Impact and Consequence of
Australia's National Assessment Program - Literacy and Numeracy (NAPLAN)
– Using Research Evidence to Inform Improvement

Submission to Australia's Senate Committee Inquiry into the
Effectiveness of Australia's National Assessment Program - Literacy and Numeracy (NAPLAN)

SUBMISSION WRITTEN BY

Professor Pauline Harris, Professor Mohan Chinnappan, Professor Geraldine Castleton,
Dr Jenni Carter, Dr Michele de Courcy & Dr Jenny Barnett

with and on behalf of the School of Education, University of South Australia

Citing this paper:

Harris, P., Chinnappan, M., Castleton, G., Carter, J., de Courcy, J. & Barnett, J. (2013) Impact and consequence of Australia's National Assessment Program - Literacy and Numeracy (NAPLAN) – Using research evidence to inform improvement. *Submission to Australia's 2013 Senate Committee Inquiry into the Effectiveness of the National Assessment Program - Literacy and Numeracy (NAPLAN),* with and on behalf of School of Education, University of South Australia. [Insert URL and date retrieved]

Executive Summary

This paper draws on an extensive range of research literature that informs our response to each Term of Reference for the Australian Government's Senate Committee Inquiry into the Effectiveness of the National Assessment Program - Literacy and Numeracy (hereafter referred to as NAPLAN).

1. Contributions of NAPLAN are acknowledged in terms of its stated objectives and ensuring literacy and numeracy as key national priorities. There are, however, significant, research-evidenced issues that are described in this paper, which constrain NAPLAN's fulfillment of its stated objectives. These issues must be addressed in any future iterations of NAPLAN if it is to realise its framing vision.

2. NAPLAN's impact on teaching and learning practices, intended or otherwise, include narrowing of curriculum and pedagogic choices; encouraging practices that focus on measurable learning outcomes; marginalising higher-order thinking that in turn lowers curriculum standards; emergence of commercial instructional materials and tests that are not quality assured; and deleterious pushdown effects in the prior-to-school and early school years.

3. In addition to these impacts, NAPLAN has had a number of unintended consequences related to the use and misuse of NAPLAN data, which further constrain its effectiveness and need to be redressed in future iterations. These consequences include pressure on school leaders to raise student performance, with threat to their jobs if this is not achieved; incidences of cheating in tests; greater attention given to students likely to achieve well, with increased absenteeism of low-achieving students on test days; and diminishment of teacher professionalism and morale, particularly devaluing teachers' judgments, knowledge and voice.

4. MySchool website's impact on teaching and learning practices is problematic, given the distancing of generalised and abstract standardised test results from classroom realities. While the MySchool website might be seen to be useful for parents, it distorts a focus on numbers and comparisons; and exacerbates tensions between teachers' and schools' broad policy conditions such as NAPLAN, and their need to engage with students in their local situations in terms of myriad factors including students' own strengths, aspirations and predispositions.

5. A number of essential improvements to Australia's national literacy and numeracy assessment program is indicated by the research-evidenced issues outlined in this paper. NAPLAN's stated objectives related to enhanced student outcomes and greater teacher and school accountability frame these suggested improvements. Specifically, it is recommended that a national literacy and numeracy assessment program:

    - Be framed by comprehensive models of literacy and numeracy as defined in the Australian Curricula for English and Mathematics;
    - Is shaped by a robust and comprehensive framework of trustworthiness (validity and reliability);
    - Duly acknowledges and develops teachers' professional capacity to gather multiple sources of evidence and to make informed, contextualised judgments about their students' learning and achievement;
    - Involves teachers in the development, design and marking of the national assessments;
    - Considers all students' needs in the design of testing and other data-gathering procedures, including cultural and linguistic considerations;
    - Engages with participants' and stakeholders' expectations and understandings;
    - Provides schools with capacity to challenge inferences drawn from NAPLAN testing and to provide evidentiary data and contextual information to contribute to the interpretation of test data;
    - Provides community education about the purposes and limitations of the national assessment program; and
    - Does not stigmatise schools that are seen to be under-achieving and does not in any way publish data that puts the professional dignity and wellbeing of teachers, students and schools at risk when such data are (mis)used by the media and other parties (e.g., league tables).

We now set out below our complete, research-evidenced response to each Term of Reference for this Inquiry.

Terms of Reference

a) Whether the evidence suggests that NAPLAN is achieving its stated objectives

It is clear that NAPLAN has helped put literacy and numeracy education in the spotlight as key national and state educational priorities. Understanding that NAPLAN's objectives relate to driving improvements in student outcomes and providing increased accountability for the community, it may be said that NAPLAN has made a contribution to providing schools with data to analyse progress; school leaders with a sense of trends occurring in their school that can inform program and policy decisions; and governments with regional data that can inform how to best support and resources areas of strength and need (Dooner, 2001).

It is also clear that national assessment programs such as NAPLAN create opportunities for meaningful exploration of teaching and learning practices and for reflective dialogue and robust discussion about the nature and purposes of assessment and its relationship to teaching and learning practices and students' lives, now and in the future.

Practices such as NAPLAN also compel us to respectfully acknowledge and reflect on the professional tensions teachers and schools experience in being publicly accountable to external demands based on a generalised notion of '*all* students', while meeting their professional and highly contextualised responsibilities to *their particular* students (Dreher, 2012; Hardy, 2013; Kostogriz & Doecke, 2013).

Less clear is the evidence that NAPLAN is achieving its stated objectives that it is benefiting *all* Australian schools and school students; and that NAPLAN is informing the allocation of education resources in ways that ensure *all* students achieve worthwhile learning.

Such achievement is at best unclear because there are a number of key issues impacting the effectiveness of NAPLAN as a national assessment program. These issues must be acknowledged and carefully addressed if NAPLAN is to realise its framing vision.

These issues, borne out in the research literature, are summarised below and elaborated upon in what follows:

- NAPLAN's limited coverage of content and skills and the time allocated for sitting the test.
- The need for other sources of evidentiary data, including data gathered by teachers in a knowledgeable and principled way, to inform practices that improve learning outcomes.
- Limitations of NAPLAN as a non-diagnostic assessment procedure for informing improved student outcomes.
- Validity issues related to attributing students' test scores to school performance and teaching effectiveness.
- Cultural and linguistic appropriateness and accessibility of NAPLAN's content.

NAPLAN covers a limited range of content and skills that assess fragments of a student's achievement, and so is not representative of the whole literacy and numeracy domains (Wu & Hornsby, 2012). Like other such tests, this fragmentation limits the usefulness of

the data in helping teachers and parents make sense of students' results and how they might inform teaching and learning practices (Boaler, 2002).

For example, NAPLAN numeracy items do not examine children's intuitive and formal understanding of patterns in their life, yet the ability to recognise and use patterns is a prerequisite for development of numeracy. While NAPLAN assesses basic understandings, it is limited in value in relation to long-term learning trajectories that are better supported by learning strategies that prompt students to reflect, revise and construct deeper understandings that are critical for transfer of knowledge. Students' responses to open-ended tasks are more effective because such tasks provide opportunities for students to exhibit alternative representations of a given problem.

Further, while NAPLAN is a test of basic literacy skills, the actual use of such skills is not divorced from other literate practices. The question we need to consider is, why have these skills been selected for particular focus, and not others? This question is significant when one considers that in the fields of literacy education policy and practice, we assess what we value and value what we assess. This relationship between assessment and value has profound implications such as struggles in classrooms and schools over what really counts (Hardy, 2013), and narrowing curriculum and pedagogic choices to what is being nationally prioritised and assessed (Klenowski & Wyatt-Smith, 2012).

As defined in the Australian Curriculum: English (ACARA, 2010), literacy is constituted by a complex of practices – it includes fundamental decoding and encoding practices, as well as meaning-making practices, critical literacy practices, and practices for using texts for real-life purposes in context. These practices work in synergy with one another and not in isolation, and are shaped by the sociocultural settings in which they are deployed, as extensively documented and long understood in the literature (Harris, Turbill, McKenzie & Fitzsimmons, 2006; Freebody, 2007; Luke & Freebody, 1999).

Furthermore, literacy is not print-based alone – multiple literacies continue to emerge in students' day-to-day lives that involve sophisticated forms of technology. This is particularly the case in many students' out-of-school experiences, and increasingly so in schools too.

NAPLAN's limited and limiting view of literacy and its print-based, basic skills focus, is therefore out of step with many students' present and future lives – thereby providing a limited and limiting experience for both the students and the purposes and objectives of this assessment program.

These shortcomings are exacerbated by test time allocations that do not always allow students to fully show what they know. Carter (2012) for example, found that on average, 17% students ran out of time in the 2010 NAPLAN tests.

Accurate interpretation of a student's test results is further limited without more fully knowing the student and without other sources of evidentiary data, including drawing on data gathered by those who know students' learning strengths and needs best – teachers and the students themselves (Anstey & Bull, 2003; Wu & Hornsby, 2012).

Moreover, having test responses alone is not equal to assessing how a student is achieving according to particular standards. As McKay (2000, p. 193) reminds us in an ESL context,

> 'ESL standards are not tests; they are dependent for their validity on the uses and interpretations that we make of the "test response", rather than on the rating which is elicited from the combined use of the standards and the assessment activity used to observe a student's performance.'

Despite NAPLAN as it currently stands being a test of basic skills in literacy and numeracy, many NAPLAN advocates claim that its results can be used to inform classroom practice. However, in order to help students achieve improved outcomes, teachers need an assessment tool that provides specific information about individual learners' strengths and weaknesses in a comprehensive range of literacy and numeracy practices. NAPLAN is not a diagnostic test, given its large-scale application to large groups of students to extrapolate how schools, not students, are performing. As such, NAPLAN does not provide adequate information on areas of individual strength and weakness to inform improved outcomes (Wu & Hornsby, 2012).

While testing in schools is important to informing teacher understanding of student knowledge, it is only one kind of data that needs to be triangulated with other sources of information; and the research evidence indicates that testing does not need to be standardised and centralised. The usefulness of NAPLAN for informing teaching and learning practices to improve student outcomes is sorely limited by the time between the test and releasing the results so that teachers and schools might act on them (Doecke, Kostogriz & Illesca, 2010; Saubern, 2010). This time lag between testing and teaching does provide teachers with reliable information about what students know at the time of teaching because students' knowledge invariably changes as a consequence of responding to test items. For instance, a students' understanding of place value in a numeracy item may be different at the point of the test and at the time of teaching. Thus, using NAPLAN results to assist students' learning may have negative consequences for the students and their learning outcomes.

This problem underscores the disconnection that research has shown exists between NAPLAN and teachers. This disconnection is exemplified in the absence of teacher presence or teacher-sourced data on the MySchool website, despite teachers being the ones who spending the most amount of time in a day interacting with the children and young people whose achievement NAPLAN purports to be measuring (Doecke *et al*, 2010).

The validity of using students' test results to infer how schools are performing is problematic. As is extensively and rigorously borne out in the research literature, many factors influence student learning – for example, socio-economic conditions and poverty; sociocultural and linguistic factors; individual predispositions, aspirations and learning needs; out-of-school learning experiences; peers, family and community factors and influences; to name but a few (Baker, Barton, Darling-Hammond, Haertel, Ladd, Linn, Ravitch, Rothstein, Shavelson & Shepard, 2010; Delpit, 2003; Harris, 2010; Jorgenson, 2010; Ladwig, 2010).

Combined with idiosyncratic factors that influence student test performance on a given day, these factors create a considerable margin of error in how performance reflects actual student competence if they are not taken into account (Wu, 2010). Children's stress and wellbeing are key considerations, too. Research evidence in the United Kingdom, for example, has shown problems with increased stress levels arising from

mandatory standardised testing, and consequences of this stress for diminishing students' test performance and contributing to a margin of error in test results reflecting student competence (Connor, 2001, 2003).

A similar issue arises with any misuse of NAPLAN data to judge teaching effectiveness. Attributing low performance to school performance or teaching effectiveness is neither valid nor reliable (Wu & Hornsby, 2012), even when value-added models of analysis are designed and used to compare schools with similar contextual factors (Baker *et al*, 2010).

Koch and DeLuca (2012), in their careful critique of high-stakes assessment practices, examine NAPLAN's test validity, pointing out that interpreting and using large scale assessments for purposes other than those for which they were designed, by multiple stakeholders and in other contexts, is neither reasonable nor valid. Because of (mis)uses of such assessments, Stobart (2009, cited in Koch & DeLuca, 2012) recommends a framework for validity that considers validity in terms of the purpose of the particular assessment/testing procedure; its fitness for purpose; its reliability; parameters for its interpretation of results; and its likely impact (intended and otherwise).

Finally, any test's content is culturally laden and NAPLAN is no exception. Consequently, NAPLAN's content is not always familiar to students. For Indigenous students living in remote communities, for example, this means content and concepts are not always accessible or relevant to them (such as test items centred on newsagency shops or cinemas that are not features in remote communities); and with many Indigenous students living in linguistically complex communities, NAPLAN actually masks their linguistic, cultural and literate competences (Wigglesworth, Simpson & Loakes, 2011). Research studies have revealed many factors in Indigenous communities that enable and constrain students' learning and performance. For example, Jorgensen's study (2010) evidences a compelling case for considering Issues of social equity in Indigenous students' access and success in mathematics learning in remote communities. Sporadic attendance, understood in terms of these students' home and community lives, never the less impacts teacher morale and student knowledge and thus lowers test results.

On the matter of language, the language and literacy demands of tests have been found to be inhibitors to test completion and achievement, thereby creating significant access and equity issues. For example, Hipwell and Klenowski (2011)'s literacy audit of assessment tasks in NAPLAN indicates what they term the 'silent assessors' – that is, the literacy demands of the assessment itself and the identity work that is implicit within assessment as a value-laden activity. Further, these researchers identify how the tests focus on individual literacy and are undertaken in conditions where students are not able to deploy and display the literacy practices that they are accustomed to using in their school, home and community lives – such as engaging in the social classroom practices of literacy work, engaging with environmental print, and being involved in collaborative learning.

Also in relation to language matters in NAPLAN, language used in mathematics tests has been found to be an inhibitor sometimes in test completion and achievement. This issue is particularly relevant to NAPLAN numeracy items that involve word problems where the critical phase of the solution attempt is to translate the text embedded in the problem to a mathematical representation. Logan and Greenlees (2008), for example, demonstrated that language challenges can have a significant impact on students'

comprehension of graph-related mathematical problems. In the case of Indigenous students, there is a further challenge in that both linguistic and visual representations of numeracy items are difficult to decode (Klenowski & Gertz, 2009; Warren & Miller, 2013). While highlighting the need for the difference between mathematical English and everyday English to be explicitly explored in classrooms, issues of linguistic accessibility for students obscure their mathematical capabilities and so distort test scores and their interpretations (de Courcy & Burston, 2000; Jorgenson, 2010; Morley, 2011; Quinnell & Carter, 2011). As stated most clearly below,

> '"any test that employs language is, in part, a measure of [test takers'] language skills. This is of particular concern for test takers whose first language is not the language of the test." [1] … In short, when English language learners take a mathematics or science test, it is not clear the degree to which the results reflect their knowledge of English instead of their content-area knowledge.' (TESOL Inc, 2005, p. 1).

Bringing linguistic and cultural factors together in our consideration here, the danger inherent in high-stakes standardised testing implemented on a large national scale, such as NAPLAN, is that their content presents a standardised view of student identities and the world, which contributes to inequality in student access to and demonstration of knowledge (Au, 2009). The limited ways for explaining why some students do not do well or why some schools have large numbers of students who do not do well, in terms of teacher quality, teaching practice and learning difficulties, ignore the longstanding research that provides social and cultural explanations for difference – such as the research we've included here, a small sample of a wealth of rigorous, highly regarded body of research that conveys similar messages and understandings.

b) Unintended consequences of NAPLAN's introduction

Negative consequences in classrooms of the (mis)use of national test results to evaluate teacher effectiveness include discouraging teachers from working with the neediest students and schools; diminishment of teacher voice and morale; narrowing curriculum and pedagogic practices and teaching to the test; encouraging practices that aim at learning outcomes that can be measured; marginalising open-ended problem solving that in turn lowers curriculum standards; and a reduction in collegiality that sees what has been called the privatisation of classrooms as teachers increasingly teach behind closed doors (Baker *et al*, 2010; Hattie, 2005; Klenowski & Wyatt-Smith, 2012; Wilson, 2012; Wilson & Hornsby, 2012).

Evidence in the UK and USA is clear on these negative consequences of national testing and high-stakes accountability (Barksdale-Ladd & Thomas, 2000). In Australia, the Australian Primary Principals Association (APPA) have identified NAPLAN's consequences at the school level, in terms of pressure exerted on or felt by leaders to raise student performance, with threat to their jobs if this is not achieved; incidences of cheating in tests; and greater attention given to students likely to achieve well, with increased absenteeism of low-achieving students on test days (Klenowski & Wyatt-Smith, 2012).

Also of concern is students' wellbeing in terms of stress brought to bear by these pressures and by some students' own test anxiety (Connor, 2001, 2003). While stress

has been found to diminish test performance for some students, enhancing students' sense of wellbeing has been found to enhance academic performance and learning (Dix, Slee, Lawson & Keeves, 2012) – and improved learning outcomes is, after all, one of NAPLAN's two stated objectives.

Broader concerns rest with the implementation of high-stakes testing in Australia and the impact this is having on education, in terms of the move to see student achievement serving economic imperatives that distorts the learning experience and an authentic understanding of student learning and achievement; and an emphasis on standardised testing as a more reliable form of judgment than the judgments of those who work most closely with students – teachers (Klenowski & Wyatt-Smith, 2012).

In regard to teachers and teacher judgment, teachers' sense of their own professional identities has declined and their professional judgments brought into question in the face of an increasingly top-down policy regime (Cormack & Comber, 2013) – NAPLAN takes a view of teacher and learner alike that is disempowering and does not value their knowledge and voice (Wilson, 2012).

Yet, It is well documented in the research literature that teachers' well-informed and discretionary professional judgments and decisions are critical to the effectiveness of their programs (Dudley-Marling, 2005; Pearson, 2007). This research highlights the importance of sustained investment in teachers' professional development and support to ensure their capacity for gathering and interpreting evidence of students' learning and making well-informed judgments as part of any national assessment program and in the local contexts where they work with their students.

c) NAPLAN's impact on teaching and student learning practices

As previously described, the usefulness of NAPLAN for informing teaching and learning practices to improve student outcomes is sorely limited by the time between the test and releasing the results so that teachers and schools might act on them (Doecke, Kostogriz & Illesca, 2010; Saubern, 2010). Nor is NAPLAN designed as a diagnostic test, and therefore should not be used as such. Indeed, as we previously argued, given the time lag between test and results, using NAPLAN results to assist students' learning may have negative consequences for the students and their learning outcomes.

A number of other concerning impacts on teaching and learning practices have been identified by the Australian Primary Principals Association (APPA), including what we have previously identified as unintended consequences – these include narrowing the curriculum; neglecting curriculum areas in favour of literacy and numeracy; neglecting higher-order thinking skills that are not subjected to NAPLAN testing; and the rise of a commercially driven education sector, including the emergence of commercial instructional materials and tests that are not quality assured (Klenowski & Wyatt-Smith, 2012; see also Cormack & Comber, 2013, who found similar results in their rural school study).

NAPLAN's effects are even being felt in the prior-to-school years in ways that are not appropriate for young children and how we know they learn and engage. What have been described as 'pushdown' and 'hothousing' effects in the prior-to-school and early school years have seen an inappropriate use of formalized basic skills instruction and

commercial teaching/learning packages that disembody literacy skills from real and meaningful acts of literacy. Such approaches alienate the young child from literacy engagement because they do not respect and nurture the voice, agency and participation of the child, and they disregard what is essentially children's work and approach to their world - their play (Sylva, Melhuish, Sammons, Siraj-Blatchford, Taggart & Elliot 2008; Winter, 2010).

d) The impact on teaching and student learning practices of publishing NAPLAN test results on the MySchool website

Hardy and Boytle (2011), in their critique of standardised tests' abstraction and quantification in education, make the very good point that education needs to be understood from the notion of 'practice' as complex social activity – and that to understand this practice one has to be engaged in the practice. That is, to understand and therefore be in a position to judge the effectiveness of a system, one must be a participant in that system.

We concur with this position. Consequently, we, as do these authors, question the validity of quantified results of social practice (education) when this quantification distances the students (and indeed their teachers and contexts) from the data. If tests are to be used to mandate structural changes in schools and communities, fuelled by the publication of test results on the MySchool website, then this overly magnifies a focus on numbers and comparisons; and exacerbates tensions between teachers' and schools' broad policy conditions such as NAPLAN that emphasise the abstract and the general (such as NAPLAN test results), and teachers' engagement with what is specific and locally situated at 'the chalkface' – students and their local circumstances that shape their ongoing, authentic, connected teaching and learning experiences and outcomes across wide-ranging whole class, small group and individual situations (Hardy, 2013; Kostogriz & Doecke, 2013).

These tensions further highlight the importance of professional judgment in resolving tensions in ways that serve students' best interests and outcomes, which surely are key elements of the Government's broader objectives in this national assessment space. Moreover, if teachers and schools are to be held accountable through means such as NAPLAN and are expected to work with the tensions that result, then NAPLAN providers need to show how they are held accountable to schools, teachers and students in how they design the tests and interpret and publish the results, and work with teachers and schools to resolve tensions that arise.

e) Potential improvements to the program, to improve student learning and assessment

In light of the concerns we have taken time and care to outline in this paper, drawing on the research evidence at hand, and given the significance of NAPLAN to the Government's objectives related to enhancing student outcomes and improving teacher and school accountability, we have a number of constructive suggestions to make.

Effective assessment of student learning – that is, assessment to inform teaching and learning practices and support and foster student learning and engagement – makes students' learning visible (Latham, Neville & Semple, 2012). Effective assessment

documents growth in student learning over time and gauges authentic transfer of learning, by incorporating information about context and process inherent in the learning experience, as well as products or outcomes of that experience. Effective assessment draws on a range of evidentiary data that make learning visible to all stakeholders and provides timely and ongoing feedback to teachers and students to inform program decisions and practices. It includes and validates students' voices and reflections; and it includes and validates the role of the teacher in the assessment process.

Whether or not such an approach to assessment is possible or even desirable at a national level is highly questionable. But if assessment is not shaped this way, then the question to be asked is, why do it?

Thus we recommend that a national assessment program:

- Is framed by a comprehensive model of literacy and a comprehensive model of numeracy, that takes account of all the practices that are involved, the synergies among these practices, and includes not only print-based but also multi-media experiences. Such framing would reflect the multiple realities of literacy and numeracy in students' lives and education and society at large, and so endow the assessment program with some degree of ecological validity.

- Is consistent with ways in which literacy is defined in the Australian Curriculum – English (ACARA, 2010); and therefore encompasses skills and knowledge to listen to, view, read, speak, write and create a growing repertoire of texts. These skills include comprehending, interpreting and creating spoken, written and multimodal texts; using the English language accurately, fluently, critically, creatively and confidently; manipulating and adapting language to meet the demands of more general or more specialised purposes, audiences and contexts; and understanding the different ways in which knowledge and opinion are represented and developed in texts, and about how more or less abstraction and complexity can be shown through language and through combinations of language and visual representation.

- Is consistent with the Australian Curriculum – Mathematics (ACARA, 2010), which has identified four process elements with which to engage students in the core content strands. These elements, referred to as Proficiency Strands, include understanding, fluency, problem solving and reasoning. Future NAPLAN items could place more emphasis on the elucidation of students' ability to make active use of these four proficiency strands.

- Is shaped by a robust framework of trustworthiness in terms of validity, clarity of and fitness for purpose, reliability, and explicit parameters for defining and limiting interpretation of results.

- Duly acknowledges and indeed further develops the professional capacity of teachers to gather multiple sources of evidence and make informed, contextualised judgments about their students' learning and achievement; and that these judgments are included to inform the interpretation of NAPLAN results. (See also Ladwig, 2010, who explores similar ideas.) In so doing, care would need to be given to minimising additional administrative burden this action might create for teachers

and schools; and appropriate investment needs to be made to ensure all teachers are equipped and supported to make such judgments.

- Critically involves teachers in the development, design and marking of the national assessments; and draws on multiple sources of evidence, including evidence gathered by teachers, not national tests alone, to achieve the objectives of the national assessment program (see also Klenowski & Wyatt-Smith, 2012).

- Considers needs of all students in how the national assessment program and its testing and other data-gathering procedures are designed, including the kinds of cultural and linguistic considerations we have identified in this paper. With respect to students learning English as an additional language or dialect (EAL/D), we have a number of specific suggestions:

    - Ensure the test data are disaggregated from the broader category of children from language backgrounds other than English, to reveal EAL/D learners' performance. This would enable nationally consistent data on EAL/D learners' performance and progress without which it is difficult, if not impossible, to monitor the impact of the NAPLAN assessment and reporting regime on EAL/D learners' educational outcomes. It would also assist teachers in interpreting NAPLAN data for their classes, and make available data that would clarify the performance of schools with significant numbers of EAL/D.

    - Carefully check NAPLAN test items for cultural and linguistic bias towards 'mainstream' or western-cultural test takers since these often prevent EAL/D learners from freely demonstrating their knowledge. The choice of texts, prompts and the wording of questions should not limit the opportunities for EAL/D students to demonstrate their understanding of the English language or of numeracy.

    - Further develop NAPLAN to provide data on language errors made by those learning English as an additional language or dialect.

    - Encourage and support teachers of EAL/D learners not to focus on the NAPLAN at the cost of attending to the development of students' overall English language development, notably with regard to listening and speaking skills, which are central to the development of English language proficiency, and not part of the NAPLAN.

    - Numerically represent as much as possible the numeracy component, rather than using word problems which increase the linguistic and cognitive demands on students and sometimes prevent EAL/D students from demonstrating their knowledge.

    - Consider constructing an alternative national assessment of English language proficiency for EAL/D learners, drawing on research into the diverse assessments across the states and territories. (This has been recommended by the House of Representatives Report, *Our Land, Our Language* - Recommendation 15).

- Develop an online adaptive version of NAPLAN and provide opportunity for students to take the test in time periods that would allow for the fact that EAL/D students are either mentally translating the test or taking time to process it in the new language.

- Provide professional development and support for teachers to learn how to mine the data to see what it is saying about EAL/D learning.

– Develops assessment items for the numeracy component that assess both procedural and conceptual knowledge of Mathematics that underpins development of numeracy. Additionally, item developers need to consider different ways students could express their answers. We also recommend the use of open-ended tasks to capture students' cognitions.

– Engages with participants' expectations and fosters their roles in terms of students understanding the assessment process and feedback and what they can do to improve; parents understanding the assessment approaches and their particular purposes, limitations and interpretations in terms of their own children; and teachers enhancing their skills in developing assessment procedures suitable to student needs based on national curriculum standards that inform and guide assessment.

– Provides schools with capacity and ability to challenge inferences drawn from NAPLAN testing, as recommended by the Australian Primary Principals Association (Klenowski & Wyatt-Smith, 2012); and indeed to provide their evidentiary data and contextual information to contribute to the interpretation of test data, thereby recognising the complexity of influences on school results in NAPLAN testing.

– Provides community education about the purposes and limitations of the national assessment program, and communicates appropriate ways in which the data may be interpreted and ways in which it may not.

– Does not stigmatise schools that are seen to be under-achieving and does not in any way publish data that put the professional dignity and wellbeing of teachers, students and schools at risk , particularly when such data are (mis)used by the media and other parties (e.g., league tables). Such misappropriation is unacceptable and measures to prevent it from happening must be taken. To this end, the MySchool website needs to be re-considered from the viewpoints of all individuals who are represented in the data *albeit* anonymously – naming schools compromises individuals' anonymity and, in the case of low performance scores, diminishes their sense of identity and worth.

f) International best practice for standardised testing, and international case studies about the introduction of standardised testing

If we are to learn from other nations' assessment practices, then we need to take a long hard look at what has proven to be exemplary and highly effective – and understand the contexts in which these practices have developed, and relate context and practice to what is achievable, desirable, and fit for the purpose in and across Australia's multi-layered, richly diverse contexts.

Finland, for example, is frequently upheld as an exemplary case in student outcomes, given its recurrently high PISA (Programme for International Student Assessment) results. Yet Finland has achieved this success without implementing a high-stakes accountability regime (Sahlberg, 2011). This success is attributable to many factors, including ones deeply embedded in Finland's social, political and cultural contexts. We are not arguing that Australia could or should transplant their practices. But we should pay heed and come to understand what we in Australia might learn from how Finland achieved success in terms of their recognition and development of teachers' autonomy, development of instructional leadership in schools, and enhancement of trust in teachers and schools supported by investment in continuing to grow and support teachers' professionalism – as we ourselves have argued in this paper.

Learning from international experiences also requires us to scrutinise what has not worked. Finland's experience stands in sharp contrast to the USA, where research continues to document the continued failure and associated problems of its high-stakes accountability regime (Darling-Hammond, 2004, 2010; Hursh 2007; Menken 2008; McNeil 2000) – a regime that Australia in no small way has emulated in its own national assessment program, NAPLAN. This US regime has been found to engender competition, diminish teacher voice and morale, deprofessionalise the teaching profession, mistrust teachers and schools, and narrow teacher responsibility to test-based accountability that in turn has narrowed curriculum and pedagogic choices. According to Linda McNeil (2008), Director of the Center for Education at Rice University:

> 'High-stakes, test-based accountability doesn't lead to school improvement or equitable educational possibilities. It leads to avoidable losses of students. Inherently the system creates a dilemma for principals: comply or educate. Unfortunately we found that compliance means losing students.'[3]

These problems in the USA notwithstanding, more successful outcomes there have been found in US states and districts that have focused on broader notions of accountability, including investments in teacher knowledge and skill, organisation of schools to support teacher and student learning, and systems of effective assessment, such as what we have described here in this paper, that can genuinely inform curriculum reform and teaching improvements (Darling-Hammond, 2010).

Immediately to the north of the U.S.A., it is interesting to note that in the Canadian province of Alberta, the Education Ministry is phasing out its benchmark province-wide achievement tests for students in favour of what it describes as more comprehensive computer-based exams to better assess a student's creativity, critical thinking and problem-solving, and not just test for core knowledge in numeracy and literacy.[2] The change in approach has been described as more student-centred and focused on informing learning, while still being able to derive comparative benchmarks as a province. These tests will still be administered to students in Grades 3, 6, and 9, as was the case with the previous testing regime. However, the tests will be undertaken at the start rather than the end of the school year, so that timely feedback can be provided and stigma for schools is removed.

Closer to home, New Zealand is an interesting case to consider. Their national standards program[4] was introduced in 2009 to avoid problems associated in national assessment programs in other countries (including issues we have identified with NAPLAN in this

paper). The program involves schools developing their own approaches to assessing the standards; and is backed by a professional development program with modules on establishing shared expectations; building effective partnerships; knowledge of the learner; meeting the needs of English language learners; knowledge of literacy learning; knowledge of mathematics learning; engaging learners with texts, and engaging learners with mathematics. The national standards website states:

> 'Data and information about learners' progress and achievement is used at all levels of the system to drive improvement. Assessment for learning, teaching as inquiry, school-wide self-review, and system use of data all focus on taking action and then looking at the impact of those actions on learners' progress and achievement.'[4]

There is also a link to their assessment online site[4] that includes a section on teachers' professional judgments about their students' learning and achievement - very resonant with a recurring theme in this paper. We find these features of a national assessment program laudable, particularly if the standards, professional development and assessment approaches incorporate the issues and recommendations we have identified in this paper.

However, we need to insert a note of caution about the New Zealand experience and learn from its problems too. A three-year longitudinal study (Thrupp, 2013) has led to the conclusion that schools developing their own approaches to assessing national standards limits comparability of data between schools – particularly problematic given the public release of data from 2012, resulting in reputational issues that we previously identified in relation to Australia's MySchool website. These issues, along with enrolment issues, have seen a distortion of assessment outcomes at the local school level. Thus, while the intentions of the New Zealand approach are commendable and the constituents of its program quite sound, it is in the publication of results that the issue rests.

Hence we reiterate our earlier recommendation that in Australia, schools are given capacity to challenge inferences drawn from NAPLAN testing and provide their evidentiary data and contextual information to contribute to the interpretation of test data; that schools with low scores are not stigmatised; and that the release of data on a website is duly and most carefully reconsidered from the viewpoints of all involved.

On a broader international scale, PISA (Programme for International Student Assessment) tests are also informative to deliberations on the effectiveness of NAPLAN. Recent PISA analyses have corroborated some issues previously identified in our paper. For example, Bautier and Rayou (2007) have found that, while the purpose of these tests is to assess students' intellectual skills and knowledge, the test requires students to draw on many points of reference in their lives beyond the texts at hand, with a focus on real-life circumstances. Thus results are enhanced or hindered by students' own social context, personal experiences, ethics and values, and cognitive processes – and do not simply reflect students' in-school teaching and learning practices alone.

That said, Finland's success in these PISA tests has been attributed in part to teachers' autonomy that allows them to cater to students' individual contextualized strengths, aspirations and predispositions (Grek, 2009) – corroborating our earlier recommendation that investment be made in nurturing teachers' professional judgments to inform their curriculum and pedagogic choices to achieve NAPLAN's

objectives of better learning outcomes and improved accountability for all students, teachers, families and schools.

g) Other relevant matters

Preparing for global citizenry

Young Australians needs skills that are relevant to solving complex problems effectively both at home and abroad. In a borderless world, we need to prepare our students to engage in the international community on multiple fronts. Towards this end, our educational programs and policies need to be sensitive to these demands on our students' capacity to become independent problem solvers, higher-order thinkers, lifelong learners and ambassadors for Australia. Assessment tools such as NAPLAN, which are implemented early in the life of our young children, could either support or hinder future developments in literacy and numeracy, and need to be framed by a clear and robust vision of what it means to be a global citizen in the 21st Century.

Endnotes

1. Standard 9 (Testing Individuals of Diverse Linguistic Backgrounds) in *Standards for Educational and Psychological Testing,* p. 91, Joint Committee of the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education, 1999, Washington, DC: American Educational Research Association.

2. From http://www.theglobeandmail.com/news/national/education/alberta-to-overhaul-standardized-testing-emphasize-creativity/article11821827/ Retrieved 5th June 2013

3. Rice University (2008, February 16). Negative Implications Of No Child Left Behind: As Graduation Rates Go Down, School Ratings Go Up. *ScienceDaily*. http://www.sciencedaily.com- /releases/2008/02/080214080530.htm Retrieved 6th June 6, 2013

4. Information about New Zealand's National Standards program can be found at http://nzcurriculum.tki.org.nz/National-Standards (retrieved 5th June 2013). An overview of their integrated system of support at http://nzcurriculum.tki.org.nz/System-of-support (retrieved 5th June 2013). There is also a link to their assessment online site at http://assessment.tki.org.nz/ (retrieved 5th June 2013).

References

ACARA 2010, *Australian Curriculum – English*. Canberra: Australian Curriculum, Assessment and Reporting Authority

ACARA 2010, *Australian Curriculum – Mathematics*. Canberra: Australian Curriculum, Assessment and Reporting Authority

Anstey, M. & Bull, G. 2003, *The literacy labyrinth*, 2nd Edn. Sydney: Prentice-Hall.

Au, W. 2009, High-stakes testing and discursive control: The triple bind for non-standard student identities, *Multicultural Perspectives*, *11*, 2: 65-71

Baker, E.L. Barton, P.E. Darling-Hammond, L. Haertel, E. Ladd, H.F. Linn, R.L. Ravitch, D. Rothstein, R. Shavelson, R.J. & Shepard, L.A. 2010, Problems with the Use of Student Test Scores to Evaluate Teachers, *Economic Policy Institute Briefing Paper, no. 278*, Washington

Barksdale-Ladd, M.A., & Thomas, K.F. 2000, What's at stake in high-stakes testing – Teachers and parents speak out. *Journal of Teacher Education, 51*, 384-397

Bautier, E. and Rayou, P. 2007, What PISA really evaluates: Literacy or students' universe of reference? *Journal of Educational Change, 8:* 359-365

Boaler, J. 2002, The development of *disciplinary relationships*: knowledge, practice, and identity in mathematics classrooms, *For The Learning of Mathematics*, *22*, 1, 42-47

Carter, M.G. 2012, Time limitations in NAPLAN numeracy tests, *Australia Mathematics Teacher*, *68*, 1: 36-4

Connor, M.J. 2003, Pupil stress and standard assessment tasks (SATs) An update. *Emotional and behavioural difficulties, 8,* 2: 101-107

Connor, M.J. 2001, Pupil Stress and standard assessment tasks. *Emotional and behavioural difficulties*, *6,* 2: 103-111

Cormack, P. & Comber, B. 2013, High-stakes literacy tests and local effects in a rural school, *Australian Journal of Language and Literacy, 36,* 2: 78-89

Darling-Hammond, L. 2004, Standards, Accountability, and School Reform, *Teachers College Record*, *106*, 6: 1047-108

Darling-Hammond, L. 2010, *Performance Counts: Assessment Systems that Support High-Quality Learning,* Standford Centre for Opportunity Policy in Education, Washington

de Courcy, M. C, & Burston, M. 2000, Learning Mathematics through French in Australia. *Language and Education*. *14*, 2: 75-95

Delpit, L. 2003, Educators as 'seed people growing a new future. *Educational Researcher, 32,* 7: 14-21

Dix, K.L., Slee, P.T., Lawson, M.J. & Keeves, J.P. 2012, Implementation quality of whole-school mental health promotion and students' academic performance, *Child and Adolescent Mental Health, 17,* 1: 45-51

Doecke, B. Kostogriz, A. & Illesca, B. 2010, Seeing 'things' differently: Recognition, ethics, praxis, *English Teaching: Practice and Critique*, *9*, 2: 81-98

Dreher, K. 2012, Tests, Testing times and literacy teaching, *Australian Journal of Language and Literacy, 35,* 3: 334-352

Dudley-Marling, C. 2005, The complex relationship between reading research and classroom practice. *Research in the Teaching of English, 40,* 1: 127-130

Freebody, P. 2007, *Literacy education in school: Research perspectives from the past, for the future.* http://www.acer.edu.au/documents/AER52.pdf Retrieved May 2011

Grek, S. 2009, Governing by numbers: The PISA 'effect' in Europe, *The Journal of Education Policy, 24,* 1: 23-37

Hardy, I. 2013, Testing that counts: Contesting national literacy assessment policy in complex school settings. *Australian Journal of Language and Literacy, 36,* 2: 67-76

Hardy, I. & Boyle, C. 2011, My School? Critiquing the abstraction and quantification of Education, *Asia-Pacific Journal of Teacher Education*, *39*, 3: 211-222

Harris, P. 2010, 'At a school like this': Implementing reading policy in a Kindergarten classroom, *International Research in Early Childhood Education, 1,* 1: 50-59

Harris, P., Turbill, J., Fitzsimmons, P. & McKenzie, B. 2006, *Reading in the primary school years.* 2nd Edition. Melbourne: Thomson.

Harris, P., Turbill, J., Kervin, L. & Harden-Thew, K. 2010, Mapping the archive: an examination of research reported in AJLL 2000-2005. *Australian Journal of Language and Literacy, 33,* 3: 173-196

Hattie, J. 2005, What is the nature of evidence that makes a difference to learning? In *Using data to support learning. Proceedings of the 10th national research conference of the Australian Councilfor Educational Research* (pp. 11-21). Camberwell, VIC: ACER.

Hipwell, P. & Klenowski, V. 2011, A case for addressing the literacy demands of student assessment ', *Australian Journal of Language and Literacy, 34,* 2: 127-146

Hursh, D. 2007, Exacerbating inequality: the failed promise of the No Child Left Behind Act, *Race Ethnicity and Education, 10, 3:* 295–308

Joint Committee of the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education, 1999, *Standards for Educational and Psychological Testing,* 1999, Washington, DC: American Educational Research Association

Jorgensen, R. 2010, Issues of social equity in access and success in mathematics learning for Indigenous students, Teaching Mathematics? Make it Count: *What Research Tells Us About Effective Teaching and Learning of Mathematics, Research Conference*, pp 27-30

Klenowski, V. and Gertz, T. 2009, Culture-fair assessment: Addressing equity issues in the context of Primary Mathematics Teaching and Learning, *Assessment and Student Learning: Collecting, Interpreting and Using Data to Inform Teaching Research Conference.*

Klenowski, V. & Wyatt-Smith, C. 2012, The impact of high stakes testing: The Australian story, *Assessment in Education: Principles, Policy and Practice*, *19*, 1: 65-79

Koch, M.J. and DeLuca, C. 2012, Rethinking validation in complex high-stakes assessment contexts, *Assessment in Education: Principles, Policy and Practice*, *19*, 1: 99-116

Kostogriz, A. & Doecke, B. 2013, The ethical practice of teaching literacy: Accountability or responsibility? *Australian Journal of Language and Literacy, 36,* 2: 90-97

Ladwig, J.G. 2010, What NAPLAN doesn't address (but could, and should), Professional Voice, Australian Education Union, vol. 8, is. 1, pp 35-39

Logan, T., & Greenlees, J. 2008, Standardised assessment in mathematics: The tale of two items. In M. Goos, R. Brown & K. Makar (Eds.), *Navigating currents and charting directions* (Proceedings of the 31st annual conference of the Mathematics Education Research Group of Australasia, Vol. 2, pp. 655-658). Brisbane, QLD: MERGA.

McNeil, L. 2000, *Contradictions of school reform: Educational costs of standardized testing,* New York: Routledge.

Menken, L. 2008, *English learners left behind: Standardized testing as language policy.* Clevedon: Multilingual Matters.

Morley, P. 2011, Victorian Indigenous children's responses to Mathematics NAPLAN items, *Mathematics: Traditions and New Practices (Proceedings of the 34th annual conference of the Mathematics Education Research Group of Australasia*), pp 523-530

Pearson, P.D. 2007, An endangered species act for literacy education. *Journal of Literacy Research, 39*, 2: 145-162

Quinnell, L. & Carter, L. 2011, Cracking the language code: NAPLAN numeracy tests in years 7 and 9. *Literacy Learning: the Middle Years, 19,* 1: 49-53

Sahlberg, P. 2011, *Finnish lessons: What can the world learn from educational change in Finland?* New York: Teachers College Press

Saubern, R. 2010, Useful assessment is timely assessment, *Teacher*, 24-27

Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I., Taggart, B. & Elliot, K. 2008, *Effective Provision of Preschool and Primary Education: final report EPPE*. London: DCSF.

TESOL, Inc., 2005, Position Paper on Assessment and Accountability of English Language Learners Under the No Child Left Behind Act of 2001 (Public Law 107-110), Alexandria, Virginia, USA

Thrupp, M. 2013, National Standards for student achievement: Is New Zealand's idiosyncratic approach any better? *Australian Journal of Language and Literacy, 36,* 2: 99-109

Warren, E., & Miller, J. 2013, Young Australian indigenous students' effective engagement in mathematics: The role of language, patterns and structure. *Mathematics Education Research Journal, 25*, 151-171

Wigglesworth, G. Simpson, J. & Loakes, D. 2011, NAPLAN language assessments for Indigenous children in remote communities: issues and problems, *Australian Review of Applied Linguistics, 34,* 3: 320-343.

Wilson, L. 2012, The NAPLAN view of reading, teachers and reader-learners. www.literacyeducators.com.au/naplan Retrieved 5th June 2013

Wilson, L. & Hornsby, D. 2012, Teaching to the test. www.literacyeducators.com.au/naplan Retrieved 5th June 2013

Winter, P. 2010, *Engaging families in the early childhood development story – neuroscience and early childhood development.* Canberra: MCEECDYA. Available at http://www.mceecdya.edu.au/mceecdya/publications,11582.html#Early%20childhood%20education%20and%20care Retrieved 8th November 2010.

Wu, M. 2010, The inappropriate use of NAPLAN data, *Professional Voice*, *Australian Education Union, 8*, 1: 21-25

Wu, M. & Hornsby, D. 2012, Inappropriate uses of NAPLAN. . www.literacyeducators.com.au/naplan Retrieved 5th June 2013