**Submission to the Senate on the Administration and Reporting of the National Assessment Program in Literacy and Numeracy (NAPLAN)**

I'm writing this submission as both a private individual and representing my organisation. I'm a specialist in assessment and reporting. I've worked in this field for more than 30 years. I'm currently the director of the Assessment Research Centre at the University of Melbourne where I occupy the Chair of Education (Assessment). I've occupied this position for 14 years after holding the equivalent chair at RMIT University for seven years. In my career I have been employed as a university teacher and researcher and as an employee of the State government education system in Victoria. In the latter role I was responsible for leading a team that developed the original literacy and numeracy profiles. The profiles became the prototype for the National Profiles and Curriculum Statements which were distributed in the early 1990s. They were the forerunner of wide spread use of developmental learning progressions which have remained in use over the past 20 years to illustrate that learning progressions of students over their time in school, and which currently inform the adoption of the developmental scale for reporting NAPLAN data.

Reporting procedures based on standards frameworks were developed in the mid-1980s by a team of researchers working in the Curriculum Branch of the Victorian Education Department. Such reporting methods were developed originally to assist schools in reporting student development in terms of competence rather than scores. The procedure was based on the idea that scores were codes and that these need to be decoded. The decoding process can help teachers use descriptions of students' competence in ways that scores cannot.

This notion of reporting what students have achieved rather than numerical scores has remained in use for more than 20 years. The NAPLAN continues to use this method of reporting in terms of bands of increasing sophistication or competence in each of the skills tested. The utility of this method of reporting focuses on informing teachers and students about learning outcomes. However, there is still a need for scores as a means of communicating aggregate data to systems or to schools as a whole. Hence a dual mode of reporting on large-scale assessment continues to be needed. The NAPLAN fulfils this need. Teachers and students (and parents) can be informed about learning outcomes and achievements of students in terms of their competence. Principals, regions and systems of education are informed about aggregate performances based on numerical scores. The score system used in NAPLAN conforms to international standards of reporting using a scale with a mean of 500 and a standard deviation of 100. This is an internationally acceptable method of reporting test scores.

The two forms of reporting enable both teachers and systems to make constructive plans for improvement of student learning outcomes and, as such, they serve a useful purpose in improving education in Australia.

The system is not perfect. The data is limited. The tests are not sufficiently long to produce data of sufficiently high reliability to enable individual intervention or clinical style decisions to be made. This is not a criticism of the test. In psychology, clinical decisions about individual people are made on the basis of test results. Strict standards are required of those psychological tests. Reliability must be above 0.9 and is optimally around 0.95. There are no such standards for educational testing. Reliabilities of the NAPLAN tests are not published. The range of NAPLAN is comprehensive and has the potential to influence the curriculum. In the field of Mathematics and English, the tests will no

doubt influence teaching. There is already anecdotal evidence of systems of education placing great pressure on teachers to improve scores and, in some systems, targets of percentage improvement have been set. Systems of education are under such pressure to improve results (because of funding arrangements) that this pressure is passed on to teachers and has the danger of narrowing the curriculum focusing only on literacy and numeracy. This is particularly a danger in the time immediately before the administration of the NAPLAN tests.

On the other hand, the effect of NAPLAN can be profound in terms of improving student achievement and providing a focus on learning outcomes in these basic skills. The information given to teachers provides them with sufficient stimulus to focus on improving the achievements of all students. In 2010 the first opportunity will arise to check individual growth patterns. Teachers will be able to see what development has taken place from grade 3 to 5, from grade 5 to 7, or from grade 7 to 9.

It is likely that there will be some anomalies in the data when individual student performances are examined. Reporting at an individual level often exaggerates the errors in the data. The message for teachers should be that individual scores need to be supplemented with other evidence of learning and teachers need to be supported in their selection and interpretation of multiple sources of evidence of student learning – one measure of which is provided by NAPLAN. In general it can be anticipated that most of the individual student results will appear logical and consistent with the teachers' knowledge, but any anomalies in the data may receive more attention than they warrant.

There have been critics of the system who have demonstrated that the reliability or measurement error information of the data can be used to undermine the confidence of systems, teachers and parents in making comparisons either over time or across jurisdictions. It is possible to examine the data from the point of view of the standard errors of measurement and conclude that the data is not sufficiently accurate to draw valid conclusions. Such a view assumes measurement error is random and that this random error undermines the capacity to draw conclusions. However, the persistence of differences across time for thousands of schools would indicate this perception about measurement error is flawed. Basic statistical tests of the probability of thousands of results being in a similar direction (improvement or different) would indicate that the probability of this occurring by chance as it is described by the critics is negligible. That is to say that it is almost impossible to conclude that the data cannot be used when consistent results are produced over thousands of cases. This consistency of the results supports the use of the data for comparison purposes.

The critics' focus on small differences over time and across schools, regions or jurisdictions can also be a flawed approach. The use of statistical significance or hypothesis testing is a curious approach to the use of the data from the national testing program. Statistical significance is based upon the use of random samples drawn from a population. These are not random samples. It is the population. The differences in the data are real differences in the performances of the students, and the schools, and systems. They are not estimates of differences based upon sample analyses. In this context, the use of significance tests and standard errors do not provide meaningful or useful information, and they should be removed from reports. It is better to focus on the change in competence levels and whether the differences are educationally important. For this to be done the small differences in scores need to be translated, or decoded, into differences in competencies being demonstrated by the students. This would enable the relative importance of differences to be evaluated in educational terms.

The national reporting website (My School) presents an interesting dilemma. Its first attempt to

report the student and school level data was criticised because of the limited nature of the data. Yet the millions of hits that the website received when it was first made available indicated an intense interest in the data on the part of the community. There is also a clear interest in examining the data at a school level. The limit of what is available to be reported has been the source of some criticism. Ranking schools, even "like" schools, might be considered to be of limited use. However, it does tend to put pressure on low performing schools and for the students and those schools this may be to their benefit. Note that this is possibly beneficial if, and only if, the resources and opportunities to improve the performances are available. However there is so much information that is not available that could be used in explaining differences in performance, that we are at risk of misinterpreting the reasons for these differences.

Perhaps the most informative reporting system that I have seen was used in Maryland in the United States. Prior to the introduction of the national programme "No Child Left Behind", the Maryland Assessment Programme reported extensively on the schools and districts within the state. In a manner similar to NAPLAN, students were assessed to have reached a particular level of competence. For each school the percentage of students at each level of competence was reported. Schools were classified according to the target proportions of students at each level of competence. There was no emphasis on a single minimal competence benchmark but there was an emphasis on the proportion of students reaching levels of excellence. The results were reported for teacher developed tasks with moderated assessments across schools together with the results of standardised testing programs. These results were reported in the context of a large range of indicators of the school. They included measures of average family income, expenditure per pupil, teacher student ratios, teacher aides per pupil, library and other facilities as well as a range of support material and a range of other indicators of the quality of school life. These data were not used in any contextual analysis that adjusted scores for opportunity and risk factors nor were they reports in groups of 'like schools' Because the data was available and explicit, the public could see for themselves the context in which the school operated and the effects on student achievement.

It was clear that there was great disparity between schools in terms of the support provided and in terms of the achievement levels. While low performances were not excused on the basis of reduced opportunity, the state and school districts had multiple targets to address in improving the context of student achievement. Reporting a limited range of indicators of the school may obscure other points of intervention that systems and jurisdictions could address in removing inequity of opportunity for students.

In the United Kingdom the school reports provide what is called a context or value-added analysis. In these reports and analyses the school results are modified by controlling for factors that the schools and teachers cannot control and including an explanatory factor for the factors that teachers can control. This is a very complex analysis and the public is presented with an actual and an adjusted score. However, it is not transparent to the general public who take 'on faith' the statistical arguments for the adjustments and the ensuing ranks of schools.

If the purpose of the national program is one of accountability, then the accountability metric may need to take into account the factors that teachers can and cannot control. Later iterations of the My School website intend to take these factors into account and may well provide a greater transparency and contextual explanation. However for the time being and under the time constraints imposed by the systems, the My School website has served an obviously valuable role amongst the community.

International research currently under way may address some of the issues of NAPLAN. In a study being conducted in six countries including Australia and supported by the Department of Education, Employment and Workforce Relations, technology based assessment strategies are being developed that will enable summative data to be collected and harvested in student learning outcomes and simultaneously provide immediate feedback to teachers for instructional intervention. These procedures are of interest for their potential importance to NAPLAN and the capacity to reduce the turnaround time for the reports to be issued to schools. The goal of the project (Assessment and Teaching of 21st Century Skills – www.atc21s.org ) is to enable large scale assessment strategies to be technology based and to provide relatively instant feedback to teachers and students. The current digital education revolution can enhance the chances of this being successfully implemented in Australian schools. The project outcome has the potential to significantly alter the way large scale assessments such as NAPLAN are conducted, the immediacy of data availability, and the breadth of measurement of skills that have relevance for the success of our youth through this century in a technologically-enabled society.