communities and the national economy, through

improvements in individual learning outcomes and consequent employability and independence.

A further area where there is potential for data linkage is with teacher databases. There is now potential for analysing data from teacher education programs to entry to teaching to career development. A critical issue is why teachers leave teaching — an important issue since it represents loss of a key and costly resource. Entry data include teaching-entry literacy and numeracy, teaching performance and graduate attributes. Linkage with administrative and outcomes databases would enable a range of important studies on the career development of teachers, influences on teacher performance, and effects of student learning and other outcomes.

7. LSIA suggests that there is a need for more effective funding of educational research to inform policy

It is inescapable that a National Education Evidence Base will cost additional money. It is unclear how much it would cost to implement the approach we have suggested. Some organisational changes towards more effective access to databases may be absorbed into normal operating expenses in the relevant organisations. However, additional resourcing will clearly be necessary.

In thinking about resourcing, it would be useful to know how much Australia expends publicly on educational research, and how this compares with expenditure in other countries. It is difficult to obtain such information and LSIA is unable to offer such a comparison. However, it is possible to discover that the estimated expenditure on educational research in the UK is in the region of 9–10 million GBP <http://www.statista.com/statistics/298905/united-kingdom-uk-public-sector-expenditure-r-and-d-education/>.

We do not have a similar statistic to hand for Australia, but it is unlikely to compare favourably. It would, however, be useful to know how Australian expenditure of education research compares with that of other countries. The USA is probably an outlier on such a measure, since its educational research publications swamp those of other countries. Finding resources to improve Australia's attention to research and development in education would seem to be important for ensuring that research findings from other countries, such as the USA, are not applied inappropriately to the local context.

Summary

LSIA supports the development of a searchable National Education Evidence Base that includes evidence of various kinds and seeks to expedite linkages between various existing databases in education and related areas. This evidence base should be used to inform the public about the functioning of the education system by the transformation of the data and the interpretation of research as simple and succinct presentations. More importantly, however, a

National Education Evidence Base should service educational research that can inform educational policy and practice. It is unreasonable to expect that databases incorporated into a National Education Evidence Base could constitute a single large database because of the idiosyncratic problems of linkage between databases. Rather, emphasis should be placed on supporting access and linkage of various databases. The potential for analyses within and between existing databases to improve our understanding of educational outcomes and processes, and how to produce better outcomes and more effective processes, is largely unrealised but could be highly beneficial.

LSIA considers that a wider range of measures is needed to properly capture the expressed aims of the Australian education system and assess its health and quality. Existing measures, such as NAPLAN, canvass limited outcomes, have doubtful validity, and produce unfortunate consequences for the wellbeing of many students. Some redirection of the way they are used is needed. They also need to be supplemented with a range of other measures of desired learning outcomes and personal attributes. The collection of additional data does not need to be by census. A sampling approach affords a greater range and depth of data and offers improved cost-benefit.

Major attention needs to be given to creating easier and simpler access to existing databases and to assisting in creating linkages between them. Access currently is not always easy and not always complete, even for NAPLAN. Missing data and incomplete data are a huge problem, as is aligning data from different databases, and matching persons between databases and longitudinal. The cost and effort in dealing with such issues needs to be recognised and supported. High quality data analysis and educational research will not occur in the absence of sufficient funding and Australia would appear to be lagging in this respect.

Professor Claire Wyatt-Smith
Director, Learning Sciences Institute Australia

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (AERA/APA/NCME). (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Australian Primary Principals Association (APPA). (2010). *The reporting and use of NAPLAN*. Kathleen, ACT: APPA.

Australian Primary Principals Association (APPA). (2013). *Canvass report: Primary principals: Perspectives in NAPLAN testing and assessment.* Kathleen, ACT: APPA.

Centre for Educational Research and Innovation (CERI). (2007). *Evidence in education: Linking research and policy*. Paris: OECD.

Collins, S., Reiss, M., & Stobart, G. (2007). *The effects of national testing in Science at KS2 in England and Wales*. London; Institute of Education.

Cumming, J., & Dickson, E. (2013). Educational accountability tests, social and legal inclusion approaches to discrimination for students with disability: A national case study from Australia. *Assessment in Education: Principles, Policy & Practice*, *20*(2), 221–239.

Cumming, J., & Goldstein, H. (2016). Handling attrition and non-response in longitudinal data with an application to a study of Australian youth. *Longitudinal and Life Course Studies Journal, 7*(1), 53–63.

Cumming, J., Dickson, E., & Webster, A. (2013). Reasonable adjustments in assessment: Putting law and policy into practice. *International Journal of Disability, Development, & Education*, *60*(4), 295–311.

Cumming, J., Goldstein, H., Harris, K., & Hand, K. (in preparation). An analysis of educational progress of Indigenous and non-Indigenous Australian school students in Queensland from Year 3 to Year 5 in Reading and Numeracy: Different analyses, different policy implications. (In preparation for submission to *Comparative Education Review,* June 2016*.)*

Cumming, J., Wyatt-Smith, C., & Colbert, P. (2016). Students 'at risk' and the National Assessment Program Literacy and Numeracy (NAPLAN): The 'collateral damage'. In R. Lingard, G. Thompson, & S. Sellar (Eds.), *National testing and its effects: Evidence from Australia* (pp.126–138). London and New York: Routledge.

Cumming, J.J. (2012). *Valuing students with impairments: International comparisons of practice in educational accountability*. Dordrecht, The Netherlands: Springer.

Cumming, J.J., & Maxwell, G.S. (2014). Expanding approaches to summative assessment for students with impairment. In L. Florian (Ed.), *SAGE Handbook of Special Education* (2nd ed.) (pp. 573–596). London: Sage.

Dasu, T., & Johnson, T. (2003). *Exploratory data mining and data cleaning* (Vol. 479). John Hoboken, NJ: Wiley & Sons.

Delandshere, G. (2002). Assessment as inquiry. *Teachers College Record*, *104*(7), 1461–1484.

Dulfer, N., Polesel, J., & Rice, S. (2012). *The impacts of high stakes testing on school students and their families: An educator's perspective*. Sydney: Whitlam Institute, University of Western Sydney.

Elliot, S., Davies, M., & Cumming, J.J. (in press). Documenting support needs and adjustment gaps for students with disabilities: Teacher practices in Australian classrooms and on national tests. International Journal of Inclusive education. Available at: DOI:10.1080/13603116.2016.1159256

Elliott, S.N., Kettler, R.J., Beddow, P.A., & Kurz, A. (Eds). (2011). *Handbook of accessible achievement for all students*. Dordrecht: Springer.

European Communities. (2007). *Key competences for lifelong learning: European Reference Framework*. Luxembourg: Office for Official Publications of the European Communities.

European Parliament and European Union Council. (2006). Recommendation on key competencies for lifelong learning. *Official Journal of the European Union* (2006/962/EC).

Finnigan, K.S., & Daly, A.J. (2014). *Using research evidence in education: From the schoolhouse door to Capitol Hill*. Drodrecht: Springer.

Fowler, C. (2016). *One and one makes three: Getting smarter with data*. Adelaide: NCVER.

Gemici, S., & Nguyen, N. (2013). *Data linkage and statistical matching: Options for the Longitudinal Surveys of Australian Youth*. Adelaide: NCVER, LSAY Publications.

Gordon Commission on the Future of Assessment in Education. (2013). *To assess, to teach, to learn: A vision for the future of assessment (Technical Report)*. Princeton, New Jersey: The Gordon Commission.

Harvard Management Update. (2006). The use and misuse of statistics. *Harvard Management Update*, 11(3).

House of Commons Children, Schools and Families Committee. (2008). *Testing and assessment: Third report of session 2007–08, Volume II*. London: Parliament, House of Commons.

Huff, D. (1954, 1991). *How to lie with statistics*. New York: Norton; Penguin (new edition).

Kim, W., Choi, B.J., Hong, E.K., Kim, S.K., & Lee, D. (2003). A taxonomy of dirty data. *Data mining and knowledge discovery, 7*(1), 81–99.

Klenowski, V. (2016). Questioning the validity of the multiple uses of NAPLAN data. In B. Lingard, G. Thompson, & S. Sellar (Eds.), *National testing in schools: An Australian assessment* (pp. 44–56). London and New York: Routledge.

Klenowski, V., & Wyatt-Smith, C. (2014). *Assessment for education: Standards, judgement and moderation*. Los Angeles: Sage.

Koretz, D. (2008). *Measuring up: What educational testing really tells us*. Cambridge, MA: Harvard University Press.

Lemke, J., & Sabelli, N. (2008). Complex systems and educational change: Towards a new research agenda. *Educational Philosophy and Theory*, 40(1), 118-129.

Lingard, B., Thompson, G., & Sellar, S. (Eds.). (2016). *National testing in schools: An Australian assessment*. London and New York: Routledge.

Maroulis, S., Guimera, R., Petry, H., Stringer, M.J., Gomez, L.M., Amaral, L.A.N., & Wilensky, U. (2010). Complex systems view of educational policy research. *Science, 330*, 38–39.

Ministerial Council for Education, Early Childhood Development and Youth Affairs (MCEEDYA). (2008). *Melbourne Declaration on Educational Goals for Young Australians*. Melbourne: MCEEDYA (now SCSEEC: Standing Council on School Education and Early Childhood).

Ng, C., Wyatt-Smith, C., & Bartlett, B. (2016). Disadvantaged students' voices on national testing: The submersion of NAPLAN's formative potential. In R. Lingard, G. Thompson, & S. Sellar (Eds.), *National testing and its effects: Evidence from Australia* (pp.152–163). London and New York: Routledge.

Nutley, S., Walter, I., & Davies, H. (2007). *Using evidence: How research can inform public services*. Bristol: Policy Press.

Organisation for Economic and Cultural Development (OECD). (2005). *The definition and selection of key competences: Executive summary*. Available at: http://www.oecd.org/pisa/35070367.pdf

Pelligrino, J.W., & Hilton, M.L. (Eds). (2012). *Education for life and work: Developing transferable knowledge and skills in the 21st century*. Washington, DC: The National Academies Press. Available at: www.nap.edu

Rice, S., Dulfer, N., Polesel, J., & O'Hanlon, C. (2016). NAPLAN and student wellbeing: teacher perceptions of the impact of NAPLAN on students. In R. Lingard, G. Thompson, & S. Sellar (Eds.), *National testing and its effects: Evidence from Australia* (pp. 72–85). London and New York: Routledge.

Science Community Representing Education (SCORE). (2009). *Government abolishes Key Stage 2 Science Sats*. Available at: http://www.score-education.org/news/press-releases-and-media-coverage/abolishment-of-ks2-sats

Senate Standing Committee on Education and Employment. (2014). *Effectiveness of the National Assessment Program — Literacy and Numeracy (Final Report).* Canberra: Commonwealth of Australia.

Strayhorn, T.M. (2009). Accessing and analyzing national databases. In T. Kowalski, & T.J. Lasley (Eds.), *Handbook of data-based decision making in education* (pp. 105–122). New York: Routledge.

Tufte, E. (1983). *The visual display of quantitative information*. Cheshire, CT: Graphics Press.

Vass, G., & Chalmers, G. (2016). NAP{LAN, achievement gaps and embedding Indigenous perspectives in schooling: Disrupting the colonial option. In R. Lingard, G. Thompson, & S. Sellar (Eds.), *National testing and its effects: Evidence from Australia* (pp. 130–151). London and New York: Routledge.

Voogt, J., & Roblin, N.P. (2012). A comparative analysis of international frameworks for 21st century competences: Implications for national curriculum policies. *Journal of Curriculum Studies*, *44*(3), 299–321.

White, J., & Rosauer, K. (2015). *Young Australians, illness and education: Report on the national database project*. Melbourne: The Victoria Institute. Available at:

https://www.vu.edu.au/sites/default/files/victoria-institute/pdfs/young-australians-illness-and-education-web.pdf

Wyatt-Smith, C., & Jackson, C. (2016, forthcoming). NAPLAN data on writing: A picture of accelerating negative change. *Australian Journal of Language and Literacy*.

Wyn, J., Turnbull, M., & Grimshaw, L. (2014). *The experience of education: The impacts of high stakes testing on school students and their families: A qualitative study*. Sydney: The Whitlam Institute, University of Western Sydney. Available at: https://www.whitlam.org/__data/assets/pdf_file/0011/694199/The_experience_of_education_-_Qualitative_Study.pdf