1996 CENSUS DATA QUALITY: QUALIFICATION LEVEL AND FIELD OF STUDY

Euan Robertson

Population Census Evaluation
August 2000

Produced by the Australian Bureau of Statistics

SUMMARY OF FINDINGS

The 1996 Qualification Level and Field of Study Paper evaluates the data quality of the qualification questions in the census. The topics analysed in the paper include: the most frequent errors made by respondents (including non-response rates and answering with incorrect or insufficient information), processing issues (including coding instructions, the edits invoked and the most frequent coding errors) and the proposed changes to questions and classification for the 2001 Census.

The main conclusions of the analyses are as follows:

- High non-response rates are a serious issue for qualification variables. Further testing needs to investigate the potential of reducing the non-response rate through improved question design.

- It was often difficult to code respondents' Qualification Field to the three-digit detailed level. For 'Education', 31.6% of respondents could not be coded to the detailed field because they provided insufficient detail in their responses.

- Coding discrepancy analyses for Qualification Level showed that coders had most difficulty coding Qualification Level for 'Skilled Vocational Qualifications' or 'Basic Vocational Qualifications'.

- Further analyses of coding discrepancies for Qualification Field showed that misallocations frequently featured the broad fields 'Business and Administration' and 'Society and Culture'.

- For the 2001 Census there have been important changes made to qualification questions, and a new classification system is to be implemented.

CONTENTS

LIST OF TABLES

1. INTRODUCTION

A question relating to education, in which respondents reported their highest level of achievement, was included in the 1911 Census. However, it was not until 1976 that a question was included in the census which directly asked respondents to provide details of the name of their highest qualification and the institution at which it was obtained. In the 1996 Census, qualification data were obtained which were used in planning and policy development in education, training and employment. These data were also used to assist in evaluating the qualifications, skill and knowledge level of the labour force, and were used by the Department of Immigration and Multicultural Affairs in guidelines for recruiting skilled migrants. Qualification data can also reflect educational advantage in different socio-economic groups and are used in the calculation of the Socio-Economic Indicator for Areas (SEIFA).

1.1 *Qualification Questions in 1996*

The aim of this working paper is to evaluate the quality of data relating to qualification collected in the 1996 Census.

The Australian Bureau of Statistics Classification of Qualifications (ABSCQ) defines a post-school qualification as an award for attainment as a result of formal learning from an accredited post-school institution. This definition was not included on the census form, although it was specified on the form that the qualification must have been completed since leaving school. There is therefore an element of discretion required by respondents in determining the relevance of their qualification.

In 1996, respondents answered five questions relating to qualification which were coded to three main variables. Question 23, Qualification Indicator, was a tick-box question which asked whether the respondent had completed a trade certificate or other educational qualification since leaving school. If respondents answered 'No' or 'No, still studying for first qualification', sequencing instructions directed them to skip the remaining qualification questions. If respondents answered 'Yes, trade certificate/ apprenticeship' or 'Yes, other qualification' they were expected to answer the subsequent questions for qualification.

Write-in responses were required for the full name and field of the highest completed qualification in Questions 24 and 25 respectively. These responses were used to code a level and field for each qualification. Question 26 asked for the institution at which the respondent's highest qualification was completed, although responses for this question were used only to help code Qualification Level or Field and were not themselves coded. Finally, Question 27 was a tick-box question asking the year of completion of the highest qualification. The complete wording and sequencing of the 1996 Census questions relating to qualification can be seen in Appendix 1.

1.2 *Changes to Qualification Questions From 1991*

Changes to form design and question wording were made from the 1991 Census. Most notably, questions in 1996 asked about the highest qualification *completed* rather than the highest qualification *obtained* (1991 wording). After the 1991 Census it was thought that some respondents had answered the qualification questions for courses in which they were enrolled and were participating but which they had not completed. This

change in wording saw a decrease of 305,294 (6.0%) from 1991 to 1996 in the number of respondents who reported that they held a qualification.

There were also changes made to the examples for the 'full name of qualification' question. In 1991 the examples were 'registered nursing certificate, bricklaying trade certificate'. However in 1996 the examples were 'trade certificate, bachelor degree, associate diploma, doctorate'. The wider variety of examples and the inclusion of commonly obtained university degrees saw the non-response rate for Qualification Level decrease from 15.3% in 1991 to 10.9% in 1996. The absence of any university qualification examples in 1991 may have led university-qualified persons to believe that they were not required to answer this question or to be unsure about the required response.

Finally, the Qualification Year question (sequentially the last qualification question) was printed at the top of a new page on the 1991 Census form, but on the same page in 1996. The placement of this question on the following page in 1991 may have led some respondents to overlook it, evidenced by a lower non-response rate in 1996 (4.4% compared to 5.2% in 1991).

1.3   Quality Issues in Qualification Data

Qualification data rely heavily on the ability of respondents to provide the correct information, so are subject to the usual quality constraints imposed by a self-enumerated questionnaire. The first issue discussed in this working paper concerns respondents' errors, such as failure to respond to questions (non-responses) and answering with incorrect or incomplete information.

The second issue involves matters of processing, such as the coding strategies used and the edits invoked. This section also includes an analysis of the accuracy of coders and the most frequently made errors and miscodings.

The final point involves a discussion of the changes to the wording of qualification questions and changes to the classification scope for the 2001 Census.

1.4   List of Acronyms Used in this Paper

SEIFA - Socio-Economic Indicator for Areas

ABSCQ - Australian Bureau of Statistics Classification of Qualifications

NFD - Not Further Defined

CAC - Computer Assisted Coding

QM - Quality Management

ICR - Intelligent Character Recognition

AC - Automatic Coding

QR - Query Resolution

TEW - Transition from Education to Work

AQF - Australian Qualification Framework

ASCED - Australian Standard Classification of Education

2.   RESPONDENT ERRORS

The 1996 Census of Population and Housing form was a self-enumerated questionnaire completed by respondents with little or no assistance from the census collector. Data therefore relied heavily on the ability of respondents to understand each question and to answer in the appropriate manner with the appropriate amount of detail. In a questionnaire of this type, there was no opportunity to probe respondents for more information or to clarify a response.

2.1   Non-Response Rates in 1996

The high non-response rate for qualification questions was the most serious issue relating to respondent error. Non-response rates for the four qualification variables, in particular Qualification Level, were some of the highest of all census questions. Table 1 contains the 1996 non-response rates for qualification variables as calculated on the 1996 Census Fact Sheets for Australia.

TABLE 1: NON-RESPONSE RATES FOR QUALIFICATION QUESTIONS, 1996 CENSUS

Qualification Variable	Persons for whom Questions were relevant	Persons from whom there was no response	Non-Response Rate (%)
Qualification Indicator	13,914,897	1,085,713	7.8
Qualification Level	4,749,063	515,525	10.9
Qualification Field	4,749,063	185,494	3.9
Qualification Year	4,749,063	210,449	4.4

Table 1 shows as many as 1 in 10 respondents who should have provided Qualification Level information were coded as 'Not Stated'. Qualification non-response rates compare unfavourably with other variables on the census form, like the tick-box 'Method of Transport to Work', which had a non-response rate of just 1.8% in 1996. The Occupation question, which required a write-in response, had a non-response rate of 1.7%.

A critical issue in calculating non-response rates is the definition of a 'non-response' and the determination of respondents for whom the question was 'relevant'. Qualification Indicator was an applicable question for all respondents over the age of 15. Data for Qualification Level, Field and Year were coded for the same group of respondents *except those who indicated that they did not have a qualification*. However, in non-response analyses (Table 1), level, field and year were deemed to be 'relevant' only if the respondents were over the age of 15, and if they responded to Qualification Indicator that they had completed a qualification. Hence if a respondent was coded as 'not stated' to Qualification Indicator, that respondent was excluded from non-response analyses.

The decision to remove respondents from non-response analyses if they had not answered Qualification Indicator was made in 1991. This strategy intended to exclude people who did not answer any qualification questions (referred to in this paper as 'topic non-respondents') because the majority of topic non-respondents were assumed not to have any post-school qualification. On reaching the qualification questions, these respondents may have thought that none of the questions were relevant and failed to answer the Qualification Indicator question. By not answering this question, they failed to show that they did not have a qualification. Topic non-respondents were excluded from

analyses so that non-response rates could better reflect the proportion of respondents who failed to answer questions when expected to do so.

This method of analysis arguably underestimates the true non-response rate. A notable number of respondents (953,192) were coded as not stated to all four qualification questions. The issue of whether 'topic non-respondents' should be included in analyses of qualification data is worthy of closer attention. Table 2 illustrates the relatively larger non-response rate for qualification questions if not only respondents who answered 'Yes' to Qualification Indicator were included, but also those who did not provide a response to Qualification Indicator (and therefore *may* have completed a qualification).

TABLE 2: NON-RESPONSE RATES FOR QUALIFICATION QUESTIONS (INCLUDING RESPONDENTS WHO FAILED TO ANSWER QUALIFICATION INDICATOR), 1996 CENSUS

Qualification Variable	Persons for whom questions could have been relevant	Persons for whom there was no response	Adjusted Non-response Rate (%)	Non-Response Rate according to Fact Sheet (%)
Qualification Indicator	13,914,897	1,085,713	NA	7.8
Qualification Level	5,834,776	1,530,815	26.2	10.9
Qualification Field	5,834,776	1,173,579	20.1	3.9
Qualification Year	5,834,776	1,175,027	20.1	4.4

NA = Not Applicable

Table 2 reveals a large increase in non-response rates if persons for whom the question *could* have been relevant are included. More than one in four respondents who may have held a qualification failed to provide codeable information relating to level of attainment. Discussion of the applicability of topic non-respondents is therefore needed.

2.2 Characteristics of Topic Non-Respondents

A large number of respondents coded as topic non-respondents in 1996 were on dummy forms completed due to an inability to contact a person or a refusal by a person to complete the form. In such instances, responses for age, marital status and usual residence were imputed and the remaining census questions were coded as Not Stated (or Not Applicable, depending on the values of the imputed variables). Therefore, the number of actual topic non-respondents is less than is implied by the raw figures. Of the 953,192 topic non-respondents, 190,758 (20.0%) were on dummy forms, while the remaining 762,434 (80.0%) were genuine non-respondents. Since dummy forms do not reflect a mistake by a person in responding to census questions, they have been removed from the following analyses.

2.2.1 Analysis of Topic Non-Respondents

The exclusion of respondents who had not answered Qualification Indicator from Qualification Level, Field and Year non-response analyses was justified by the belief that the majority of these people did not have a post-school qualification. It therefore becomes pertinent to analyse the characteristics of these respondents. Table 3 shows the pattern of responses to the three other qualification variables as a function of response to Qualification Indicator.

TABLE 3: FREQUENCY OF 'STATED' RESPONSES FOR QUALIFICATION FIELD, LEVEL AND YEAR BY RESPONSE TO QUALIFICATION INDICATOR, 1996 CENSUS

| Number of 'Stated' Responses to Other Qualification Variables | Response to Qualification Indicator[1] | | | |
	Yes	% of 'Yes' Respondents	Not Stated	% of 'Not Stated' Respondents
0	95,549	2.0	762,434	85.2
1	80,834	1.7	39,849	4.5
2	463,153	9.8	28,679	3.2
3	4,109,527	86.5	63,993	7.2
Total[2]	**4,749,063**	**100.0**	**894,955**	**100.0**

[1] All respondents (8,080,121) who answered 'no' to Qualification Indicator were coded as 'Not Applicable' to the remaining qualification questions, so have been excluded from this table.

[2] Some totals do not add up due to rounding.

Table 3 shows that 85.2% of respondents who failed to answer the Qualification Indicator question did not answer any of the other Qualification questions either. Only 7.2% of those who did not answer Qualification Indicator answered all three remaining qualification questions. These data show that if a respondent failed to answer Qualification Indicator then that respondent was also highly unlikely to provide a response to any of the other qualification questions.

It is also possible to cross tabulate qualification topic non-respondents with other census variables to determine which members of the population are failing to complete the qualification questions. For example, Question 4, Age:

TABLE 4: TOPIC NON-RESPONSE FOR QUALIFICATION QUESTIONS BY AGE GROUP, 1996 CENSUS

Age[1]	Number of Topic Non-Respondents	% of Topic Non-Respondents	% of Age Level
15	49,626	6.5	19.5
16	34,975	4.6	14.0
17	26,589	3.5	10.7
18	12,125	1.6	4.9
19	8,547	1.1	3.4
20-29	73,164	9.6	2.7
30-44	119,163	15.6	2.9
45-59	105,852	13.9	3.5
60-74	166,354	21.8	8.6
75-89	148,369	19.5	18.0
90+	17,670	2.3	29.0
Total	**762,434**	**100.0**	

[1] Qualification questions are only applicable to respondents aged 15 years or more

The above table shows that the topic non-response rate decreases as the likelihood of a population holding a qualification increases. It can be seen that 43.6% of topic non-respondents were over the age of 60. The frequency of post-secondary qualifications in this age group would be fewer than in younger age groups. The proportion of this age

5

group topic non-responding was also relatively high. Respondents between the ages of 15 and 20 formed 17.3% of the total topic non-respondents. This group would be less likely to hold a post-school qualification due to the length of time needed to complete a course (although not impossible, since many courses are as few as 12 months in length). Furthermore, a large number of respondents in this age group would still have been attending school and may have been confused about how to respond.

Analysis of Question 22, Age Left School also suggests that most topic non-respondents do not have a formal qualification. Table 5 shows the distribution of topic non-respondents as a function of Age left School. For this variable, too, the likelihood of a group holding formal qualifications is inversely proportional to the non-response rate.

TABLE 5: TOPIC NON-RESPONSE FOR QUALIFICATION QUESTIONS BY AGE LEFT SCHOOL, 1996 CENSUS

Age Left School	Number of Respondents	% of Topic Non-Response	% of 'Age Left School' Level
Still at School	82,536	10.8	12.6
Never Attended School	4,224	0.6	4.2
14 years and under	55,227	7.2	2.9
15 years	45,016	5.9	1.6
16 years	34,016	4.5	1.3
17 years	23,097	3.0	0.9
18 years	15,336	2.0	0.9
19 years and over	8,144	1.1	1.6
Not Stated	494,838	64.9	NA
Total	**762,434**	**100.0**	

NA = Not Applicable

Table 5 shows that 11.4% of topic non-respondents were either still at school or had never attended school. The majority of these respondents can be assumed not to have a post-school qualification due to their reduced likelihood of participation in tertiary study. Similarly, 7.2% of topic non-respondents left school at the age of 14 or under. These respondents were also unlikely to hold a post-school qualification.

2.2.2   *Conclusions About Non-Response Rates*

Analyses of topic non-respondents support the hypothesis that most topic non-respondents did not hold a qualification and that these respondents failed to respond to Qualification Indicator. Firstly, 85.2% of respondents who did not answer Qualification Indicator also failed to answer any other qualification question. Secondly, analysis of the variables 'Age' and 'Age Left School' showed that the less likely a respondent was to hold a qualification, the more likely they were to topic non-respond. Specifically, the young (aged 15-20), the old (aged over 60), those still at school, those who had never attended school or those who left school at 14 years of age or under were the most likely to topic non-respond.

However, the above analyses do not intend to imply that none of these topic non-respondents held a qualification. Moreover, 14.8% of respondents failed to answer Qualification Indicator but still supplied an answer to at least one other qualification

question. There is concern that Qualification Indicator is particularly confusing to respondents. The wording of Qualification Indicator ('has the person completed a trade certificate or any other qualification since leaving school?') may be interpreted as inquiring primarily about trade certificates. For example, it may not be immediately obvious that the question is intended to include university degrees, resulting in a number of false-negative (incorrect 'no' responses) answers. The likelihood of such a misinterpretation was increased because the words 'has the person completed a trade certificate' were on the first line of the question. Respondents failing to scan the second line of the question would overlook the words 'or any other educational qualification'. In 2001 the first line of the question is: 'has the person completed a trade certificate or any' with the words 'or any' prompting respondents to read the second line of the question.

2.3   'Not Further Defined' Codings for Qualification Field

The principles of coding to Australian Bureau of Statistics Classification of Qualifications (ABSCQ) required responses given on the census form to be coded to the most detailed level of the classification possible (see Appendix 2 for an example of the structure of the ABSCQ). If a response was not detailed enough to allow coding to the 3-digit level, an 'NFD' (not further defined) code was allocated. The coding was structured as follows:

- the Detailed Field, called the 3-digit level (for example Personnel Management is 113);
- the 'NFD' category of the Narrow Field, called the 2-digit level (for example Management NFD is 110);
- the 'NFD' category of the Broad Field, called the 1-digit level (for example Business and Administration NFD is 100); or
- the inadequately described category.

NFD coding, also known as dump coding, mainly occurs when the level of information provided on the census form is not detailed enough. As discussed, respondents might overlook some questions or provide a response which does not contain sufficient information. Responses may also be assigned a NFD code due to a coder not following correct procedures or failing to use all information on the forms. The following table shows the distribution of NFD (dump) coding during 1996 Census processing.

TABLE 6: DISTRIBUTION OF NOT FURTHER DEFINED RESPONSES IN 1996 CENSUS

ABSCQ	% of responses coded to Broad Field (1-digit code)	% of responses coded to Narrow Field (2-digit code)	% of responses coded to Detailed Field (3-digit code)	Total
Business & Administration	11.6	8.2	80.2	833,190
Health	1.1	2.4	96.5	535,391
Education	5.3	26.3	68.4	460,638
Society & Culture	5.6	7.0	87.4	573,019
Natural & Physical Sciences	12.9	0.5	86.6	274,144
Engineering	11.1	13.8	75.1	1,155,637
Architecture & Building	0.3	17.2	82.5	365,538
Agriculture & Related Fields	0.3	1.3	98.4	103,972
Miscellaneous Fields	0.1	1.3	98.6	304,440
Inadequately Desc.	NA	NA	NA	55,228
Not Stated	NA	NA	NA	1,173,579

NA = Not Applicable

Table 6 shows that within the broad field 'Education' only 68.4% of responses were coded to the 3-digit detailed field, the lowest percentage in the table. This is largely due to the great number of responses (120,988) dump coded at the 2-digit level as 'School Teacher Training NFD'. This was the 2-digit dump code to which the greatest number of responses were coded and would be used when a respondent indicated his/her qualification was in school teaching, but failed to provide more specific information. Thus trained teachers are frequently failing to specify the type of teaching in which they are trained, despite the example 'primary school teaching' accompanying the question on the census form. More detailed instructions for school teachers on the census form, or in the census guide, may improve the quality of these responses.

The second lowest percentage of 3-digit level coding took place for 'Engineering', for which only 75.1% of responses were coded to a detailed field. 13.8% of responses were dump coded to the 2-digit level. A great proportion of these responses were coded to the narrow field 'Electrical and Electronic Engineering NFD'. The 11.1% who were dump coded to the 1-digit level were those respondents who answered simply 'engineering', or who used a similarly broad term like 'drafting'. Since 'Engineering' was the single largest group in the classification the use of an example like 'Mechanical Engineering' on the census form may be useful.

Coding to the 3-digit level for 'Business and Administration' took place for just 80.2% of responses. 11.6% of responses were dump coded to the 1-digit level and 8.2% to the 2-digit level. The high percentages of 1-digit NFD coding can be attributed to responses of merely 'Business'. Dump coding to 'Management NFD' took place on 37,753 occasions and formed the majority of dump coding for Business and Administration at the 2-digit level. While this indicates that some respondents may not be providing sufficient information, these dump codes do not necessarily imply an incomplete answer from a respondent. A 'Bachelor of Business' or a 'Diploma of Management' may not have a specific type of business or management associated with them and the qualification itself may only be codeable at the 2-digit level.

Respondents with an 'Architecture and Building' qualification were coded to the 3-digit level on 82.5% of occasions. Once again, a large percentage (17.2%) were dump coded at the 2-digit level. The majority of this dump coding was for 'Building Construction NFD', to which 42,630 responses were assigned.

2.4 'Inadequately Described' and 'Not Stated' Responses to Qualification Level

In addition to the 1,530,815 respondents (including topic non-respondents) who failed to provide a response to Qualification Level, another 124,812 respondents provided a level that could not be fully coded and were classified as 'Inadequately Described'. Of these 1,655,627 respondents who could not provide a codeable qualification level, 513,816 (31.0%) provided a response to Qualification Field that was suitably coded. It is worthwhile to consider why more than half a million people were able to provide Qualification Field information, but could not provide codeable Qualification Level data.

There has been concern that some respondents may be reporting qualifications which are not of sufficient Qualification Level to be classified by the ABSCQ, and are therefore out of the scope of the qualification questions. As stated earlier, the census form did not provide a definition of what levels of qualification were in-scope. Table 7 cross tabulates respondents who failed to provide enough qualification level information by their response to qualification field.

TABLE 7: FREQUENCY OF 'UNDEFINED' QUALIFICATION LEVEL RESPONSES BY QUALIFICATION FIELD, 1996 CENSUS

Qualification Field[1]	Response to Qualification Level			
	Inadequately Described	Not Stated	Total Undefined	% Undefined
Business & Administration	46,335	161,968	208,303	25.0
Health	28,274	42,192	70,466	13.2
Education	4,678	21,316	25,994	5.6
Society & Culture	14,240	39,747	53,987	9.4
Natural & Physical Sciences	3,400	19,051	22,451	8.2
Engineering	8,115	53,138	61,253	5.3
Architecture & Building	4,227	14,774	19,001	5.2
Agriculture & Related Fields	1,402	10,620	12,022	11.6
Miscellaneous Fields	10,176	30,163	40,339	13.3
Total	**120,847**	**392,969**	**513,816**	**11.2**

[1] Not Stated and Inadequately Described responses to Qualification Field have been removed

The table shows that 25.0% of responses coded to 'Business and Administration' did not provide a response which could be coded to a level category. A high percentage came from the detailed field 122 'Keyboard and Shorthand' (101,588 responses, or 52.9% of all 'Keyboard and Shorthand' responses). The frequency of 'level undefined' responses and the nature of this field may indicate that a proportion of these respondents did not complete a post-school qualification, but completed a short-term introductory course to typing or shorthand. A smaller number of 'Business and Administration' respondents who failed to define a qualification level came from the detailed field 'Accounting' (24,437 respondents or 13.0% of all 'accounting' responses). A possible explanation is that some

of these respondents, may have completed a basic bookkeeping course or a brief course in using a particular accountancy software. However, this cannot be stated with certainty.

The Qualification Level for 'Miscellaneous Fields' was undefined on 13.3% of occasions. The detailed (3-digit level) fields contained within this broad field were of the type that might be held as a brief introductory course, rather than a formal qualification. For example, although 'beauty-therapy' and 'waiting and bar services' can constitute a proper post-school qualification, they can also be completed as a basic introductory course, which does not qualify as a vocational qualification.

The broad field of 'Health' also had a moderately high number of respondents whose Qualification Level could not be classified. The detailed field to which the majority of these respondents were coded was 'Basic Nursing' (38,104 or 14.2% of all 'Nursing' respondents). One hypothesis might be that respondents who completed a first aid course would describe their qualification as nursing or basic health care.

There should also be some discussion of the large number of 'level undefined' responses (11,229 respondents) to 'Computer Science', detailed field 541. This figure represents 12.4% of all respondents who described their qualification field as computer science. Respondents who incorrectly reported basic computer courses (e.g. word processors or spreadsheets) would be likely to be coded to this field.

From the above data there seems to be some evidence that respondents may be reporting qualifications which do not lie within the ABSCQ definition of a post-school qualification. Such incidences are difficult to avoid due to the self-enumerated nature of the census. However the extent of this misreporting is not precisely quantifiable. Many of these respondents who failed to describe a qualification level and who were included in the above analyses might hold formal qualifications. Similarly, many respondents may have reported a qualification that falls beyond the scope of the ABSCQ as a 'certificate' and have been coded normally, along with applicable qualifications.

Some respondents who reported a qualification field but not a level may also have been confused because Question 24 asked for the 'Full name of qualification' and not for 'Qualification Level' (although it was implied by the example responses). In the 2001 Census, this question will be changed to specifically ask for 'Level of qualification'.

Similarly, it has been noted that a number of respondents answer qualification questions with the details of their occupation, assuming that this provides some detail of their qualification. For example, a hairdresser may simply describe their qualification as 'hairdresser'. This, too, could explain the large number of people for whom a Qualification Field was successfully coded, but who did not provide an adequate Qualification Level.

3. PROCESSING ISSUES

Tick-box responses to Qualification Indicator and Qualification Year were coded through Optical Mark Recognition, while the write-in responses to Qualification Level and Field were processed by coders using Computer Assisted Coding (CAC). The following discussion of processing procedures concentrates on the coding of Qualification Level and Field information, given the relatively greater complexity of processing write-in responses.

3.1   *Coding of Qualification Level and Qualification Field Responses*

Qualification Level responses indicate how advanced a qualification was. Qualification Field responses describe the content of the qualification. In 1996, coders were not restricted to information contained in the appropriate question to code these two variables. For example, if a respondent's answer to the 'Full name of qualification' question was 'Bachelor of Business' but that respondent provided no answer to the Qualification Field question, then 'business' could be used to code the field of study. Similarly, if the respondent had answered at the Qualification Indicator that they completed a trade certificate or apprenticeship, this information could be used in coding Qualification Level. Question 26, which asks for the institution at which the highest qualification was completed, was included on the census form specifically to facilitate the coding of Level and Field variables. For example, if a respondent described his/her level of attainment as 'diploma' that person's response to the institution question could determine whether this was coded as 'undergraduate diploma', 'associate diploma' or 'post-graduate diploma'.

Coding of Field of Study and Level of Attainment took place using CAC. Coders would begin by entering the 'basic word' of a stated qualification. This basic word was the word that best answered the question: 'what is the qualification about?' Some examples of basic words were: pharmacy, engineering, management, science or hairdressing. Coders also entered any qualifying words that the respondent provided. A qualifying word was a word that added meaning to the basic word: for example, if the Qualification Field response was 'Civil Engineer', 'Engineer' was the basic word and 'Civil' the qualifying word.

Coders were provided with a basic word heirarchy to determine which word in a response was the basic word, and which the qualifying word. For example, if a respondent describes the Qualification Field as 'Nursing Aide' then 'Aide' is used as the basic word and 'Nursing' the qualifying word because 'Aide' is higher in the basic word heirarchy than 'Nursing'.

After entering information about the field of qualification, coders were prompted to select an appropriate field from a number of similar entries. After selecting the relevant field entry, coders were prompted to select from a number of applicable levels of attainment. As a result of these coding procedures a single digit number was assigned to each response for Qualification Level, and a 3-digit number assigned for Qualification Field.

3.2   *Edits Used in Processing of Qualification Data in 1996*

At times during processing an 'edit' could be invoked which would systematically provide a code for one variable based on an answer to another variable. The most straightforward example of an edit would be if a respondent answered 'no' to Qualification Indicator, then Qualification Level, Field and Year were systematically coded to 'not applicable'.

Edits are invoked for three main reasons:

- to remove inconsistencies within a respondent's answers. For example, a respondent cannot logically indicate that they do not have a qualification and also answer that they completed a qualification in 1993-4;
- to balance categories when data are aggregated;
- to maintain consistency between data and the Australian Bureau of Statistics Classification of Qualifications (ABSCQ) - for example, the classification does not classify persons under 15 years of age; and
- to save time and money during census processing by removing coding that is not necessary.

There were three main types of edits invoked for qualification questions which may have implications for qualification data quality.

Firstly, a number of edits were invoked to code any respondents under the age of 15 as 'Not Applicable' to all four qualification variables. A file was retained which captured all information written on the census form (except name and address) for 2% of all respondents. Examination of this file suggests that a large number of persons younger than 15 answered the Qualification Indicator question (12,273 or 3.9%). However, 12,222 (or 99.6%) of these respondents reported that they did not have a qualification. Qualification Year, the last of the qualification questions, was answered by just 114 respondents who were aged less than 15. This edit did not, therefore, have a negative effect on overall data quality and was valuable in maintaining the consistency of the data with the classification - the (ABSCQ) is not intended to classify respondents under the age of 15.

Secondly, a number of edits were invoked to code respondents who answered that they did not have a qualification (to Qualification Indicator) as 'Not Applicable' to the remaining qualification questions. This edit reinforces the sequencing of questions and ensures consistency within responses. If a respondent answers that they do not have a qualification, they cannot logically hold (for example) a Bachelor Degree. Conceivably this edit could result in the loss of information if a respondent mistakenly marked Qualification Indicator as 'no' but then provided details of a qualification. For example, a respondent with a bachelor degree may have interpreted Qualification Indicator as 'do you have a trade certificate?' This respondent would then respond 'no' but would complete the level, field and year of their degree. This information would then be lost, although this was likely occur infrequently. It has also been thought that many of the respondents who answered 'no' but then provided details of a qualification may be providing details of a qualification in which they were currently enrolled, or which they had only partially completed. The edit would remove the details of these qualifications.

The final edit of interest balanced the respondent's provided age with Qualification Year. This edit was based on a minimum age of 15 to have a qualification. If a 25 year old

respondent replied to Qualification Year that they completed their degree before 1986 (i.e. more than ten years previously, when they were less than 15) then their response to Qualification Year would be recoded as 'Not Stated'. Analysis of the file showed 451 responses invoked this edit (0.5%). The majority (255) of invocations of this edit involved respondents between the ages of 31-40 who answered that they completed their qualification before 1971. Again this confirms consistency between collected data and the ABSCQ, in which respondents under 15 years of age cannot hold a post-school qualification.

3.3   Detection of Discrepancies

A Quality Management (QM) system was established to identify systematic discrepancies in processing, to provide feedback to coders on discrepancies and to produce and analyse discrepancy rates by topic.

During the processing of the 1996 Census data, a sample of each coder's work on Collection Districts (the smallest census unit for collection, processing and output of data) was selected for reprocessing by another coder and any mismatches were looked at by an adjudicator who would decide on the correct code. If the adjudicator disagreed with the initial coder, a discrepancy would be recorded. These discrepancy analyses were performed for a number of different variables, including both Qualification Level and Qualification Field. There were 5,834,776 applicable census counts from which 382,888 Qualification Field and Level responses (6.6%) were recoded by QM coders. Altogether 20,526 discrepancies were recorded for Qualification Field (5.4% of all responses) and 15,873 discrepancies were recorded for Qualification Level (4.1% of all responses).

The QM system in place during processing allowed the detection of discrepancies and the calculation of a crude discrepancy rate. This crude discrepancy rate differs from a true discrepancy rate for the following reasons:

- a higher proportion of 'poor' coders' work was included in the quality monitoring sample;
- the QM check coders could make the same mistake as the original coder and therefore an error would not be detected; and
- there is not always an absolutely correct code for each response.

Note that there are likely to be sustantial changes to the QM system in 2001 due to the use of Intelligent Character Recognition (ICR) and Automatic Coding (AC) technology.

3.4   Discrepancy Analyses

3.4.1   General Information

When a coder and a QM coder reached different codes for a qualification response an adjudicator would decide on the correct code and a discrepancy would be recorded whenever the initial coder and the adjudicator disagreed. These discrepancy reports were used to set qualification discrepancy rates for coders.

Discrepancy profile tables could also be used to examine which codes had been determined by the adjudicator and which codes had been incorrectly allocated by the system through the coders' work. Unlike the discrepancy reports these tables recorded

discrepancies made by the initial coder as well as the QM coder so that two discrepancies could be recorded for one qualification response if the adjudicator disagreed with both the initial coder and the QM coder. These tables have been used for the following analyses of discrepancies as they present more detailed information.

The following section presents tables showing the highest frequencies of discrepancies for Qualification Level and for the one-digit level of Qualification Field. Analyses for Field take place at the one digit level only, as these represent the most serious miscodings that could be made. For example, coding a 'Health' qualification (with broad field 2) as an 'Engineering' qualification (broad field 6) is a relatively more serious mistake than coding a 'Hairdressing' qualification (detailed field 911) as 'Beauty Therapy' (detailed field 912).

In order to determine which, among the Australian Bureau of Statistics Classification of Qualification (ABSCQ) groups, were more prone to coding discrepancies, a normalised crude discrepancy ratio has been calculated for both tables. First the frequency of discrepancies for each group in the tables has been divided by the total number of persons reporting that level of attainment or field of study. Then the group with the smallest proportion of discrepancies was used as a normaliser which by definition has the value of 1.0. The use of this normaliser was due to incomplete records of the QM recodings. Data were not available for the number of responses to each level or broad field that was recoded, therefore a direct percentage of discrepancies could not be calculated.

3.4.2 Qualification Level Discrepancies

The discrepancy profile table for Qualification Level contained 50,007 discrepancies where the adjudicator disagreed with either the initial coder or the QM coder. These discrepancies include 17,029 queries (34.1%) in which coders had incorrectly raised a query and which were resolved by Query Resolution (QR) staff. Since these queries were ultimately resolved and had no effect on the quality of qualification data, they have been removed from the total number of discrepancies. Table 8 illustrates which Qualification Levels were incorrectly allocated most frequently as a result of coders' selections.

TABLE 8: CODING DISCREPANCIES FOR QUALIFICATION LEVEL IN ORDER OF NORMALISED DISCREPANCY RATIO, 1996 CENSUS

	Correct Qualification Level					*Incorrectly allocated to:*	
Level & ABSCQ code	*Frequency in population*	*% of all quals*	*Frequency of discrepancies within code*	*% of total discrepancies (32,978)*	*Normalised discrepancy ratio[1]*	*Level & ABSCQ code*	*%*
Postgraduate Diploma (2)	183,087	3.1	1,380	4.2	3.6		
						Bachelor Degree (3)	55.7
						Undergraduate Diploma (4)	29.9
						Higher Deg. (1)	7.0
Basic Vocation. (7)	398,744	6.8	2,740	8.3	3.2		
						Skilled Vocational (6)	27.0
						Undergraduate Diploma (4)	26.3
						Assoc. Dip. (5)	7.5
Assoc. Diploma (5)	359,701	6.2	1,588	4.8	2.1		
						Skilled Vocational (6)	32.1
						Basic Vocational (7)	22.3
						Undergraduate Diploma (4)	16.4
Undergraduate Diploma (4)	486,843	8.3	2,018	6.1	2.0		
						Basic Vocational (7)	21.7
						Bachelor Degree (3)	13.7
						Postgraduate Diploma (2)	13.2
Skilled Vocational (6)	1,483,000	25.4	5,810	17.6	1.9		
						Basic Vocational (7)	33.9
						Associate Diploma (5)	6.3
						Inadequately Described (8)	2.5
Higher Degree (1)	190,840	3.3	535	1.6	1.3		
						Bachelor Degree (3)	56.3
						Postgraduate Diploma (2)	6.9
						Undergraduate Diploma (4)	6.5

15

Bachelor Degree (3)	1,076,934	18.5	2,281	6.9	1.0		
						Undergraduate Diploma (4)	24.8
						Postgraduate Diploma (2)	11.2
						Higher Deg. (1)	7.8
Inadequately Described (8)	124,812	2.1	618	1.9	NA		
						Undergraduate Diploma (4)	13.3
						Basic Vocational (7)	11.8
						Skilled Vocational (6)	10.2
Not Stated	128,595	2.2	2,801	8.5	NA		
						Skilled Vocational (6)	45.8
						Basic Vocational (7)	12.0
						Bachelor Degree (3)	10.8
A Query should have been raised	NA	NA	12,456	37.1	NA		
						Skilled Vocational (6)	18.4
						Basic Vocational (7)	15.4
						Undergraduate Diploma (4)	8.8

[1] The normalised discrepancy ratio for 'Bachelor Degree'= 2,281/1,076,934* 1,076,934/2,281 = 1.0. Therefore the normalised discrepancy ratio for 'Postgraduate Diploma' is 1,380/183,087 * 1,076,934/2,281= 3.6. NA= Not Applicable.

The qualification level 'Postgraduate Diploma' (2) recorded the highest normalised discrepancy ratio (3.6). These discrepancies were most frequently miscoded (769 times) as 'Bachelor Degree' (3). This constituted 55.7% of the miscodings for 'Postgraduate Diploma'. 29.9% of the discrepancies recorded for 'Postgraduate Diploma' were codes allocated to 'Undergraduate Diploma' (4), while 7.0% were incorrectly coded as 'Higher Degree' (1).

'Basic Vocational Qualification' (7) recorded the second highest discrepancy ratio (3.2). High percentages of discrepancies were coded to 'Skilled Vocational Qualification' (6) (27.0%) and 'Undergraduate Diploma' (4) (26.3%). These constituted 740 and 721 discrepancies respectively. 7.5% of discrepancies were miscodings to 'Associate Diploma' (5).

'Associate Diploma' (5) recorded the third highest discrepancy ratio (2.1). Discrepancies were most frequently due to miscodings as 'Skilled Vocational Qualification' (6) (32.1%), 'Basic Vocational Qualification' (7) (22.3%) and 'Undergraduate Diploma' (4) (16.4%).

Although the Qualification Level 'Skilled Vocational Qualification' had only the fifth highest discrepancy ratio it was the level which contained the largest number of discrepancies (5,810, or 17.6% of all discrepancies). The lower discrepancy ratio was due to the great frequency of this level in the population (1,483,000 respondents). Therefore, although the coding within this level was proportionately better than other levels it was the group that had the greatest single influence on the overall quality of qualification data.

The incorrect allocations of Qualification Levels indicated that coders had difficulties in classifying 'Skilled Vocational Qualifications' and 'Basic Vocational Qualifications'. As can be seen in Table 8, Basic Vocational was one of the three most frequent discrepancies for six out of nine categories, while Skilled Vocational was one of the three most frequent discrepancies for five out of nine. There were 16,421 miscodings involving Basic or Skilled Vocational Qualifications (that is, were coded as Basic or Skilled Vocational and should not have been, or were not coded as Basic or Skilled Vocational and should have been). This represented 49.8% of all discrepancies involving Qualification Level.

Of the university-type Qualification Levels, the most problematic classification was 'Undergraduate Diploma' (4), which was one of the three most frequent discrepancies for seven of the nine categories. There were 5,517 discrepancies involving 'Undergraduate Diploma' (16.7% of the total number of discrepancies).

12,456 queries (37.7%) needed to be raised if the coders had followed the correct procedures. The codes allocated instead were most frequently Qualification Levels (6) 'Skilled Vocational, (7) 'Basic Vocational' and (4) 'Undergraduate Diploma' (18.4, 15.4 and 8.8% of the number of queries respectively).

3.4.3   Qualification Field Broad Field (1-digit) Discrepancies

The most serious level of discrepancies for Qualification Field occurred when a response was coded to an incorrect broad field (i.e. at the one-digit level). As stated earlier, it is a more serious mistake to code 'Health' (broad field 2) as 'Engineering' (broad field 6) than to code Hairdressing (detailed field 911) as 'Beauty Therapy' (detailed field 912). The discrepancy profile table at the broad field level contained 43,122 discrepancies where the adjudicator disagreed with either the initial coder or the QM coder. These discrepancies included 17,475 queries (40.5%) which coders had raised incorrectly and which were resolved by QR staff. Since these queries had no effect on the quality of Qualification data they have been removed from the total number of discrepancies.

Table 9 illustrates which Qualification Fields had been incorrectly allocated at the one-digit level as a result of coders' selections.

TABLE 9: CODING DISCREPANCIES AT ONE-DIGIT LEVEL FOR QUALIFICATION FIELD IN ORDER OF NORMALISED DISCREPANCY RATIO, 1996 CENSUS

	Correct Qualification Field					*Incorrectly allocated to:*	
Field & ABSCQ code	*Frequency in population*	*% of all quals*	*Frequency of discrepancies within code*	*% of total discrepancies (25,647)*	*Normalised discrepancy ratio[1]*	*Level & ABSCQ code*	*%*
Natural & Physical Sciences (5)	274,144	4.7	1290	5.0	2.4		
						Business & Admin. (1)	18.8
						Society & Culture (4)	15.1
						Health (2)	10.5
Society & Culture (4)	573,019	9.8	2624	10.2	2.3		
						Inadeq. Desc. (0)	25.3
						Business & Admin. (1)	20.1
						Education (3)	16.9
Education (3)	460,638	7.9	1812	7.1	2.0		
						Society & Culture (4)	42.1
						Natural & Physical Sciences (5)	13.0
						Inadeq. Desc. (0)	8.9
Agriculture & Related (8)	103,972	1.8	329	1.3	1.6		
						Natural & Physical Sciences (5)	16.7
						Business & Admin. (1)	13.1
						Engineering (6)	9.4
Architecture & Building (7)	365,538	6.3	926	3.6	1.3		
						Engineering (6)	25.6
						Society & Culture (4)	7.6
						Business & Admin. (1)	6.6
Miscellaneous (9)	304,440	5.2	740	2.9	1.2		
						Business & Admin. (1)	16.6
						Engineering (6)	13.0
						Society & Culture (4)	6.1

Business & Admin. (1)	833,190	14.3	1985	7.7	1.2		
						Society & Culture (4)	21.4
						Natural & Physical Sciences (5)	11.7
						Miscellaneous (9)	7.4
Health (2)	535,391	9.2	1126	4.4	1.1		
						Natural & Physical Sciences (5)	27.5
						Society & Culture (4)	18.3
						Business & Admin. (1)	14.5
Engineering (6)	1,155,637	19.8	2293	8.9	1.0		
						Architecture & Building (7)	9.7
						Business & Admin. (1)	9.1
						Society & Culture (4)	8.8
Not Stated	1,173,579	20.1	914	3.6	NA		
						Engineering (6)	18.8
						Business & Admin. (1)	16.6
						Society & Culture (4)	12.8
A Query should have been raised	NA	NA	10313	40.2	NA		
						Business & Admin. (1)	23.3
						Engineering (6)	17.5
						Society & Culture (4)	12.1

[1] The normalised discrepancy ratio for 'Engineering' = 2,293/1,155,637 * 1,155,637/2,293 = 1.0. Therefore the normalised discrepancy ratio for 'Natural & Physical Sciences' is 1,290/274,144 * 1,155,637/2,293 = 2.4. NA= Not Applicable.

The broad field 'Natural and Physical Sciences' (5) recorded the highest normalised discrepancy ratio (2.4). These discrepancies were most frequently miscoded (243 times or 18.8%) as broad field 'Business and Administration' (1). 15.1% of the discrepancies recorded for 'Natural and Physical Sciences' were codes allocated to broad field 'Society and Culture' (4), while 10.5% were incorrectly coded to broad field 'Health' (2).

Broad field 'Society and Culture' (4) recorded the second highest discrepancy ratio (2.3). High percentages of discrepancies were coded to 'Inadequately Described' (0) (663 times, or 25.3%), broad field 'Business and Administration' (1) (527 times, or 20.1%) and broad field 'Education' (3) (443 times, or 16.9%). 'Society and Culture' was also the broad

field that had the single largest number of discrepancies (2624 or 10.2% of all discrepancies) and therefore had the largest influence on the overall quality of qualification data.

Broad field 'Education' (3) recorded the third highest discrepancy ratio (2.0). Discrepancies were due most frequently to miscodings as 'Society and Culture' (4) (42.1%). 'Education' was also mistaken as broad field 'Natural and Physical Sciences' (5) (13.0%) and 'Inadequately Described' (0) (8.9%).

Although the Broad field 'Engineering' had the lowest discrepancy ratio it was the level which contained the second largest number of discrepancies (2,293, or 8.9% of all discrepancies). The lower discrepancy ratio was due to the great frequency of this level in the population (1,155,637 respondents). Therefore, although the coding within this level was proportionately better than other levels it was one of the two groups that had the greatest influence on the overall quality of qualification data.

The incorrect allocations of Qualification Field listed above indicated that coders frequently had difficulties in classifying 'Business and Administration' (1) and 'Society and Culture' (4). As can be seen in the table above, both 'Business and Administration' and 'Society and Culture' were incorrectly allocated for eight out of ten categories. There were 6,167 basic level miscodings involving 'Business and Administration' (that is, were coded as 'Business and Administration' and should not have been, or were not coded as 'Business and Administration' and should have been). This represented 24.0% of the total number of discrepancies for Qualification Field. There were 6,128 discrepancies involving 'Society and Culture', 23.9% of the total discrepancies for Qualification Field.

10,308 queries (40.2% of discrepancies) needed to be raised if the coders had followed correct procedures. The codes allocated instead were within broad fields 'Business and Administration' (1), 'Engineering' (6) and 'Society and Culture' (4) (23.3%, 17.5% and 12.1% respectively).

3.4.4 Comparison of Qualification Discrepancies with Other Census Variables

To evaluate the accuracy of qualification coding, an overall discrepancy rate for Qualification Level and Qualification Field at the one-digit level was calculated. This overall figure was derived by dividing the number of discrepancies in the above discrepancy profile tables by the total number of forms that were recoded (382,888 for qualification variables). The resultant discrepancy rate for Qualification Level was 8.6% (32,978 discrepancies) and for Qualification Field was 6.7% (25,647 discrepancies).

Equivalent figures were calculated for other CAC coded census variables. For the Industry variable this discrepancy rate at the one-digit level was 11.1% (57,723 discrepancies from 517,370 forms), while for the Occupation variable this rate was 13.5% (70,091 discrepancies from 519,772 forms). The lower rates for Qualification Level and (in particular) Qualification Field relative to Industry and Occupation show qualification coding to be of a high standard. Furthermore the lower rate of discrepancies suggests that the quality and detail of responses to qualification questions was high. It seems, therefore, that if respondents provided an answer it was usually of sufficient detail to be accurately coded. However, as stated in the earlier analysis of non-response rates (section 2.1), the greatest problem was the failure of respondents to provide an answer at all.

4. RECONCILIATION OF 1996 CENSUS QUALIFICATION DATA WITH TRANSITION FROM EDUCATION TO WORK SURVEY

4.1 Data Reconciliation Methodology

The purpose of this section is to explain the differences in the collection of Qualification Level and Field of Study data between the Transition from Education to Work (TEW) survey and the census, to outline the steps taken to reconcile these two data collections and to present the findings from this reconciliation. The TEW was run as a supplementary survey to the monthly labour force survey for May 1996.

Although the census and the TEW both collect data on Qualification Level and Field of Study, they are not strictly comparable due to differences in the scope, coverage, timing, measurement of underlying concepts and collection methodology. Factors contributing to differences in estimates include:

- under-enumeration in the census for which census qualification data were not adjusted;
- the use in TEW of population benchmarks derived from incomplete information about population change;
- differing methods of adjustment for non-response rates to the survey or census;
- the personal interview approach using any responsible adult in the household adopted in the survey as opposed to self-enumeration in the census; and
- sampling variability.

To enable reconciliation, the scopes of the 1996 Census and the May 1996 TEW were reduced to a common population. Firstly, data were restricted to respondents between the ages of 15 and 64 to match the scope of the TEW. Secondly, 125,406 visitors to Australia were deducted from census figures because overseas residents in Australia are out of scope of the TEW. Finally, 33,483 defence force personnel were subtracted from census figures because members of the Australian Defence Forces are not included in the TEW.

4.2 Results of Data Reconciliation

Census codings included the additional categories 'not stated' and 'inadequately described' to be used when respondents provided no information, or insufficient information to be coded. 1,098,961 respondents were coded as not stated to Qualification Level, while 806,439 respondents were coded as not stated to Field of Study. 104,065 respondents were coded as inadequately described to Qualification Level, while 48,595 were coded as inadequately described to Field of Study. These respondents have been removed from analyses.

Table 10 presents Qualification Level cross-tabulated by Field of Study for the census, while Table 11 presents these figures for the TEW. Cell figures represent the number of respondents in each category as a percentage of all respondents with a stated qualification. Tables A1 and A2 in Appendix 3 show the raw figures used to derive these proportions.

TABLE 10: DISTRIBUTION OF RESPONSES FOR QUALIFICATION LEVEL BY FIELD OF STUDY, 1996 CENSUS

Field of Study	Higher Degree	Postgrad. Diploma	Bachelor Degree	Undergrad. Diploma	Associate Diploma	Skilled Vocational	Basic Vocational	Total
Business & Administration	0.8	0.6	4.5	1.8	3.1	0.6	4.4	15.8
Health	0.7	0.4	4.1	3.8	0.4	0.2	1.7	11.3
Education	0.5	2.5	4.1	2.8	0.8	0.0	0.0	10.8
Society & Culture	1.2	0.7	7.0	1.1	1.3	0.8	0.8	13.0
Natural & Physical Sciences	0.9	0.3	3.5	0.5	0.6	0.1	0.5	6.4
Engineering	0.5	0.1	2.4	0.7	1.7	19.5	0.8	25.7
Architecture & Building	0.0	0.0	0.5	0.2	0.3	6.9	0.3	8.3
Agriculture & Related Fields	0.1	0.0	0.3	0.3	0.3	0.8	0.5	2.3
Miscellaneous Fields	0.0	0.0	0.0	0.3	0.1	5.2	0.8	6.5
Total	**4.6**	**4.6**	**26.5**	**11.5**	**8.7**	**34.2**	**9.8**	**100.0**

Qualification Level (spanning header)

TABLE 11: DISTRIBUTION OF RESPONSES FOR QUALIFICATION LEVEL BY FIELD OF STUDY, MAY 1996 TRANSITION FROM EDUCATION TO WORK SURVEY

Field of Study	Higher Degree	Postgrad. Diploma	Bachelor Degree	Undergrad. Diploma	Associate Diploma	Skilled Vocational	Basic Vocational	Total
Business & Administration	0.6	0.7	3.5	0.7	3.2	1.9	10.4	21.1
Health	0.5	0.7	3.6	2.2	0.9	0.6	2.5	11.0
Education	0.5	1.9	3.2	2.0	1.5	0.6	0.1	9.9
Society & Culture	1.0	0.9	5.4	0.7	1.5	1.1	0.4	11.0
Natural & Physical Science	0.8	0.4	2.8	0.3	1.0	0.4	0.2	5.8
Engineering	0.3	0.2	2.0	0.4	3.9	15.3	0.4	22.6
Architecture & Building	0.0	0.0	0.5	0.0	0.7	6.7	0.0	8.3
Agriculture & Related Fields	0.0	0.0	0.3	0.0	0.6	1.0	0.4	2.5
Miscellaneous Fields	0.0	0.0	0.0	0.0	1.0	5.8	0.8	7.7
Total	**3.9**	**5.0**	**21.5**	**6.6**	**14.3**	**33.4**	**15.4**	**100.0**

Row totals indicate that there was a similar distribution of qualifications by Field of Study for the 1996 Census and the May 1996 TEW. The most notable differences by Field of Study were for broad fields 'Business and Administration' (21.1% of all qualifications for the TEW, 15.8% of qualifications for the census) and for 'Engineering' (25.7% of all qualifications for the census, 22.6% of all qualifications for the TEW). The differences between the census and TEW for all other basic fields were within two percentage points.

Column totals indicate a number of inconsistencies for Qualification Level. Census percentages exceeded those of the TEW for 'Bachelor Degree' (26.5% for census, 21.5% for TEW) and 'Undergraduate Diploma' (11.5% for census, 6.6% for TEW). TEW percentages exceeded those of the census for 'Associate Diploma' (14.3% for TEW, 8.7% for census) and 'Basic Vocational' (15.4% for the TEW, 9.8% for the census). Other qualification levels showed approximately equivalent percentage distributions.

Within cross-categories 'Qualification Level by Field of Study', differences in percentages were highest for Basic Vocational Qualifications in 'Business and Administration' (10.4% for TEW, 4.4% for census) and for Skilled Vocational Qualifications in 'Engineering' (15.3% in the TEW, 19.5% for the census). Other important differences were visible for Associate Diplomas in 'Engineering' (3.9% in TEW, 1.7% in census) and Bachelor Degrees in 'Society and Culture' (7.0% in census, 5.4% in TEW).

These differences between the census and the TEW are likely to reflect the interviewer-based approach of the TEW. In an interview situation it is possible to probe for more information and to clarify responses. When completing the self-enumerated census respondents may not consider their qualification to be relevant to the question (particularly if it is a lower level qualification like a Basic Vocational Qualification). In some respects these problems in qualification data for the census are unlikely to be overcome, because an interviewer-based collection on such a large scale is impractical. However further instructions on the census form or in the census guide which clarify the definition of a 'trade certificate or any other educational qualification' and specify minimum criteria may significantly improve the quality of data.

5. CHANGES FOR 2001

A number of important changes have been made for the 2001 Census in form design and in the index used to code qualification responses. The following section discusses the most salient issues.

5.1 Changes in Form Design

As in 1996, five questions will be asked in 2001 pertaining to qualification and these will be coded to three main variables: Qualification Level, Qualification Field and Qualification Year. One of the most significant changes to questions is the use of Intelligent Character Recognition (ICR) boxes for all write-in responses. The use of this technology is dependent on respondents understanding that they must write in clear, unambiguous block letters within the boxes provided. This change will allow approximately 50% of responses to be automatically coded, while the remaining responses will be coded using Computer Assisted Coding (CAC).

In 2001, the first question will again be the Qualification Indicator question, which sequences respondents to answer qualification questions if they completed a qualification, or to ignore these questions if they have not completed a qualification. For the first time, qualifications completed while the respondent was still at school are in scope. Therefore, respondents who undertook some form of vocational training while in high school will be expected to answer qualification questions.

The second question will specifically ask 'what is the level of the highest qualification the person has completed?' In 1996 this question asked for the 'Full name of qualification' but did not specifically ask for a qualification 'level' (see Appendix 1). The specific reference to 'level' is hoped to improve the quality of data for this variable. The number and range of example responses has also been increased: certificate 2 and advanced diploma have now been included and doctorate excluded.

As in 1996, the third question in 2001, will ask for the main field of study. A change will be made to the question, with beauty salon practice, civil works and hospitality management added to the range of example responses. The fourth question will again ask for the institution at which this qualification was completed.

The final qualification question will ask for the year in which respondents completed their highest qualification. In 1996, this question was answered by choosing a range of years from the tick boxes provided. However, in 2001, respondents will be required to write-in a four-digit year which will be ICR coded.

5.2 Changes to the Classification

In 2001, a different classification will be used to replace the Australian Bureau Statistics Classification of Qualifications (ABSCQ) which has been used for the past two censuses. The ABSCQ was first implemented during the 1991 Census and was intended to be used for approximately ten years to allow for comprehensive time series data. However, developments in education and training, particularly in the vocational education and training sector and the adoption in 1995 of a new framework, the Australian Qualifications Framework (AQF), have necessitated a new classification standard.

This new classification, known as the Australian Standard Classification of Education (ASCED) is not limited to classifying data collected by the census, or even to any data collected by the Australian Bureau of Statistics (ABS) as a whole. This new standard is intended to classify all forms of education, including high school and primary school, and can be used by any interested agency.

The scope of ASCED extends beyond that of the census variable because ASCED is intended as a classification of all education. Therefore, a respondent in the 2001 Census reporting (for example) a Statement of Attainment at Certificate III Level, would not be included in census output. Although this qualification *can* be classified according to the standard, it falls below the basic vocational requirements and is therefore out of the scope of the census. A bridging or enabling course would be treated in the same way.

Unlike the ABSCQ classification of Qualification Level, ASCED is a hierarchical classification which can code at the one, two or three-digit level. Nine broad fields exist, of which five are applicable to these census variables. These broad fields are:

- Postgraduate Degree Level;
- Graduate Diploma and Graduate Certificate Level;
- Bachelor Degree Level;
- Advanced Diploma and Diploma Level; and
- Certificate Level.

Qualification Level data from the 2001 Census will be available at the two-digit level only. For example, it will not be possible to distinguish between Doctorates and Masters by research or coursework, nor between pass and honours bachelor degrees (all of which are base level categories).

The classification of Qualification Field continues to have a three level hierarchy. However, the first level of this classification now contains 12 broad fields. Eight of these fields remain the same or similar to the ABSCQ:

- Natural and Physical Sciences;
- Engineering and Related Technologies;
- Architecture and Building;
- Agriculture, Environmental and Related Studies;
- Health;
- Education;
- Management and Commerce; and
- Society and Culture.

Four broad fields have been added to the classification:

- Information Technology;
- Creative Arts;
- Food, Hospitality and Personal Services; and
- Mixed Field Programmes.

The addition of these fields has updated the classification in keeping with changes in the pattern of education and training. For example, in the ABSCQ 'Computer Science' was a

detailed (three-digit) field whereas in the ASCED, 'Computer Science' is a two-digit field which includes 11 different detailed fields. The broad field 'Mixed Field Programmes' has been added largely to enable the coding of broad types of qualification, such as primary schooling, secondary schooling, social and employment skills courses.

More information about the ASCED can be obtained from the ABS publication: *Information Paper- Australian Standard Classification of Education*, Cat. No. 1271.0 (not yet released).

APPENDIX 1: 1996 Census Sequencing of Questions Relating to Qualification

23   **Has the person *completed* a trade certificate or any other educational qualifications since leaving school?**

( ) No ➜**Go to 28**
( ) No, still studying for first qualification ➜**Go to 28**
( ) Yes, trade certificate/ apprenticeship
( ) Yes, other qualification

24   **What is the *highest* qualification the person has *completed* since leaving school?**
● For example, trade certificate, bachelor degree, associate diploma, doctorate.

Full name of qualification
....
....
....

25   **What is the main field of study for the person's *highest* qualification *completed*?**
● For example, history, plumbing, primary school teaching.

Field of study
....
....
....

26   **At which institution was the person's *highest* qualification *completed*?**
● If completed overseas also state which country.

Name of Institution
....
....
....

27   **In which year did the person *complete* their *highest* qualification?**

( ) Before 1971
( ) 1971 - 1980
( ) 1981 - 1985
( ) 1986 - 1990
( ) 1991 - 1992
( ) 1993 - 1994
( ) 1995 - 1996

APPENDIX 2: ABSCQ - Example of Broad, Narrow and Detailed Qualification Field

1. BUSINESS AND ADMINISTRATION

10. Business and Administration NFD
100. Business and Administration NFD
11. Management
110. Management NFD
111. Business Management
112. Public and Institution Management
113. Personnel Management
114. Hospitality Management
119. Management NEC
12. Management Support Services
120. Management Support Services NFD
121. Office Management
122. Keyboard and Shorthand
129. Management Support Services NEC
13. Sales and Marketing
130. Sales and Marketing NFD
131. Wholesale and Retail Sales
132. Marketing
133. Real Estate
134. Tourism
139. Sales and Marketing NEC
14. Financial Services
140. Financial Services NFD
141. Accounting
142. Banking and Finance
143. Insurance
149. Financial Services NEC

APPENDIX 3: Reconciliation between Census and Transition from Education to Work

TABLE A1: FREQUENCY OF QUALIFICATION LEVEL BY FIELD OF STUDY, 1996 CENSUS

	Qualification Level									
Field of Study	Higher Degree	Grad. Dip.	Bach. Deg.	Under Grad. Dip.	Assoc. Dip.	Skilled Voc.	Basic Voc.	Not Stated	Inadeq. Desc.	Total
Business & Admin	28,057	20,607	166,979	66,395	116,158	23,957	165,743	133,591	39,911	761398
Health	24,998	14,772	153,831	141,421	15,421	6,588	64,579	33,944	22,422	477,979
Education	18,436	94,358	154,276	105,890	29,943	0	604	14,626	3,571	421,704
Society & Culture	46,100	27,770	261,323	42,677	49,268	30,402	28,132	33,362	11,256	430,290
Natural & Physical Science	32,695	10,705	129,643	16,818	23,166	5,634	20,069	17,321	3,027	259,078
Engineering	16,957	3,332	90,082	25,129	64,007	728,025	30,048	41,121	6,486	1,005,187
Architecture & Building	1,260	815	18,759	6,875	12,651	257,804	10,596	12,327	3,669	324,756
Agriculture & Related	3,122	882	12,499	10,457	11,078	31,194	16,735	9,245	1,320	96,532
Misc.	97	68	1,027	12,187	4,137	193,263	31,023	26,217	9,470	277,489
Not Stated	1,892	350	2,463	787	2,127	24,803	935	772,580	502	806,439
Inad. Desc.	2,035	1,130	20,611	2,619	3,907	7,532	3,703	4,627	2,431	48,595
Total	**175,649**	**174,789**	**1,011,493**	**531,258**	**331,863**	**1,309,202**	**372,167**	**104,065**	**1,098,961**	**5,009,447**

TABLE A2: FREQUENCY OF QUALIFICATION LEVEL BY FIELD OF STUDY, MAY 1996 TRANSITION FROM EDUCATION TO WORK SURVEY

	Qualification Level							
Field of Study	Higher Degree	Grad. Dip.	Bachelor Degree	Undergrad Diploma	Assoc. Dip.	Skilled Vocation.	Basic Vocation.	Total
Business & Admin	32,672	33,983	177,621	37,804	164,738	97,108	531,042	1,074,969
Health	24,065	35,943	184,716	111,023	46,051	28,744	130,018	560,560
Education	25,161	96,330	164,899	102,807	75,900	32,522	7,258	504,878
Society & Culture	51,222	47,196	275,101	33,445	77,647	55,616	20,580	560,807
Natural & Physical Science	40,166	21,194	141,305	12,573	50,189	19,278	8,102	292,806
Engineering	17,815	11,110	104,407	21,900	197,471	780,287	20,289	1,153,280
Architecture & Building	3,435	4,291	26,631	5,630	34,255	342,545	4,681	421,469
Agriculture & Related	5,549	2,244	16,085	5,039	29,197	48,833	21,591	128,537
Miscellaneous	83	0	1,930	4,824	52,063	295,693	38,888	393,481
Total	**200,169**	**252,291**	**1,092,695**	**335,046**	**727,512**	**1,700,625**	**782,450**	**5,090,787**

Reference List

Australian Bureau of Statistics (1993) *Australian Bureau of Statistics Classification of Qualification (ABSCQ)* Cat. No. 1262.0

Australian Bureau of Statistics (1994) Census Working Paper 94/2, *1991 Census Data Quality: Education*

Australian Bureau of Statistics (1996) Census Working Paper 96/2, *1996 Census Form Design Testing Program*

Australian Bureau of Statistics (1999) Census Working Paper 99/6, *1996 Census Data Quality: Occupation*

Census Working Papers

96/1   1991 Census Data Quality: Income

96/2   1996 Census Form Design Testing Program

96/3   1996 Census of Population and Housing: Digital Geography Technical Information Paper

97/1   1996 Census: Homeless Enumeration Strategy

99/1   1996 Census: Industry Data Comparison

99/2   1996 Census: Labour Force Status

99/3   1996 Census Data Quality: Housing

99/4   1996 Census: Review of Enumeration of Indigenous Peoples in the 1996 Census

99/5   2001 Census: Indigenous Enumeration Strategy

99/6   1996 Census Data Quality: Occupation

00/1   1996 Census Data Quality: Journey to Work

If you would like a copy of any of these papers, or have any other queries, please contact Rosa Gibbs on (02) 6252 5942 or Email: rosa.gibbs@abs.gov.au

The papers are also available on the ABS website at www.abs.gov.au