Geolocation Questions and Coding Index

A technical report submitted to the

Performance Measurement and Reporting Taskforce
of the
Ministerial Council on Education, Employment, Training and Youth Affairs

by

Dr Roger Jones

Quantitative Evaluation and Design Pty Ltd

November 2004

Table of contents

List of Abbreviations

ABS   Australian Bureau of Statistics
ARIA   Accessibility/Remoteness Index of Australia
ASGC   Australian Standard Geographical Classification
CD   Collection District
LGA   Local Government Area
MCEETYA   Ministerial Council on Education, Employment, Training and Youth Affairs
NEPMT   National Education Performance Monitoring Taskforce
NLI   National Localities Index
PMRT   Performance Measurement and Reporting Taskforce
RRMAC   Rural, Remote and Metropolitan Areas Classification 1991 Census Edition
SD   Statistical Division
SLA   Statistical Local Area
StatD   Statistical District
SSD   Statistical Sub-division

Geolocation Questions and Coding Index

1. Objectives

The objectives of this project are to:

- Develop appropriate geolocation question(s) to elicit information from secondary school students on the locality and postcode of their permanent home address;
- Develop a coding index to code the information on locality and postcode to the *MCEETYA Geographical Location Classification*; and
- Provide advice to ensure that the coding index is capable of being maintained over time, with complete instructions on managing the system.

2. Background

When Ministers agreed, in 1999, to new National Goals for Schooling, they identified eight priority areas for reporting progress against the national goals, and specified that these should be reported on in a nationally consistent way. Reporting was to occur not only for students overall but also by gender, for Indigenous and geographically isolated students, for students from language backgrounds other than English, students with disabilities, and students from socioeconomically disadvantaged backgrounds.

The Commonwealth/State/non-government taskforce set up to achieve this reporting (initially the National Education Performance Monitoring Taskforce or NEPMT and currently the Performance Measurement and Reporting Taskforce or PMRT) established a Nationally Consistent Definitions Sub-group to produce agreed definitions of the various student groups including geographically isolated students.

In 1999 the Taskforce commissioned the author to develop an approach to defining the geographic location of students. The project report (Jones, 2000) recommended a geographic classification that deliberately sought to maintain reasonable comparability with the Rural, Remote and Metropolitan Areas Classification 1991 Census Edition (RRMAC) developed jointly by the then Department of Primary Industry and Energy (DPIE) and the Department of Human Services and Health (DHSH) (1994). In particular, Jones retained the concept, while changing the specific categories, of a *Metropolitan Zone* based on the Statistical Division (SD) or Statistical Sub-division (SSD) of the State/Territory capital cities and major urban centres. In addition, the RRMAC concept of Large Rural Centres was maintained (renamed as *Provincial City Regions*), based on the knowledge that the Australian Bureau of Statistics (ABS) was in the process of revising the Statistical District (StatD) Structure of the 2001 Australian Standard Geographical Classification (ASGC) to define regions around all provincial cities of 25,000 or more population.

For the remaining categories, defined as *Other Provincial Areas* and the *Remote Zone*, Jones drew on the work of the Accessibility/Remoteness Index of Australia (ARIA) project. Following the revisions to ARIA and the release of the ABS Remoteness Structure, and recent up-dates to the definition of Statistical Districts incorporated in the 2001 ASGC, the revised definition of geographic location adopted by MCEETYA is shown in Figure 1.

Figure 1: MCEETYA Geographical Location Classification, July 2001

1. Metropolitan Zone

> 1.1 Mainland State Capital City regions (Statistical Divisions)
> Sydney, Melbourne, Brisbane, Adelaide and Perth SDs

> 1.2 Major urban Statistical Districts (100,000 or more pop'n)
> ACT-Queanbeyan, Cairns, Gold Coast-Tweed, Geelong, Hobart,
> Newcastle, Sunshine Coast, Townsville, Wollongong

2. Provincial Zone (non-remote)

> 2.1. Provincial City Statistical Districts plus Darwin SD

>> 2.1.1 Provincial City Statistical Districts and Darwin SD (50,000 or more
>> pop'n)
>> Albury-Wodonga, Ballarat, Bathurst-Orange, Bundaberg, Burnie-
>> Devonport, Bendigo, Darwin, Launceston, La Trobe Valley, Mackay,
>> Rockhampton, Toowoomba, Wagga Wagga

>> 2.1.2 Provincial City Statistical Districts (25,000 - 49,999 pop'n)
>> Bunbury, Coffs Harbour, Dubbo, Geraldton, Gladstone, Shepparton, Hervey
>> Bay, Kalgoorlie-Boulder, Lismore, Mandurah, Mildura, Nowra-Bomaderry,
>> Port Macquarie, Tamworth, Warrnambool

> 2.2 Other provincial areas (CD ARIA index value $\leq$ 5.92)

>> 2.2.1 Inner provincial areas (CD ARIA index value $\leq$ 2.4)

>> 2.2.2 Outer provincial areas (CD ARIA index value > 2.4 and $\leq$ 5.92)

3. Remote Zone

> 3. Remote Zone (CD ARIA index value > 5.92)

>> 3.1 Remote areas (CD ARIA index value > 5.92 and $\leq$ 10.53)

>> 3.2 Very Remote areas (CD ARIA index value > 10.53)

As in RRMAC, areas included in the *Metropolitan Zone* are defined by their Statistical Division (SD) or Statistical Sub-Division (SSD). The *Major Urban* regions have been expanded by the

inclusion of the SSDs of Cairns and Sunshine Coast, while Darwin SD is assigned to the *Large Provincial City* regions with other urban centres of similar population size. The Sunshine Coast and Cairns were included in the *Metropolitan Zone* on the belief that both would achieve the status of Major Urban Centres of population 100,000 or more in the 2001 Census. In the case of the Sunshine Coast, which in 1996 had three large, but separate, urban centres of Caloundra, Maroochydore-Mooloolaba and Tewantin-Noosa, this occurred with the ABS defining a new single urban centre of Sunshine Coast with a population of 165,089. Cairns, however, with an urban centre population of 90,085 has not achieved Major Urban Centre status, since Cairns Northern Beaches remains a separate urban centre of 15,245 population within the Cairns SSD. It is not proposed, however, that any change should be made to the classification on account of this.

The *Large Provincial Cities* and *Small Provincial Cities* of the classification reflect the definition of Large Rural Centres of population 25,000 or more in RRMAC. The fact that all centres of population 25,000 or more have now been incorporated into the Statistical District (SDist) Structure of the 2001 ASGC (and into the Main Structure as SSD) was an important consideration in the definition of these categories.

Outside of these major and provincial city centres, the definition of *Inner* and *Outer Provincial*, *Remote* and *Very Remote* areas follows the criteria adopted by the ABS for the definition of Inner and Outer Regional, Remote and Very Remote classes in the ASGC Remoteness Structure. For the remote areas, the ABS defined cut-off ARIA score of 5.92 was a slight adjustment to the score originally proposed in the NEPMT report (Jones, 2000) to maintain comparability with RRMAC (ABS, 2001a).

Provincial City regions are grouped into large and small Provincial City Statistical Districts on the basis of their regional (SSD/StatD) population, the division being made originally at the 50,000 population size. However, recently released 2001 Census estimates of the usually resident population of these regions (ABS, 2002a and 2002b) identify Mandurah SSD with a population of 56,926 which should place it in the large Provincial City Statistical Districts category. Wagga Wagga, with an urban centre population of 44,272 and a SSD/StatD population of 49,993, and Bunbury, with an urban centre population of 45,153 and a SSD/StatD population of 47,610, appear likely to exceed the 50,000 population size boundary by the next Census in 2006. As above, it is not proposed that changes should be made to the classification at this stage.

ARIA scores and the ASGC Remoteness Structure will be updated by the ABS for release in conjunction with 2006 Census data. The MCEETYA Geographical Location Classification will then also need updating, to maintain consistency with the ASGC Remoteness Structure and take account of any changes in the population of metropolitan and provincial city regions.

3. Coding requirements

A further issue that had to be considered in the derivation of the *MCEETYA Geographical Location Classification* was the ability to code home address data to the categories. While it is expected that geo-coding systems will be developed in the next few years which will allow every address to be linked to its latitude and longitude coordinates and hence to its CD and Remoteness Structure class, automated matching of rural and remote addresses to CDs still has some way to go. Postcode alone is not sufficient to determine accurately the Remoteness Structure class of all addresses, particularly in rural and remote areas.

The ABS Geography Section maintains the National Localities Index (NLI), a comprehensive list of the locality names in current use and links between these locality names, their postcode and SLA. Given the place name, State/Territory and postcode of an address, the NLI then provides all that is required to code an address to its SLA. In particular, addresses within the *Metropolitan* and *Provincial City* regions of the classification can be assigned to their appropriate category, since these regions are defined by aggregates of complete SLAs within SD and SSD. In the majority of cases, addresses can be assigned to these categories on the basis of postcode alone, with the NLI only required where postcodes overlap the category boundaries.

In addition, the ABS have recently added Remoteness Structure categories to some 22,000 of the localities in the NLI for which latitude and longitude were known. For addresses outside the *City* regions, the ARIA score of the suburb/urban centre/locality can be identified in the great majority of cases and thus assigned to its *Other Provincial* or *Remote Zone* category. For the remaining cases where the NLI does not include the ARIA score of a locality, there will often be sufficient information on other localities in the same postcode to accurately determine the appropriate category.

4. Coding to the MCEETYA Geographical Location Classification –nli-ra coder

As indicated above, coding into the categories of the *MCEETYA Geographical Location Classification* is based on a version of the NLI that includes Remoteness Structure codes for those localities whose latitude and longitude were known. It should be noted that localities are not in general single points, but areas which may in some cases be quite large and whose boundaries are not clearly defined, so it is possible for some parts of the localities to be misclassified. In addition, a substantial number of the localities included in the NLI have not been assigned to a Remoteness Structure category, although many of these unassigned localities represent alternative spellings of assigned localities or are places with very small or no permanent population. More information on the NLI is given in *Explanatory Notes for the National Localities Index (NLI)* (ABS (2001), Cat No 1252.0).

The copy of the NLI Localities Index, with Remoteness Structure codes assigned, used to derive the *MCEETYA Geographical Location Classification* coding index was supplied to the author by the ABS Geography Section in October 2002. It has not been updated since that time, nor is it likely to be updated in the future, although the Localities Index alone, without Remoteness Structure codes, is updated and released quarterly.

The author has matched this October 2002 release to the most recent Localities Index quarterly release available from the ABS, that of April 2004, adding Remoteness Structure codes to the current Localities Index. Uncoded locality names have, in some cases, been assigned geographical location codes by the author on the basis of their SLA (in the case of Metropolitan and Provincial City categories), the evidence of other codes assigned in the same postcode and/or SLA, or in some cases from ARIA scores assigned to localities by GISCA or by ABS in their coding of the MCEETYA Schools Geographic Location Database 2002.

This worksheet, *nli-ra coder* included in the Excel workbook *PMRT_NLI_Coding*, includes 31,956 locality names and is the primary coding index for the *MCEETYA Geographical Location Classification.* The fields (columns) in the coding index worksheet, and a brief description, are as follows (see *Explanatory Notes for the NLI* (ABS, 2001d) for more details on fields A-F):

A. SLA_2001   the code for the major SLA which contains the locality.

B. SLA_NAME   the name for the major SLA.

C. Pcode   the Australia Post postcode for the locality. Note that a few localities have more
than one postcode and only the first postcode has been retained. Note also the
disclaimer that "While the ABS regularly checks postcodes on the NLI against
Australia Post postcode lists, the postcode associated with a Locality on the NLI
may not be the Locality's current postcode."

D. LOCALITY   the locality name.

E. PI   property indicator field.

F. STATE   the State/Territory code.

G. PC code   the *MCEETYA Geographical Location Classification* code assigned on the basis
of postcode alone, when all localities within the postcode are in a single category
of the classification. See 9. below. If a postcode overlaps a boundary between
categories, "Locality?" is entered for all localities within the postcode.

H. blank

I. Enter locality   entries use the function CONCATENATE to create a unique field combining
"PC1", "_" and "LOCALITY". This is used as an index for matching postcode
and locality name when assigning a code to a locality.

J. Code   the *MCEETYA Geographical Location Classification* code assigned to the
locality. Codes are:

Code	Name
1.1	Metropolitan Zone Mainland State Capital City regions
1.2	Metropolitan Zone Major urban Statistical Districts
2.1.1	Provincial Zone Provincial City Statistical Districts 50,000 to 99,999
2.1.2	Provincial Zone Provincial City Statistical Districts 25,000 to 49,999
2.2.1	Provincial Zone Inner provincial areas
2.2.2	Provincial Zone Outer provincial areas
3.1	Remote Zone Remote areas
3.2	Remote Zone Very Remote areas

K. Name   Names are assigned to codes as follows:
For categories 1.1 to 2.1.2 above, the name of each region within the categories
is given – eg Sydney SD, Port Macquarie StatD, etc
For categories 2.2.1 through 3.2 the names used are those assigned by the ABS
on the original file –ie INNER REGIONAL AUSTRALIA, OUTER REGIONAL
AUSTRALIA, REMOTE AUSTRALIA, and VERY REMOTE AUSTRALIA.

A simple procedure for assigning the *MCEETYA Geographical Location Classification* codes has
been included at the top of the worksheet for illustration. The procedure works as follows:

Step 1   Enter postcode in E8. If the postcode can be assigned directly to a category, the appropriate code will be returned in F8. Otherwise:

Step 2   If #N/A is returned, the postcode is not included in the NLI and is invalid. This may be a post office (PO) box postcode, which are not included in the NLI, or because no such postcode exists currently. Many PO box postcodes can be assigned to a category and could be included in the coding index, although the recommended question aims to discourage PO box addresses and encourage location addresses.

Step 3   If "Locality?" is returned, additional information – ie the locality name – is required to determine the geographical location category. Enter the locality name in H8. This entry is case insensitive – ie upper or lower case, or a combination, may be used. If the locality name is found <u>within the postcode listing</u>, a code will be returned in I8.

Step 4   The return of "9" indicates that the locality was found but there is NO RA CODE ALLOCATED – ie that the ABS has not yet assigned that locality to a Remoteness Structure category. The return of "9" can only occur for localities in categories 2.2.1 through 3.2, and the population associated with these localities is expected to be small.

Step 5   The return #N/A indicates that the locality was <u>not</u> found within that postcode. This may be a postcode error, a locality name misspelling, or a locality that is not included in the NLI. To check these cases, use the 'Find' procedure (in the Edit pull-down menu of Excel or Ctrl+F) to search the LOCALITY column (column D) for postcode error, or browse the locality listing for the given postcode for alternative spelling.

Suggested additional codes for uncoded cases are:
>9   Locality listed in valid postcode but code not yet assigned (Step 4)
>98   Valid postcode but Locality not listed (#NA returned in Step 5)
>99   Invalid postcode (#NA returned in Step 2)

5. Coding to the MCEETYA Geographical Location Classification –pc-book coder

A second coding index has been developed using the pc_book file available from the Australia Post website (www.auspost.com.au). This file is a listing of place names and their postcodes current at the time of issue, as printed in the White Pages telephone directory. An edited version of this data, current at 30 April 2004, is included in the Excel workbook *PMRT_NLI_Coding* as the worksheet *pc-book coder*. The editing of the pc-book file included the deletion of all PO box and PO bag postcodes, and entries with locality names which refer to District Council (DC) or Borough Council (BC) areas.

In effect this is a shortened version of the nli-ra coder, with 9861 postcode/locality name combinations listed and coded to the *MCEETYA Geographical Location Classification.* Codes have been assigned using the nli-ra coder, showing that it is a relatively quick and simple task, given a listing of almost 10,000 locality names and postcodes, to assign *MCEETYA Geographical Location Classification* codes to the vast majority of cases.

The fields (columns) in the coding index worksheet, and a brief description, are as follows:

A. Pcode   the Australia Post postcode for the locality.

B. Locality   the locality name assigned in the Australia Post listing.

C. State   the State/Territory acronym.

D. PC_code   codes obtained by entering the appropriate VLOOKUP function statement into any empty cell of row 12, namely

=VLOOKUP(A12,'nli-ra coder'!$C$12:$G$31956,5,FALSE)

This Excel function matches the contents of A12, the postcode value 800, against the entries in the first column (C) of the array 'nli-ra coder'!C$12:G$31956 and returns the contents of the fifth column (G) of that array, the PC_code, from the first row in which an exact match is found. Copying and pasting this function down then returns the appropriate PC_code for the other postcode entries in column A. Function entries are then converted to values using Copy and Paste Special + Values.

E. blank

F. Enter locality   Entries use the Excel function CONCATENATE to create a unique field combining "Pcode", "_" and "Locality". This is used as an index for matching postcode and locality name when assigning a code to a locality.

G. Code or #NA   When "Locality?" is returned for PC_code (as for example for the entries in postcode 822 starting in row 47) the codes in Column G are obtained by entering the appropriate VLOOKUP function statement into an empty cell in that row, namely

=VLOOKUP(F47,'nli-ra coder'!$I$12:$J$31956,2,FALSE)

This function matches the contents of F47, the concatenated postcode_locality name, against the entries in column I of the nli-ra coder and returns the contents of column J from the first row in which an exact match is found. Copying and pasting this function down then returns the appropriate code for other rows with "Locality?" in Column D. Function entries are converted to values using Copy and Paste Special + Values.

Column G combines the codes obtained either directly from the postcode or from the "Locality?" match.

6. A test coding

The *MCEETYA Schools Geographic Location Database 2002* has been used to test the accuracy and performance of both the *nli-ra coder* and the *pc-book coder*. The database variables

L_addr_2	Location Address Line Two
L_pcode	Location postcode
State	State/Territory

are indicative of the locality data that would be obtained from a survey question, and schools have already been coded to the *MCEETYA Geographical Location Classification* using ARIA+ scores derived on the basis of each school's latitude and longitude coordinates.

Of the 9861 schools included in the database, 91 had postcodes which returned "#N/A" indicating PO box or PO bag postcodes, or postcodes not included in the current pc-book listing. 48 of these schools are in remote or very remote parts of the Northern Territory, where the variable L_addr_2 is often used to record the main urban centre in the region, for example Alice Springs or Tennant Creek, and the school locality is given either in L_addr_1 or simply in the name of the school. Using this information, it was possible to code all of the 91 schools to the correct category using the *nli-ra coder*, although fewer localities (64) were included in the *pc-book coder*.

Of the remainder, 7235 schools are coded to their *MCEETYA Geographical Location Classification* category using their Location postcode only, with agreement between both coding indexes and the Schools database coding in 7220 cases. Only 15 cases have non-matching codes, of which 5 are clearly identified as potential problems by the inclusion of "Via" in the locality name variable. Four errors arise when a wrong, but valid, postcode is given and codes are assigned on the basis of that postcode alone. One of these cases, the Yintarri Remote Community School, could not be correctly coded as Very Remote on the basis of the information given, Yintarri not being included in the NLI and an address of Boulder 6432 being coded to the Kalgoorlie-Boulder StatD (although Remote Community School provides a strong hint of the likely error). The remaining six cases were referred to DEST who have advised that the codes will be revised in the 2003 edition of the database.

One approach to reducing this type of error is to require that both the postcode and locality name match before a code is assigned. Appling this criterion, 6678 of the 7220 cases coded on the basis of postcode alone have an exact match on the locality name using the *pc-book coder* with the remaining 542 cases returning "#N/A", while the *nli-ra coder* gives 6878 exact matches and 342 unmatched "#N/A" returns. Each of the "#N/A" returns should then be examined to identify whether it is simply the result of a misspelt or different form of the locality name, or is the locality listed in a different postcode and thus possibly in a different category.

In the Schools database, a substantial number of unmatched locality names result from the inclusion of "Via" in the location address. For example, all schools in the Torres Strait Island have "Via Thursday Island" in the location field. If this is a common feature of addressing in remote areas, it may be important to provide some form of instruction to students to avoid this style of address, either on the questionnaire itself (although this may be cumbersome) or through staff who are distributing forms and should be in a position to know whether this problem is likely to occur.

Another common reason for unmatched locality names is associated with the use of North, South, East and West, Upper and Lower etc in locality names. The NLI generally puts these qualifiers after the main locality name (eg Cooma North not North Cooma) whereas Australia Post tends to do the reverse, although neither are consistent. The NLI also excludes all punctuation (eg Oconnor rather than O'Connor and Kuring gai rather than Kuring-gai) while pc-book does not. These are easily checked, but checking adds to the time required for coding.

The remaining 2525 schools listed in the database require both postcode and locality name matching. Using the *nli-ra co*der, 2176 are assigned codes matching those on the database, 80 are assigned non-matching codes and 279 return "#N/A". Similarly, using the *pc-book c*oder, 2081 have matched codes, 76 have non-matching codes (included in the 80 above) and 378 return

"#N/A". These "#N/A" returns are again mostly due, as above, to addressing "Via" larger nearby towns, alternative versions of locality names or postcode error. After checking, only one school could not be coded using the *nli-ra coder*, Wallungurru (purportedly in Alice Springs 872) not being included in the NLI.

The 80 assigned codes not matching those on the Schools database were of more concern. More than half (46) appear to be errors in the Schools database coding. These were referred to DEST who advise that all coding is being checked for the 2003 update of the database. Twenty-eight of the remainder could be correctly coded if the more precise locality, often given in the school name, was used. For example, Kingston Public School, Bundarra-Bendemeer Road, Uralla NSW 2358 is coded inner provincial, the correct category for Uralla (ARIA plus 2.34) but Kingston itself is outer provincial (ARIA plus 4.239) in the schools database and in the locality coders. Another 4 schools were of this type, but the precise locality name (Drinane v Dubbo, Blackall Range v Nambour, Miltaburra v Streaky Bay, Yuna v Geraldton) was not included in the NLI and therefore they could not be coded. The remaining two miscoded schools were on the division between inner and outer provincial areas – Taralga 2580 has two schools, one inner (ARIA plus 2.35) and one outer (ARIA plus 2.458) and is coded inner in the locality coders, and Kadanga Creek State School in Upper Kadanga 4570 is outer provincial if in Upper Kangdanga but inner provincial if in Kadanga or Kadanga Creek.

In summary, the positive results of this test indicate that, of the 9861 schools listed on the MCEETYA Schools database,
- only 8 could not be coded to their correct *MCEETYA Geographical Location Classification* category on the basis of their correct locality name and postcode; and
- only 9 of the 7235 schools assigned codes on the basis of their postcode only were wrongly classified.

The first of these findings indicates that there is little need for concern about the coverage of the locality coders, particularly the *nli-ra coder* with almost 32,000 localities listed. Nevertheless, there are some localities which appear not to be listed, and there are others that have not yet been assigned to a category of the classification. The significance of these omissions can be determined and, if necessary, addressed over time. The second finding is important in showing that coding on the basis of postcode alone has not been over-extended and that, while some errors may arise, they should be very few. Given appropriate locality and postcode information, the evidence of this test is that coding to the *MCEETYA Geographical Location Classification* will, in the vast majority of cases, be accurate.

One concern raised by this test is the frequency with which schools located in small communities in remote areas are given or give addresses "via" another larger locality. Where the information is being collected through a survey question, it is important that the response relates to the actual locality of the respondent and not an alternative larger locality somewhere in the vicinity.

7. The geolocation question

The information required to code a location to the *MCEETYA Geographical Location Classification* is the postcode and, in many cases, the NLI locality name, defined very broadly as "the name of a place where people live or work – or say they live or work" (See *Explanatory Notes for the NLI,* p1). The information required from survey respondents is therefore, the name and postcode of the place where they live.

For surveys of secondary school students, a suggested question is:

Q. **Where do you live?** Please write in below the <u>place name</u>, <u>State/Territory</u> and <u>postcode</u> of your ***permanent home address*** (ie the last line of your <u>home</u> address).

[If you are boarding away from home, please think of your *permanent home address*.]

[If you have a PO Box, please think of your *home* rather than the PO Box address.]

⬜⬜⬜	⬜⬜⬜⬜

(Place name)   (State/Territory eg NSW)   (Postcode)

An almost identical question to this was successfully trialed in the *Survey on Parental Background of Students Commencing Higher Education in 2001* (Jones, 2001) with only 154 of 20,713 students not responding to the question.

The question places considerable emphasis on *permanent home address* and *home*, attempts to exclude PO Box postcodes from responses and includes State/Territory in addition to the locality name and postcode, drawing the parallel between this information and the way that the last line of an address is usually written.

One concern raised by that parallel is that it may, in remote areas, encourage "via" responses if that is common practice. Responses such as 'Melville Island via Darwin" or "Santa Teresa via Alice Springs" would not be a problem, but simply "via Darwin" clearly would be. Instructions to schools and teachers who are distributing the forms may be one approach to addressing this concern, but there is no certainty that instructions will be relayed to students.

An alternative approach which places greater emphasis on the need to obtain accurate locality information first, and then seeks State/Territory and postcode information, is the following:

Q. **Where do you live?** Please write in the <u>place name</u> of your ***permanent home address****,* not the street but the name used to refer to the **local** area or community where you live.

[If you are boarding away from home, please think of your *permanent home address*.]

_____

Q. What is the <u>State/Territory</u> and <u>postcode</u> of your ***permanent home address***?

[If you have a PO Box, please think of your *home* rather than the PO Box address.]

⬜⬜⬜	⬜⬜⬜⬜

(State/Territory eg NSW)   (Postcode)

The use of two questions is felt, on reflection, likely to yield better locality name information from younger respondents and is preferred to the single question.

8. The coding procedure

Where the geolocation question(s) is used in a questionnaire such as for the 2004 Year 10 national sample civics and citizenship education test form, it is recommended that:

- the locality, state and postcode information elicited by the question be entered verbatim as text for subsequent coding to the *MCEETYA Geographical Location Classification.*

All geolocation data can then be coded together in a single operation, using procedures essentially identical to those discussed above. Three separate variables will then be created, namely:

- State/Territory of permanent home address;

- Postcode of permanent home address; and

- *MCEETYA Geographical Location Classification* category of permanent home address.

Suggested procedures for coding to the *MCEETYA Geographical Location Classification* are as follows. Note that steps 1-4 and 8-9 are the essential steps for allocating codes to the vast majority of cases. Steps 5-7 are a check on ineligible postcodes, and step 10 deals with unmatched locality names. Step 11 combines the codes obtained into a single column.

1. Insert the case identifier, locality name, State/Territory and postcode information into columns A, B, C and D respectively of an Excel worksheet.

2. The *nli-ra coder* and *pc-book coder* give identical results when assigning codes on the basis of postcode alone, and use of the smaller *pc-book coder* is recommended. Assuming data is entered from row 2 down (postcodes in column D starting in cell D2), enter the VLOOKUP function statement in cell E2:

    =VLOOKUP(D2,'[PMRT_NLI_Coding.xls]pc-book coder'!$A$12:$D$9872,4,FALSE)

3. Copy cell E2 and Paste down column E to the end of the data. Copy all of the entries in column E and Paste Special + Values to their current location. All cases that can be coded on the basis of postcode only will then have been coded in column E.

4. Highlight all of the data in columns A-E and sort on column E in descending order. Ineligible postcodes indicated by "#N/A" will be listed first, followed by those requiring "Locality?" and finally those which have been coded to their geolocation category.

5. Check locality names in ineligible postcodes (indicated by "#N/A") against the Locality names in *nli-ra coder* to try to determine eligible postcodes as follows. In cell F2, enter the VLOOKUP function statement

    =VLOOKUP(B2,'[PMRT_NLI_Coding.xls]loc by pc'!$A$12:$B$31967,2,FALSE)

    This checks for the locality name in B2 against a copy of the locality names in *nli-ra coder* that has been stored in alphabetic order in the worksheet *loc by pc* and returns the postcode of the first exact match found. Copy F2 and Paste down against all ineligible postcode entries. Copy all of the entries in column F and Paste Special + Values to their current location.

6. NB. Step 5 returns need to be checked. In most cases, the ineligible and returned postcode will be similar and it will be clear that the returned eligible postcode is correct. However, some locality names have multiple entries in *loc by pc* in different postcodes and Step 5 returns the postcode of the first matched entry. This may be in a different state and clearly wrong, or it may be more difficult to be sure whether the return is correct or not. Step 5 may also return "#N/A" showing that there is no exact match for the locality name entered. Check for alternative spellings or forms manually in *loc by pc* and/or *nli-ra coder*.

7. For all cases where eligible postcodes are identified, enter the correct postcode in column D, group the corrected rows together (using Sort) and repeat steps 2 and 3 on these rows. Clear column F. Replace any remaining "#N/A" with 99. Repeat step 4. Cases requiring "Locality?" will now be listed first.

8. Enter the following CONCATENATE function in cell F2,

    =CONCATENATE(D2, "_", B2)

    Copy and Paste against all cases requiring "Locality?" coding; Copy all of these entries and Paste Special + Values to their current location.

9. For "Locality?" coding, the use of both the *pc-book coder* and the *nli-ra coder* is useful, the results from either being the correct code and the use of both resolving cases such as "Armidale East" (in *nli-ra coder*) and "East Armidale" (in *pc-book co*der).
    For *pc-book coder*, enter the VLOOKUP function statement below in G2, then Copy and Paste against all cases requiring "Locality?" coding; Copy all of these entries and Paste Special + Values to their current location.

    =VLOOKUP(F2,'[PMRT_NLI_Coding.xls]pc-book coder'!$F$12:$G$9872,2,FALSE)

    For *nli-ra coder*, enter the VLOOKUP function statement below in H2, then Copy and Paste against all cases requiring "Locality?" coding; Copy all of these entries and Paste Special + Values to their current location.

    =VLOOKUP(F2,'[PMRT_NLI_Coding.xls]nli-ra coder'!$I$12:$J$31956,2,FALSE)

10. Highlight all data in cases requiring "Locality?" coding and Sort by column G descending and column H descending, bringing the cases with "#N/A" in both columns to the top. These are cases where the postcode – locality name combination given cannot be matched exactly in either the *nli-ra coder* or the *pc-book coder*. This may be due to
    - a different postcode;
    - a difference in the way that the locality name is spelt or worded; or
    - both of the above.

    A postcode difference can be checked using the approach outlined in Steps 5 and 6 above, or manually, by finding the locality name in *loc by pc* and checking the postcode. A locality name difference, particularly a misspelt name, may be more difficult to resolve. Checks against cases from the same class or school that have been successfully coded may indicate the error.
    Code any unresolved cases to 98 in column H.

11. "Locality?" codes in columns G or H need to be combined – replace "#N/A" codes in column H with valid codes from Column G.
Combine codes in columns E and H in column I. Enter the IF function statement below in I2, then Copy and Paste against all cases; Copy all of these entries and Paste Special + Values to their current location.

=IF(EXACT(E2,"Locality?"),H2,E2)

Sort the data in columns A-I by the case identifier in column A and combine with other case data.

9. Maintaining the coding indexes over time

The primary coding index *nli-ra coder* is based on the April 2004 release of the NLI Localities Index, with Remoteness Structure codes assigned, primarily by the ABS Geography Section in October 2002 but with additional geographical location codes assigned by the author on the basis of their SLA (in the case of Metropolitan and Provincial City categories), the evidence of other codes assigned in the same postcode and/or SLA, or in some cases from ARIA scores assigned to localities by GISCA or by ABS in their coding of the MCEETYA Schools Geographic Location Database 2002. The *pc-book coder* uses the 30 April 2004 release of pc-book, with *MCEETYA Geographical Location Classification* codes derived from, and thus consistent with, the *nli-ra coder*.

The NLI Localities Index is updated quarterly and made freely available on the ABS website. The primary purpose of the NLI is to code address information to the Statistical Local Area (SLA) based on geographical units of the Australian Standard Geographical Classification (ASGC).Updates to the Localities Index are undertaken primarily to ensure that any changes to SLA, Local Government Areas (LGA) and other units are reflected in the latest edition of the NLI. New locality names are added, but there are relatively few additions, and others are deleted at times, usually in regions where changes to ASGC unit boundaries are being made or new units created and there is a reason to check the localities in that region against current reference sources.

Similarly, pc-book is updated quarterly and made freely available on the Australia Post website. Postcode changes or new postcodes usually occur in newly developing areas on the fringes of the larger cities, but this is not always the case – a new postcode 4895 has recently been created for Cooktown in far North, and Very Remote, Queensland. The NLI tries to take account of changes to postcodes in its updates but may not be fully up-to-date with all postcode revisions.

Accounting for postcode changes can be achieved as follows:

1. Identify new postcodes in the latest version of pc-book, by sorting and matching the current and previous edition of pc-book. Only new Delivery Area postcodesare required , not Post Office Box or Bag postcodes..

2. Using the Locality name assigned to the postcode, search the *pc-book coder* to determine if the locality is included (highlight column B and enter the locality name in Find).

3. If it is included, it is recommended that both the old and new postcode be retained. Copy the row containing the current entry, insert that copy as a new row in the appropriate place in

the postcode – locality order, change the postcode (column A) to the new postcode, and change column D to "Locality?". This latter change will ensure that only cases which match on both postcode and locality name use the new postcode to assign their *MCEETYA Geographical Location Classification* category.

4. Similarly, insert a matching row in the *nli-ra coder*. Copy the row containing the current entry, insert that copy as a new row in the appropriate place in the postcode – locality order, change the postcode (column C) to the new postcode, and change column G to "Locality?".

5. If the Locality name is not included in the *pc-book coder*, search *nli-ra coder* to determine if it is included there (highlight column D and enter the locality name in Find).

6. If the locality is included in *nli-ra coder* but not *pc-book coder*, insert a new row in *pc-book coder* in the appropriate place in the postcode – locality order, insert the postcode (column A), locality name (column B), state acronym (column C), "Locality?" (column D), Copy Down the concatenate function (Column F) and insert the geolocation category from the *nli-ra coder* (Column G).

7. Similarly, insert a matching row in the *nli-ra coder*. Copy the row containing the current entry, insert that copy as a new row in the appropriate place in the postcode – locality order, change the postcode (column C) to the new postcode, and change column G to "Locality?".

8. If the new locality name is not included in *nli-ra coder*, there is insufficient data to assign it to a geolocation category and additional information is required. Request ABS Geography Section assistance or wait until the new entry appears in the next NLI Localities Index update.

Updating changes made to the NLI Localities Index are made more complex by the fact that it is considerably larger and that changes are more likely to be driven by changes to SLA boundaries and names than by changes to locality names and postcodes, and that new locality names that are added do not include the Remoteness Structure codes needed to assign localities to the Other Provincial and Remote sectors of the *MCEETYA Geographical Location Classification*.

A systematic process of updating the *nli-ra coder* to take account of changes to the NLI is not therefore recommended. The vast majority of changes that are made do not affect the geolocation category coding. When data collection activities show an increase in responses from new localities that do require coding, new entries can be made to overcome the problem.

The procedure then required to update the *nli-ra coder* would then be as follows:

1. Identify new localities not included in the current *nli-ra coder*, and search the most current NLI Localities Index to determine if the locality is included.

2. If it is included, insert a new row in the nli-ra coder in the appropriate place in the postcode – locality order and copy SLA, postcode, locality, PI and State data into columns A-F.

3. If the SLA data indicates that the locality can be assigned to one of the City categories, enter the appropriate codes and description in columns G, J and K and copy down the concatenate function (column I).

4.   Otherwise, the locality has not been assigned to a category, although the ABS Geography Section who maintain the NLI will, in most cases, be able to assign it. Send an e-mail or letter requesting assistance.

5.   In cases where the new locality had appeared in pc-book but where the *pc-book coder* had not been updated (see step 8 above), update *pc-book coder* as indicated in step 6 above.

A complete revision to the coding index will be required following the 2006 Census when an updated edition of ARIA plus will be calculated by the ABS. There may also be some minor changes to the Provincial Cities categories of the *MCEETYA Geographical Location Classification* based on revised population estimates. The new edition of ARIA plus will probably be available in 2008, and ABS should be commissioned around that time to allocate the revised Remoteness Structure codes to the localities listed in the current (July 2008 edition) of the NLI Localities Index. A revised coding index can then be derived using the procedures outlined above in Sections 4 and 5.

References

Australian Bureau of Statistics (2001a), 'ABS views on remoteness, 2001'. ABS Information Paper, Cat. No. 1244.0. Canberra: ABS.

Australian Bureau of Statistics (2001b), 'Outcomes of ABS views on remoteness consultation, Australia, 2001'. (ABS Information Paper, Cat. No. 1244.0.00.001). Canberra: ABS.

Australian Bureau of Statistics (2001c), *Statistical Geography Volume 1 Australian Standard Geographical Classification (ASGC) 2001*. (Cat. No. 1216.0). Canberra: ABS.

Australian Bureau of Statistics (2001d), *Explanatory Notes for the National Localities Index (NLI)*. (Cat. No. 1252.0). Canberra: ABS.

Australian Bureau of Statistics (2002a), 'Selected Social and Housing Characteristics Australia, 2001'. ABS, Cat. No. 2015.0. Canberra: ABS.

Australian Bureau of Statistics (2002b), 'Selected Characteristics for Urban Centres Australia, 2001'. ABS, Cat. No. 2016.0. Canberra: ABS.

Department of Health & Aged Care/GISCA (1999), 'Measuring remoteness: Accessibility/ Remoteness Index of Australia (ARIA)', DH&AC Occasional papers: New Series, No. 6. Canberra: DH&AC.

Department of Primary Industries and Energy/Department of Human Services and Health (1994), *Rural, remote and metropolitan areas classification: 1991 census edition*. Canberra: AGPS.

Jones, R G (2000), 'Development of a common definition of, and approach to data collection on, the geographic location of students to be used for nationally comparable reporting of outcomes of schooling'. Report prepared for the MCEETYA National Education Performance Monitoring Taskforce, unpublished.

Jones, R G (2001), *Identifying Higher Education Students from Low Socio-Economic Status Backgrounds and Regional and Remote Areas*. Canberra: Higher Education Division, DEST.