Productivity Commission Inquiry into the National Education Evidence Base

Submission by Associate Professor Ben Edwards and Mark Sipthorp at the Australian Institute of Family Studies

This submission to the Productivity Commission's inquiry into the National Evidence base addresses the following issues:

1. New education outcomes that are relevant: the importance and relevance of social and emotional skills and the home learning environment.
2. Measuring the home learning environment and social and emotional skills.
3. The need for a new LSAC cohort
4. The promise and the barriers in the use of administrative data.
5. Challenges and impediments to data linkage in education
6. Support for administrative data linkage for research purposes: Evidence from LSAC.

The relevance and importance of social or emotional skills (non-cognitive skills) and the home learning environment

While there are a large number of factors that could influence children and youth educational outcomes, we argue that there is evidence to support consideration of a national collection of information of social or emotional skills or personality (non-cognitive skills) and the home learning environment.

Increasingly non-cognitive skills are being recognized as a critical component in education. For example, there is some evidence to suggest that non-cognitive skills predict academic achievement and grades (Borghans et a., 2011a), in part because elements of non-cognitive skills foster cognitive development but also because they independently predict achievement and grades in their own right (e.g. Kautz, Waal, Heckman, Borghans & Diris, 2014). There is also emerging evidence that non-cognitive skills are important contributors to the employability of young adults and independently predict life long earnings (Borghans et a., 2011a). Another important rationale behind the collection of non-cognitive skills is that there is some emerging evidence that these skills are more malleable during adolescence than cognitive skills such as IQ, which tends to be more stable during this period (Kautz et al., 2014).

Non-cognitive skills are also increasingly being recognized as an important means by which the long-term benefits of early childhood education are transmitted. Although many evaluations of Early Childhood Education and Care (ECEC) show that there are short-term benefits to ECEC, these tend to fade in the primary school years (Leak et al., 2013). Re-analyses of influential ECEC programs such as the Perry Preschool Program and the Abecedarian Project have shown that it is non-cognitive skills that are largely responsible for the long term benefits observed in these programs (Chetty et al., 2011; Elango et al., 2015; Heckman, Pinto, & Savelyev, 2013; Love, Chazan-Cohen, Raikes, & Brooks-Gunn, 2013).

Families play an important role in preparing children to be ready for school, but they also support learning throughout the school years. An important component of how families support children's learning is the home learning environment. Several studies suggest that the home learning environment in early childhood has been found to account for much of the early achievement gaps that lead to longer term gaps in later years (Melhuish et al. 2008; Yeung & Pfeiffer, 2009). Exposure to stimulating activities within the home environment, such as having books in the house, reading and counting are all associated with higher levels of academic achievement (Bradley, McKelvey, & Whiteside-Mansell, 2011; Melhuish et al., 2008). For example, recent research by Yu and Daraganova (2015) using data from the Longitudinal Study of Australian Children (LSAC) found that daily reading to children at 2-3 years, when compared to 0-5 days per week, was associated with higher levels of Year 3 Reading and Maths NAPLAN scores (equivalent to 20 weeks of schooling in Year 3) even after adjusting for socio-demographic factors[1]. Having a national picture of the variation in families capacities to support their children's learning, particularly in the early years, enables formulation of policies and programs to support home learning in families that may not have capacity to support their child in preparing for school[2].

Measuring the home learning environment and non-cognitive skills

The measurement technologies capturing the home learning environment are well established with question banks that capture the relevant dimensions. Numerous large-scale surveys, such as LSAC, have collected this information over the preschool and primary school years (Edwards, 2014).

Different elements of non-cognitive skills have been measured in high profile national and international research including locus of control, grit, self-esteem and measures of personality as well. While there is no consensus at this stage of the optimal elements of social and emotional skills that need to be assessed, work the OECD has been undertaking with input from AIFS suggests that the Big Five personality measures of openness, conscientiousness, extraversion, agreeableness and neuroticism may form key elements of an international framework of assessment in the future. In fact, the OECD's planned longitudinal study of social and emotional skills will use this framework to organise planned assessments (OECD, 2015). An online national assessment could provide valuable information to education systems to better understand the factors that promote social and

[1] The socio-demographic factors included child's gender, family type, mother's language spoken at home, family's socio-economic position, region of residence, and neighbourhood disadvantage status.

[2] The Home Interaction Program for Parents and Youngster is one intervention that supports parents in reading to their children (Liddell, Barnett, Roost & McEachren, 2011). Lower cost initiatives that "nudge" parents into increased reading have also been trialed overseas (Mayer, Kalil, Oreopoulos, & Gallegos, 2016).

emotional skills in ECEC, primary and secondary schools so that lasting gains of educational investments have a better chance of being sustained across the life course. The move to online testing of NAPLAN opens the opportunity to trial a concise assessment of non-cognitive skills with the potential for national coverage.

The need for a new birth cohort for *Growing Up in Australia*: The Longitudinal Study of Australian Children (LSAC)

The Birth (B) cohort for LSAC are now 12-13 years of age and it has been over a decade since Australia has had a new, nationally representative birth cohort of children. While there are still many areas of public policy that the Birth (born 2004) and Kindergarten (born 2000) cohorts will contribute to, particularly in the areas of the primary, secondary and tertiary education where LSAC now provides very detailed and rich data to support development, there is an emerging gap in the evidence base in the area of early childhood education. For instance, the National Quality Framework (NQF), which applies to most long day care, family day care, preschool and kindergarten, and outside schools hours care services in Australia started on 1 January 2012. The key changes under this new framework are improved staff-to-child ratios, new staff qualification requirements, a new quality rating system to ensure Australian families have access to transparent information relating to the quality of early childhood education and care services, and the establishment of a new National Body to ensure early childhood education and care is of a high quality (COAG, 2009). This is just one area of policy where a new birth cohort would play an important role in the monitoring and evaluation of the potential impacts of quality improvements. Other possible areas of policy relevance to the education sector include understanding variations in state differences in school starting age policies (Edwards, Taylor & Fiorini, 2011) and the implications of paid parental leave for changes to childcare arrangements.

Potential features of a new cohort: Building on LSAC's strengths but advancing into the "big data" age

One of the great achievements in the LSAC study has been data linkage. With consent of participants, LSAC has led the way in terms of linkage of administrative data and making such data available to the research community (Edwards, 2014). At the time of this submission the following data sets had been linked:

- the Medicare Benefit Scheme (MBS);
- the Pharmaceutical Benefit Scheme (PBS);
- the Australian Childhood Immunisation Register (ACIR);
- National Assessment Program –Literacy and Numeracy (NAPLAN);
- the Australian Early Developent Census (AEDC);
- MySchool;
- National Childcare Accreditation Council quality data;
- and area level data from the Census of Population and Housing.

As a result, the current LSAC cohorts are a hybrid of administrative data along with in-depth observational assessments and surveys. Future cohorts are likely to benefit from investing further in data linkage at the outset of the survey. Direct assessments and surveys will be a critical component of any cohort because there are many developmental outcomes that are not able to be captured via administrative data, however, more limited banks of survey questions around demographic characteristics can be achieved. Additionally, a larger sample size and over sampling of disadvantaged groups are likely to be other key features that would be needed in any future cohort of LSAC.

The promise and barriers to the use of administrative data

There are tremendous potential benefits for Australia in Federal and state governments investing in linked national data sets. The national or state coverage of many data assets make them appealing to answer many policy questions and the low cost of data linkage means that the potential returns could be significant. In many contemporary evaluations of education policies, having baseline data that enable pre and post implementation assessments would provide much more information than is routinely collected in many evaluations. For example, an AIFS review of Commonwealth funded evaluations of place based initiatives for Prime Minister and Cabinet found that very few evaluations collected adequate baseline information to enable any sort of analysis of potential causal impacts (Wilks, Lahausse & Edwards, 2015).

There are however many barriers to adequately harnessing the power of administrative data assets. To be adequate for research purposes there needs to be detailed information about the nature of the data included in data sets and about how the data is inputted into databases. Rules and interpretation of definitions about the inclusion of data items mean that there is the potential for variation in the nature of specific data items between different areas within the one state. Policy changes or changes in the interpretation of inputting rules over time also may limit the temporal comparability of data.

The precise benefits of large-scale administrative data linkage are largely unknown. It is likely they will hinge of the quality of the data assets and realistic expectations from policy makers about what conclusions can and cannot be drawn from administrative data assets in the short and longer terms. Vanguard data sets that have wide public release amongst the education research community will be important in developing research capacity in using such data resources, and more fully understanding the strengths and limitations of particular data resources. These understandings can then be used to refine or develop new administrative data assets. An ongoing learning approach, such as the one outlined above, is likely to be

required to fully yield the benefits of the administrative data and to also build the capacity to analyse such assets.

In the context of understanding the strengths and limitations of educational data assets, such as an Australian Longitudinal Learning Database, data assets that include administrative data coupled with other sources of information become incredibly valuable. LSAC and other longitudinal surveys that combine both sets of information, will enable testing of the applicability of administrative data to inform educational policy, as well as highlight what other information will be required. The likely outcome of such work will be the conclusion that hybrid models of administrative information coupled with surveys and direct assessments will provide the most powerful evidence for decision-making that has significant and long-term policy importance. A nuanced understanding about when administrative data is most useful and when other information is required, will be a key learning from such activity.

Challenges and impediments to data linkage in the education sector

With respect to data linkage and integration of education data, AIFS has completed linkage to National Childhood Accreditation Council (NCAC), National Assessment Program Language and Numeracy (NAPLAN), Australian Early Development Census (AEDC) and My School data.

In our experience some of the challenges and impediments in data linkage are:

- Governance, policy and procedures regarding the release of data.
  There is no nationally consistent approach to how data linkages can occur. Most custodians have differing governance arrangements and release protocols and a significant amount of time is spent `liaising with these custodians to access their data, agree to have it linked with other sources and provide it to researchers.

- The number of data custodians that need to be approached to complete data linkages.
  When linking State and Territory data, each jurisdiction needs to be contacted and agree to provide their data. Issues about governance as specified above need to be address. For example the jurisdictions are the custodians of the NAPLAN data. This data is held centrally, but can't be provisioned without their consent.

There are potentially a number of benefits of expanding the Unique Student Identifier nationally to students in school and early childhood education and care.

Expanding the Unique Student Identifier would be beneficial for the following reasons:

- Linkage would be easier, as long as the USI's could be utilised.
- Linking data together to get a complete picture of an individual's education history would be possible.
- Mobility between jurisdictions becomes less of an issue.

Support for administrative data linkage for research purposes: Evidence from LSAC.

Much of the data linkage work the Institute has conducted over the years has involved asking consent of participants. In particular, this has been the approach adopted in the data linkage work the Institute has undertaken for LSAC. A by-product of this activity is to assess the consent rate for data linkage in a nationally representative group of parents in the Australian population. Table 1 shows that consent rates are above 90% regardless of the type of administrative data linked (health or education data). This provides suggestive evidence that the general population, or at least the general population of parents of children, support the linkage of administrative data for research purposes.

Table 1: Consent rates to link administrative data to the Longitudinal Study of Australian Children

Linked data	Who provided consent	Consent rate	Wave consent given
MBS	Parent 1	97%	Wave 1
PBS	Parent 1	97%	Wave 1
ACIR	Parent 1	97%	Wave 1
MBS	Study Child	93.7%	Wave 6
PBS	Study Child	92.3%	Wave 6
AEDC	Parent 1	95.5% (B only)	Wave 4
NAPLAN	Parent 1	95.4% (K) 95.5% (B)	Wave 3 & 4 Wave 4

Notes: MBS –Medicare Benefit Scheme; PBS –Pharmaceutical Benefit Scheme; ACIR - Australian Childhood Immunisation Register; AEDC –the Australian Early Development Census; NAPLAN -

References

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., & Yagan, D. (2011). How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star. The Quarterly Journal of Economics, 126(4), 1593-1660.

Edwards, B., Taylor, M., & Fiorini, M. (2011). Who gets the 'gift of time' in Australia? Exploring delayed primary school entry. *Australian Review of Public Affairs, 10(1),* 41-60.

Elango, S., García, J. L., Heckman, J. J., & Hojman, A. (2015). Early Childhood Education: National Bureau of Economic Research.

Leak, J., Duncan, G. J., Li, W., Magnuson, K., Schindler, H., & Yoshikawa, H. (2013). Is Timing Everything? How Early Childhood Education Program Impacts Vary by Starting Age, Program Duration and Time Since the End of the Program, Working Paper: Center on the Developing Child, Harvard University.

Liddell, M., Barnett, T., Roost, F. G., & McEachren, J. (2011). Investing in our future: An evaluation of the national roll out of Home Interaction Program for Parents and Youngsters (HIPPY). Final report to the Department of Education, Employment and Workplace Relations **,** Brotherhood of St Laurence.

Love, J. M., Chazan-Cohen, R., Raikes, H., & Brooks-Gunn, J. (2013). What makes a difference: Early Head Start evaluation findings in a developmental context. Monographs of the Society for Research in Child Development, 78(1), vii-viii, 1-173.

Mayer, S. E., Kalil, A., Oreopoulos, P., & Gallegos, S. (2016). Using behavioral insights to increase parental engagement: The Parents and Children Together (PACT) Intervention. National Bureau of Economic Research Working Paper 21602.

Melhuish, E. C., Phan, M. B., Sylva, K., Sammons, P., Siraj-Blatchford, I., & Taggart, B. (2008). Effects of the Home Learning Environment and Preschool Center Experience upon Literacy and Numeracy Development in Early Primary School. Journal of Social Issues, 64(1), 95-114.

OECD (2015). *Skills for Social Progress: The Power of Social and Emotional Skills*, OECD Skills Studies, OECD Publishing. http://dx.doi.org/10.1787/9789264226159-en

Wilks, S., Lahausse, J., & Edwards, B. (2015). *Commonwealth Place-Based Service Delivery Initiatives: Key Learnings project. Final report* (Research Report No. 32). Melbourne: Australian Institute of Family Studies.

Yeung, W.-J. J., & Pfeiffer, K. M. (2009). The black-white test score gap and early home environment. Social science research, 38(2), 412-437.