# Executive summary

Recent evaluations of the Longitudinal Surveys of Australian Youth (LSAY) have recommended investigating the potential for combining LSAY data with external data sources as a way to improve the breadth of information in the survey, but without adding respondent burden (Gemici & Nguyen 2013). Linking administrative data from the education, training and health sectors to LSAY data would greatly enhance the ability to explore the key drivers of young people's transition outcomes.

The aim of this project is to assess the feasibility (and practicability) of linking National Assessment Program — Literacy and Numeracy (NAPLAN) scores to LSAY data (which contain data from the Organisation for Economic Co-operation and Development's (OECD) Programme for International Student Assessment (PISA)). A second aim is to determine the similarity between NAPLAN and PISA in measuring underlying academic achievement and whether the two measures rank individuals similarly across the distributions of NAPLAN and PISA.

## LSAY, PISA and NAPLAN

The Longitudinal Surveys of Australian Youth (LSAY) tracks young people as they move from school into further study, work and other destinations using large nationally representative samples of 15-year-olds. Surveys are conducted annually over a ten year period to capture information about young people's transitions from school to tertiary education and the labour market. Since 2003 the initial survey wave has been integrated with PISA.

The Programme for International Student Assessment (PISA) is a triennial international survey that aims to evaluate education systems worldwide by testing the skills and knowledge of 15-year-old students.

National Assessment Program — Literacy and Numeracy (NAPLAN) is the annual assessment of literacy and numeracy performance undertaken by all students in Years 3, 5, 7 and 9. The data from the NAPLAN tests provide schools with information to measure their students' achievements against the national minimum standards.

Many researchers use literacy and numeracy scores from PISA as key predictors of post-school transition outcomes as these scores are available as part of the LSAY dataset. Given that both PISA and NAPLAN scores are routinely used in research studies that inform national education and training policy, it is important to verify that the two measures have a reasonable degree of overlap.

## Methodology

The LSAY data are owned by the Australian Government Department of Education and Training and specific arrangements have been established by the Commonwealth Government to manage the risks associated with integrating Commonwealth data. As the custodians of the NAPLAN data, each of the jurisdictions were also required to provide approvals for linking their state or territory's NAPLAN scores to the LSAY data.

In order to link LSAY records to their NAPLAN scores it was necessary to obtain consent from individual LSAY respondents. Three methods for obtaining consent were used — written, oral (via telephone) and online.

The data were analysed through comparisons of summary statistics, graphs and regressions between PISA and NAPLAN to determine the relationship between the two measures.

## Findings

The project demonstrated that it is technically feasible to link NAPLAN scores to LSAY records. About four out of five LSAY respondents who had the opportunity to respond to the consent question via their telephone or online interview agreed to have their data linked. We found that obtaining consent using written methods was far less effective, with only one in ten respondents providing consent in this way. Of those providing consent, a matching rate of 98% was achieved overall.

The analysis undertaken in this paper was restricted to a small sub-group of LSAY participants from the 2009 commencing cohort (Y09). The sub-sample comprised those who participated in the 2014 wave of LSAY and provided consent to link to NAPLAN. The analysis showed that this group of participants had higher NAPLAN and PISA scores than the average of all respondents (national average for NAPLAN). The likely reason for this is that higher-performing and more successful individuals are more likely to remain in the LSAY survey over time and may be more likely to provide the required consent to match their NAPLAN and LSAY data.

The secondary purpose of the linkage project was to investigate how similar the PISA and NAPLAN measures are. The statistical analysis showed there is a reasonable level of agreement between the two measures. The weighted correlations were in the range of 0.7 for both maths and reading. The correlations between the NAPLAN reading scores and the PISA reading scores were slightly higher than those for maths.

## The future

Despite the high rate achieved when linking the data, it is important to consider how rates of consent can be improved and to develop other strategies to maximise the pool of LSAY respondents available for data linkage. To this end, the following strategies are suggested:

- Avoid where possible the use of written methods in obtaining consent. Telephone and online methods achieve higher rates of consent.
- Obtain consent early to maximise the number of records available for linking, which also helps to remove bias. This could be done by gaining consent during the PISA assessment, or seeking consent during the first round of LSAY interviews.
- Simplify the questions used and information provided during the consent-gathering stage to reduce the burden for interviewers and respondents while ensuring respondents are fully informed.
- Consider obtaining approvals through the existing national governance processes established to support the work of the Commonwealth Government's Education Council rather than separately for each state and territory.

The success in matching NAPLAN scores to the LSAY data means that we can now consider joining multiple years of NAPLAN results. This would allow for the creation of an expanded linked dataset which could be made accessible to researchers and would enable analyses of important policy issues related to the effects of early education outcomes on young people's transition from school to work. Further developments might also include consideration of linkages with other datasets, such as the ABS Census of Population and Housing data (to obtain data on the areas in which respondents live, attend school or undertake further post-school study), and Medicare data.