



Barbara Preston

25 June 2010

Submission to the Parliament of Australia Senate Inquiry into the
Administration and Reporting of NAPLAN Testing:
Does the Index of Community Socio-Educational Advantage have systematic bias?

1. Introduction

This submission is concerned with a central aspect of the presentation of NAPLAN results on the My School website: the determination of 'statistically similar' schools. It is thus most relevant to the Term of Reference (c) (iv).

My argument is that the major component of the Index of Community Socio-Educational Advantage (ICSEA) - that which is based on Australian Bureau of Statistics (ABS) 2006 Census Collection Districts (CDs) - is, most certainly, systematically and substantially biased.

Over the coming weeks I plan to carry out a detailed analysis that would provide evidence in relation to this argument. To do this I need to have access to an appropriate index. Some weeks ago I wrote to the Australian Curriculum, Assessment and Reporting Authority (ACARA) seeking access to the CD-based index component of ICSEA. I have not yet received a reply from ACARA. If I am not given access to the CD-based index component of ICSEA, I plan to use appropriate Australian Bureau of Statistics SEIFA indexes (ABS 2008). The paper arising from this work will be completed by 23 July, and further work may be carried out over subsequent months.

I have attached a brief paper that I prepared in March 2010 that covers issues with which this submission is concerned, but utilises 2001 ABS Census data.

2. The determination of 'statistically similar' schools on the My School website

This section explains the way 'statistically similar' schools are determined, and those familiar with this process may go directly to section 3.

A central element of the My School website is the comparison between 'statistically similar' schools regarding their NAPLAN results. On each school's page on the website are charts for the 2008 and 2009 years that set out the school's average NAPLAN scores for each domain (reading, writing, spelling, grammar and punctuation, and numeracy) at each relevant year level (3, 5, 7 and/or 9). Immediately below are the average scores for 'statistically similar schools' (SIM) and for all Australian schools (ALL). There is a bar above each of the SIM and ALL scores that indicates by its colour whether the particular school is 'substantially above' (green), 'above' (pale green), 'close to' (white), 'below' (pink), and 'substantially below' (red) 'statistically similar' schools and all schools respectively. In addition, a page is linked to each school's main page that lists up to 60 'statistically similar' schools and provides similar data and colour coding. Thus, a scan of the dominant colours can quickly indicate to the viewer whether or not the school with which they are concerned is substantially above or substantially below 'statistically similar' schools, or somewhere between.

Schools are classified as 'statistically similar' if they have a similar ICSEA score. ICSEA has several components, the most important of which is an area-based index, with 14 variables drawn from ABS 2006 Census data at the CD level (a CD has around 220 dwellings). This index is broadly similar to ABS Socio-Economic Indexes for Areas (SEIFA), but its particular components are selected from the 35 variables that are used in the SEIFA indexes according to their high correlation with school level NAPLAN performance. The index is applied to CDs, and the value for a school is derived from the weighted average of all the CDs of students' home addresses. The final ICSEA value for a school is adjusted for remoteness and the proportion of Indigenous students, and a small proportion of schools have additional adjustments if considered warranted (ACARA 2010).

3. Area-based indexes and the ecological fallacy

Area-based SES¹ indexes can be of great value. They are very administratively convenient if based on CD-level ABS Census data (or similar) and geocoding (of, for example, the home address of school students or hospital patients, or the location of facilities). However, such area-based indexes are vulnerable to the ecological fallacy.

The 'ecological fallacy' involves inferring individual-level relationships from relationships observed at the aggregate level². (Macintyre & Ellaway 2000, p. 332)

The ecological fallacy can be involved in the determination of the SES score of a school (as a whole) because the mechanism involves the aggregation of the weighted average of the area-based scores of the areas (CDs) in which *individual* students live. This can be presented as 'aggregate level (CD) ⇒ individual level (student) ⇒ aggregate level (school)'.
⇒

¹ In this submission the term 'SES' is used generically for socio-economic or socio-educational advantage/disadvantage.

² The 'aggregate level' may involve any large group, and does not necessarily have to be area-based. However, in this submission an area-basis is assumed.

The degree to which the fallacy occurs and is a problem depends on the actual empirical situation and the policy purposes (and unintended consequences) of the analysis.

For example, an ICSEA-like index applied to a schooling system in which all schools are comprehensive and every school is zoned to exclusive areas more or less contiguous with CD boundaries is likely to involve very little, if any, of the ecological fallacy. If the schooling system is more complex and there is some involvement of the ecological fallacy, this may not be a serious problem if any likely bias is taken into account and/or the policy outcomes (intended and unintended) of the analysis are not 'high stakes' and potentially damaging. This occurs with some school authorities' own internal analyses of test and other school-level results³.

The alternatives to area-based measures are individual level measures (see Baker & Adhikari 2007). Indexes can be developed using direct measures such as individual students' parental occupation (or any other relevant attribute for which data of adequate quality can be collected with appropriate ease and maintained with appropriate confidentiality). There can be an 'atomistic' or 'individualistic' fallacy that is parallel to the ecological fallacy if group-level relationships are inappropriately inferred from individual-level data (Macintyre & Ellaway 2000, p. 334). However, though such a fallacy may be involved, it is less likely than if an area-based index is used because there would be one less element in the mechanism: 'individual level (student) ⇒ aggregate level (school)'. In addition, systematic bias does not appear to be likely in the context of Australian schools. There are difficulties, however. The most appropriate index (or scale) may be the Australian Socioeconomic Index 2006 (AUSEI06) (McMillan, Jones & Beavis 2009) (which, for estimates of school-level SES, is based only on parental occupation) because the data collection involves a simple question for students (or parent/caregiver). However, as it is occupation-based there may be anomalies in some regions or for some school communities where dominant occupations are anomalous in the Index.

4. Investigating any systematic bias in the application of ICSEA

Systematic bias of ICSEA can only be investigated indirectly and in a general, indicative way, and conclusions cannot be directly applied to specific schools. The data available for such an investigation can only cover certain classifications of schools, and not others that may be relevant.

It is possible to score or rank CDs according to relevant indexes (the CD index component of ICSEA, or any of the SEIFA indexes, for example). All primary or secondary school students within each CD can be analysed according to the type of school attended according to the Census classification 'TYPP': government, Catholic, or other nongovernment. This classification does not, of course, indicate the actual school attended, or whether the school attended is comprehensive or selective, or high or low fee (if nongovernment). For every student, in addition to the SES scores of the CDs in which they live and the types of schools attended, data can be obtained from the Census on family income (FINF) and/or

³ Similarly, the former Disadvantaged Schools Program (DSP) utilised area-based indexes of disadvantage to determine the clusters of schools (government and nongovernment) that can access the program. Apparent anomalies in the outcomes of the application of the indexes were resolved by local committees. The DSP was conceptualised as an 'area-based' program (not just a program for discrete individual schools, and certainly not for individual disadvantaged students), and thus the *ecology* of the area was important (Interim Committee for the Australian Schools Commission 1973, pp. 91-106).

other characteristics such as family structure (for example, whether one parent family) or dwelling attributes (for example, whether there is an internet connection).

The attached short paper reports on analysis using 2001 Census data.

This indicates that in the most disadvantaged CDs there are usually students in high income families, and in the most advantaged CDs there are students in low income families. Whether from disadvantaged or advantaged CDs, students in high income families tend to be more likely to attend nongovernment schools (especially independent schools), and students in low income families tend to attend government schools. In some cases this pattern is so marked that if an independent school drew only from very disadvantaged CDs, the majority of students would still be in high income families. But in some localities the pattern may be the reverse, and, of course, there is great diversity within the sectors which Census data cannot illuminate.

I plan to carry out similar, but much more detailed analysis using 2006 Census data. I will be very happy to report to the Committee on the outcomes of this analysis.

5. Other matters

There are a very large number of other matters related to the administration and reporting of NAPLAN testing with which I am concerned. I understand that these matters are well covered in other submissions to this Inquiry, and will not elaborate here. Some include:

- The detrimental consequences for curriculum and pedagogy of high stakes reporting of NAPLAN results.
- The individualistic connotations and real consequences of the name and content of the My School website, thus undermining or underplaying the importance of co-operative and collegial relations between and within schools and their communities.
- The inappropriateness of encouraging parental judgements about or choice of school based primarily on aggregate test results, especially if those results are presented in a way that does not indicate any 'value-added', does not take account of differential resources of schools, is systematically biased, and/or has a high level of unreliability, especially for small schools. This (implicit) encouragement of judgement and choice based on simple quantitative data militates against (a) active engagement and participation by parents and others, and (b) a recognition of the complexity of schools and education, and of the needs, aspirations and learning of individual students and the circumstances of their families and communities in relationship with schools.

References

Australian Bureau of Statistics (ABS) 2008, *Census of Population and Housing: Socio-Economic Indexes for Areas (SEIFA), Australia - Data only, 2006*, Cat. No. 2033.0.55.001, viewed 20 June 2010,

<http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/2033.0.55.0012006?OpenDocument>

Australian Curriculum, Assessment and Reporting Authority (ACARA) 2010, *Technical Paper: Index of Community Socio-Educational Advantage*, viewed 20 June 2010,

<http://www.myschool.edu.au/Resources/pdf/My%20School%20ICSEA%20TECHNICAL%20PAPER%2020091020.pdf>

Baker, Joanna & Adhikari 2007, 'Socio-Economic Indexes for Individuals and Families', ABS Cat. No. 1352.0.55.086, Research Paper, ABS Methodological Advisory Committee, Australian Bureau of Statistics, Canberra, viewed 2 February 2009,
[http://www.ausstats.abs.gov.au/Ausstats/subscriber.nsf/0/87EA1D92F0BB437CCA25733F001F541D/\\$File/1352055086_jun%202007.pdf](http://www.ausstats.abs.gov.au/Ausstats/subscriber.nsf/0/87EA1D92F0BB437CCA25733F001F541D/$File/1352055086_jun%202007.pdf).

Interim Committee for the Australian Schools Commission (Chair: Peter Karmel) 1973, *Schools in Australia*, AGPS, Canberra

Macintyre, Sally & Ellaway, Anne 2000, 'Ecological approaches: Rediscovering the role of the physical and social environment', in Lisa F. Berkman & Ichiro Kawachi (eds), *Social Epidemiology*, Oxford University Press, Oxford.

McMillan, Julie, Jones, Frank L. & Beavis, Adrian 2009, 'A New Scale for Measuring Socioeconomic Status in Educational Research: Development and validation of the Australian Socioeconomic Index 2006 (AUSEI06)', Paper presented at the 2009 AARE International Education Research Conference, Canberra: National Convention Centre, viewed 10 June 2010,
<http://www.aare.edu.au/09pap/mcm091513.pdf>

ATTACHMENT

Notes on the ecological fallacy when area-based indexes of disadvantage/advantage are applied to schooling in Australia⁴

Barbara Preston

barbara.preston@netspeed.com.au Phone: (02) 6247 8919

28 March 2010

The Australian Bureau of Statistics has strongly cautioned against the inappropriate use of area-based indexes of disadvantage (and/or advantage) such as the ABS's own four Socio-Economic Indexes for Areas (SEIFAs) to indicate the disadvantage (and/or advantage) of individuals living in the area. This is in recognition of the heterogeneous nature of areas as small as the ABS Census Collection District (CD) of around 250 households. Such inappropriate use involves the 'ecological fallacy'. The appendix provides detail of ABS's concerns.

Area-based indexes are used as an indicator of the relative disadvantage or advantage of Australian schools for the purposes of providing funding to nongovernment schools (the DEEWR SES measure) or for providing the basis for judging the relative performance of every Australian school (the My School Index of Community Socio-Educational Advantage, ICSEA). These indexes use the most recent ABS Census data and are similar to the SEIFAs, but with some different components and weightings judged more appropriate to education-specific disadvantage/advantage. The indexes are used to obtain a score for each CD, in the same way as SEIFA is used to obtain a score for each CD. A school is then scored according to the average score of the home CDs of every student (or a representative sample of students). Thus a very small school that drew students from just one CD would have a score directly based on the score of that CD, and one that drew equally from ten CDs would have a score directly based on the average scores of those ten CDs. While these indexes are not making assessments of the disadvantage/advantage of individuals as their final outcome, their *central mechanism* involves making such assessments of the disadvantage/advantage of individuals. Thus the cautions of ABS are relevant.

Not only is there great diversity of family SES levels within CDs – the heterogeneity that ABS refers to – but, most importantly, *the different schools attended by children in the same CD is not randomly determined*. In discussions around the Commonwealth Government's SES funding model for nongovernment schools some very obvious cases were raised. In particular, the situation where high fee metropolitan boarding schools were assessed as not very high SES because of the large number of students who were the children of high income/high wealth broad acre farmers who in live in CDs classified as relatively low SES because of the many low SES station hands and rural village residents in those CDs. The children in those lower SES families attend local schools, not high fee boarding schools. The corollary is that those local schools, not attended by the farm-owners' children, would be classified as higher SES because of those high SES farm-owners in the same CDs as the local students. Such instances, though stark and easy to explain, are not anomalies. The same pattern operates throughout Australia – to a greater or lesser degree.

⁴ The ABS 2001 Census data used in these notes was obtained as custom tables by the author for another purpose six years ago. Similar analyses could be carried out using 2006 Census data, which is the basis of the current DEEWR SES and ICSEA indexes.

The Census does not (of course) provide data on the actual schools attended by school age students, but it does provide data on level of schooling (primary or secondary) and type of school (government, Catholic and other nongovernment/independent) in a classification, 'Type of Educational Institution Attending' (TYPP). The Census also provides data on Family Income (FINF) which can be applied to individual school students⁵. It is thus possible to analyse school students in every CD (scored by an appropriate index according to advantage/disadvantage) by FINF⁶ and TYPP (the type of school attended by level and sector).

Custom 2001 Census data for the population of all secondary school students in Statistical Local Area (SLA) of Penrith in NSW was obtained (Preston 2004), with variables of FINF (divided into approximate thirds – LOW, MEDIUM and HIGH), TYPP (government, Catholic or independent secondary school), and CD. The CDs were classified according to the SEIFA index of advantage/disadvantage. Some of the findings from this dataset include:

- Across the whole Penrith SLA, a public school secondary student is *equally likely* to have a HIGH or LOW family income, while a private (Catholic or independent) school student is *more than twice as likely* to have a HIGH as a LOW family income.
- In the ten most *advantaged* CDs, a public school student is *four* times as likely to have a HIGH as a LOW family income, while a private school student is *ten* times more likely to have a HIGH as a LOW family income.
- In the ten most *disadvantaged* CDs, a public school student is *sixteen* times more likely to have a LOW as a HIGH family income, while a private school student is only *three* times more likely to have a LOW as a HIGH family income (and an independent school student is *equally likely* to have a HIGH as a LOW family income).

Thus, if drawing from just the ten most disadvantaged CDs, a representative public school would have 16 disadvantaged (LOW family income) students for every one advantaged (HIGH family income) student, while a representative independent school drawing only from the same CDs would have equal numbers of disadvantaged and advantaged students. Yet if an index that was similar to the SEIFA index of advantage/disadvantage (such as the DEEWR SES index or the ICSEA) was applied, both schools would be classified as equally disadvantaged.

The Census does not differentiate between selective/specialist and comprehensive public schools, or between high and low fee private schools. Thus the differences between, say, comprehensive public schools and high fee private schools are likely to be more pronounced than the general differences between public and private schools noted above. In addition, analysis of successive Censuses, including the 2006 Census, shows increasing divergences between the school sectors, with an increasing concentration of LOW family income students in public schools, and an increasing concentration of HIGH family income students in Catholic and independent schools (Preston 2007).

In conclusion, the use of area-based indexes of disadvantage/advantage are not appropriate for measuring the disadvantage/advantage of schools in Australia, especially for high stakes purposes and where there will be comparisons ('matching') between schools from different sectors or which have different formal or informal selection and exclusion practices.

⁵ But not those at boarding school or otherwise living away from home on Census night, and some other categories.

⁶ Or a family SES measure such as that developed by Baker & Adhikari (2007)

Appendix

The ABS cautions against the use of SEIFA indexes as a proxy for individual disadvantage/advantage.

The Australian Statistician, Brian Pink, points out that ‘SEIFA says nothing about the socio-economic status of a particular person or household. ...It is not correct to use SEIFA to say any particular person is relatively disadvantaged’ (Pink 2006, p. 15).

In an ABS research paper, Pramod Adhikari develops the argument further:

The relative disadvantage (or advantage) that SEIFA indexes summarise at CD level should ... be used as contextual variables and not as indexes for individuals living in that area. If we attempt to explain individual level disadvantage from area level measure such as SEIFA then we are assuming that the relationships observed for areas hold for individuals. As we know people living in a CD are not homogeneous – everyone living in a disadvantaged CD will not all be equally disadvantaged. There will be some people who are less disadvantaged than others even in the most disadvantaged areas. The ecological fallacy is a result of the assumption that relationships observed for areas also hold for the individuals who live there. If we assign an area level index to an individual then there is a risk of an ecological fallacy. (Adhikari 2006, p.6)

The evidence for the substantial nature of the potential problem of the ecological fallacy in the use of SEIFA (or any similar area-based index) as a proxy for individual measures of disadvantage in Australia is provided in a 2007 ABS Methodological Advisory Committee research paper (Baker & Adhikari 2007) which uses Census data for Western Australia and ‘explores the feasibility of creating individual and family level socio-economic indexes using the same conceptual and methodological basis as SEIFA’ (p. 1). The authors conclude:

We used the individual and family indexes to examine whether there is a high risk of an ecological fallacy if the IRSD [Index of Socio-Economic Disadvantage] is used as a proxy for individual or family level disadvantage. Our analysis found that individual and family relative socio-economic disadvantage was quite diverse within areas. This means that there is a high risk of an ecological fallacy if we use the SEIFA indexes as a measure of individual level disadvantage, rather than a measure of area level disadvantage. (p. 25)

References

- Adhikari, Pramod 2006, ‘Socio-Economic Indexes for Areas: Introduction, Use and Future Directions’, Research Paper, Australian Bureau of Statistics, Canberra, viewed 2 February 2009,
[http://www.ausstats.abs.gov.au/Ausstats/subscriber.nsf/0/12AC236185F4B054CA2571F60017AE2C/\\$File/1351055015_sep%202006.pdf](http://www.ausstats.abs.gov.au/Ausstats/subscriber.nsf/0/12AC236185F4B054CA2571F60017AE2C/$File/1351055015_sep%202006.pdf).
- Baker, Joanna & Adhikari 2007, ‘Socio-Economic Indexes for Individuals and Families’, ABS Cat. No. 1352.0.55.086, Research Paper, ABS Methodological Advisory Committee, Australian Bureau of Statistics, Canberra, viewed 2 February 2009,
[http://www.ausstats.abs.gov.au/Ausstats/subscriber.nsf/0/87EA1D92F0BB437CCA25733F001F541D/\\$File/1352055086_jun%202007.pdf](http://www.ausstats.abs.gov.au/Ausstats/subscriber.nsf/0/87EA1D92F0BB437CCA25733F001F541D/$File/1352055086_jun%202007.pdf).
- Pink, Brian 2006, ‘An Introduction to Socio-Economic Indexes for Areas (SEIFA)’, Information Paper, Australian Bureau of Statistics, Canberra, viewed 6 March 2009,

[http://www.ausstats.abs.gov.au/Ausstats/subscriber.nsf/0/D729075E079F9FDECA2574170011B088/\\$File/20390_2006.pdf](http://www.ausstats.abs.gov.au/Ausstats/subscriber.nsf/0/D729075E079F9FDECA2574170011B088/$File/20390_2006.pdf).

Preston, Barbara 2004, 'Choice and national schools policy', A paper presented at the annual conference of the Australian Association for Research in Education, Melbourne, 28 November to 2 December 2004

Preston, Barbara 2007, *The Social Make-up of Schools: Family Income, Religion, Indigenous Status, and Family Type in Government, Catholic and other Nongovernment Schools*, Australian Education Union, Melbourne, viewed 21 April 2008,
<http://www.aeufederal.org.au/Publications/Bprestonrep2007.pdf>.