Language & Equity

A discussion paper for writers of school-based assessment instruments

E J Bell
N Simpson

November 1995

Contents

Foreword

Few readers of the ministerial statement on equity published in Board syllabuses would disagree with the principles and ideals in that statement. However, this is not to say that we all know how to put our ideals into action.

This study has been designed to offer writers of assessment instruments a tool for exploring issues of language use in assessment, particularly as they relate to gender equity and quality in assessment. It aims to offer some information about the kinds of questions that writers of assessment instruments could ask when they scrutinise their own instruments. The study identified these questions using a broad range of experienced Queensland teachers who looked at real assessment instruments used in senior schools. For these reasons we can have some confidence that the questions about language use in assessment instruments identified in this study are both relevant to the Queensland context and based on the perceptions of teachers themselves.

To my mind, achieving quality and equity in assessment is a goal that requires continuous effort and a willingness to ask questions of the assessment instruments we design and write. I hope that Queensland teachers will find this study one of a range of useful resources in discussing the equity and quality of their own assessment instruments.

John A Pitman
Director

1 Summary

This study explores one aspect of gender equity in educational assessment — how gender is represented in the language of assessment items. A sample of assessment instruments from schools across Queensland was scanned by state review panellists. The results of the study are useful for teachers wishing to examine issues of language use, gender equity, validity and quality in the writing of assessment instruments.

The research method and sample size of this study do not support broad and sweeping conclusions about language use in any one school or in any one subject across the State. Rather, the study was specifically designed to lead to further constructive discussion of gender and language in assessment instruments in the teaching community.

The data obtained show that for most categories of language use most scanners did not record any observations. That is, more often than not, no gender references of the kind scanners were asked to identify were found in assessment instruments. This may suggest a general absence of human figures in assessment instruments or a reluctance to record these references on the part of scanners (or both), or that these references were not visible to many scanners. This result indicates the value of scrutinising assessment instruments to see if it is appropriate to include more references to human figures.

Where scanners found references of the kind listed on the scan sheet, the data show:

- scanners found more references to men when gender inclusive language could have been used, as well as references to the achievements of men:

    for writers of assessment instruments this result indicates the value of scrutinising language to see if it reflects *contemporary standards of gender inclusive language*

- for every sample of assessment instruments from each subject included in this study, not only are there more references to males, they are more often 'positive references':
    for writers of assessment instruments this result indicates the value of scrutinising assessment instruments to see if they are *sufficiently equal in terms of the extent to which they contain positive or affirming references to men and women*

- there are gender differences in the number of positive references to males/females that do not appear to be prescribed by the syllabus (i.e. work program writers appear to have had a 'choice' about inclusion of these references); again, the data suggest relatively more of these kinds of positive references to men than women:
    for writers of assessment instruments this result indicates the value of scrutinising assessment instruments to see if they suggest *good choices about ways of representing gender* that are informed by the educational equity statement provided by the Minister for Education for inclusion in Board syllabuses

- in relation to the visibility of references to gender, subject experts appear more likely to find more references to gender than non-experts; there are no significant differences between male and female scanners in terms of the *total* number of references to gender they found; however, female scanners, particularly non-experts, appear much more likely to find certain *kinds* of references to gender:
    for writers of assessment instruments these results indicate the value of *asking different teachers to read assessment instruments* to see if the language of these instruments has any equity, validity and quality implications that may not be apparent to any one reader.

The study was designed to stimulate discussion about how language is used to represent gender. It offers writers of assessment instruments a basis for scrutinising assessment instruments in particular ways as part of the process of evaluation of the quality of these instruments. The contribution this study has to make is by way of showing that there are particular areas of language use in assessment that could usefully be the focus of evaluation and discussion of assessment instruments. These are areas identified in a scan of some real assessment instruments by a broad range of experienced Queensland teachers.

2   Parameters and aims of the project

2.1   What was the study designed to do?

Any study of gender issues will provoke vigorous discussion and disagreement, no matter how it is designed and reported. This 'representations of gender project' was designed to stimulate discussion about how language is used to represent gender in assessment instruments in Queensland senior secondary schools. One hundred and eighteen school sets of assessment instruments collected as part of another project (the 1993 Pilot Random Sampling Project[1]) were scrutinised in this study of how language is used to describe gender in assessment instruments.

This study explores just one area of gender equity. That is, the project explores a particular aspect of gender equity in educational assessment — how gender is represented in the language of assessment items. The results of a scan of assessment instruments performed by BSSSS state review panellists offer information to the teaching community which should be a useful point of departure for further discussion about issues of gender representation in a range of subjects.

2.2   What does language use have to do with gender equity?

One question that might be asked is 'what relationship does the representation of gender in assessment instruments have to gender equity?'.

One of the premises of gender equity in assessment has been that stereotyped representations of women, and curriculum that is not gender inclusive or does not recognise the life experiences and contributions of women and girls, as well as other factors to do with how women and girls are represented in texts, do influence educational outcomes. Addressing unequal educational outcomes involves addressing issues to do with how gender is represented in the language of assessment instruments. This position is articulated in the statement on educational equity supplied by the Minister and published in BSSSS syllabuses. It is a position on gender equity about which there is some agreement in educational research into gender equity and equal outcomes, whether it concerns school or tertiary education. Yet it is not the same as saying that the language of assessment instruments is the single factor producing unequal outcomes: common sense tells us this is not so.

Precisely how important the representation of gender in language is to educational outcomes is not an answerable question. Yet the relationship of the two continues to be the focus of educational debate, with arguments also about the relative importance of other factors to educational outcomes. For example, a recent study by the Scottish Examination Board concludes that differences in outcomes in that senior school system 'are more likely due to gender differences in the experience of and attitudes to schooling than to the wording of examination questions'.[2] However, the Scottish study (as well as other studies) begs the question of what does shape students' experience of, and attitudes to, schooling and what we can do about these apparently elusive influences on differences in outcomes. If we

---

[1] See *Random Sampling of Student Folios: A Pilot Study*, BSSSS, May 1994.
[2] Scottish Examination Board, Research Bulletin 3, 1995.

accept that human realities are shaped by and in language, then it seems logical to say that the language we use to represent gender in assessment instruments is one way in which students' experience of school (and their self-perception) is shaped.

While it might not identify all these elusive influences, it could be argued that a study that examines representations of gender in assessment instruments can offer some indications about the nature and presence of at least some factors that are thought to affect educational outcomes for girls and boys, and ultimately equal participation in particular professions.

2.3 How can the findings be used?

The findings of a study that involves interpretive decisions made by scanners about the meaning of language representing gender in a limited number of assessment instruments must be carefully weighed. That is, the results of this study must be weighed by an awareness that the data are based on readings of a relatively small number of assessment instruments *in each subject*. Further, scanners drawn from the Queensland teaching community have considered only one, albeit very important, source of information about teaching practices — assessment instruments. They have not considered a whole range of other teaching practices relevant to gender equity, such as verbal encouragement. Given this fact and the fact that only a small number of assessment instruments in each subject are involved, the data obtained do not offer a basis for broad and sweeping conclusions about teaching practices in any one school or in any one subject across the State. That is, this study is one that has been quite specifically designed to lead to further constructive discussion of gender and language in assessment instruments in the teaching community. To the extent that this study is used for this purpose, its findings will have been interpreted correctly.

One way of understanding the parameters and aims of this study is to ask 'what is the value, to teachers, of further discussion of this report?'. The answer to this question involves many areas of teacher interest, not least of which is the quality of assessment instruments used in the classroom. Teachers and the Queensland community generally have an interest in the quality of the assessment instruments used in classrooms across the State; that is, the extent to which they allow students to show what they know and can do. Equity issues are of particular interest because while equity is not synonymous with quality, it is often argued that you cannot have one without the other. For these reasons, information about the language used to represent gender in assessment instruments is information that can be used to discuss issues to do with the quality of those assessment instruments.

Accordingly, the information obtained in this study about how state review panellists interpreted the language of assessment instruments should be of use to all teachers wishing to reflect upon issues to do with quality teaching; that is, how teaching practices may promote equal opportunity for all students to show what they know and can do, as well as equal participation of all groups in education and employment.

3   Background

As suggested earlier, the use of language in assessment instruments is only one of the broad range of factors in senior secondary assessment thought to affect equality of participation and outcomes. The question of how gender is represented in language is in turn only one aspect of how communication systems (language and other sign systems such as pictures and cartoons) are used in assessment instruments. Any analysis of language, whether about gender or not, can raise very complex issues indeed. However, the particular approach used in this study is intended to offer an exploratory discussion of some reasonably accessible language issues relevant to assessment instruments. The purpose of this section is to describe some features of the debate about particular ways of representing gender that can usefully preface a reading of the results of this study.

3.1   What is the relationship of language use to validity and quality in assessment?

In a recent discussion of the sources of unfairness in tests, Gipps and Murphy (1994) note with reference to the American context:

> One source of poor performance on a test is the test material itself, e.g. the inadequate or stereotyped representation of some groups . . . In the USA this is referred to as facial bias, defined as when particular words or item formats appear to disfavour one group whether or not they have that effect. For example, using the male pronoun 'he' throughout or involving only male figures in the items would count as facial bias whether or not this affects the scores of women. (p. 25)

The representation of gender is one area of potential bias in assessment items that is understood to affect the validity of assessment items — whether assessment items measure what they are designed to measure. In equity terms the issue of validity ('do these assessment items measure what they are supposed to measure?') involves asking a range of questions about whether assessment items have features that could unfairly disadvantage particular groups of students. That is, the validity and quality of assessment are perceived to be affected if assessment instruments present inappropriate barriers to particular groups of students showing what they know and can do. The quotation from Gipps and Murphy suggests that the ways in which we refer to gender can present a barrier to the equal opportunity of all students. Interestingly, Gipps and Murphy also suggest that we can identify a facial bias or an *apparent* bias of assessment items without being absolutely certain it disadvantages a particular group in terms of educational outcomes. *Since it is not always possible to identify how the language of assessment items affects educational outcomes, the practice of scrutinising the language of assessment to identify facial bias may well be one of several ways of evaluating the validity and quality of assessment instruments. The particular issues raised by this study, some of which are discussed below, may be useful to teachers wishing to identify possible bias in assessment instruments.*

3.2   What about gender inclusive language?

Gender inclusive language has been given some attention in this study, not least because it continues to be the subject of discussion in educational equity research as well as in the general community.

For example, in an article by Whitehouse entitled 'Research into gender equity and assessment' (1992), the following perspective is given:

> Some people continue to argue that the generic term 'man' does mean human. I suppose a statement like 'man has walked on the moon' is in itself a complete truth (no woman has set foot there). But, as Dale Spender points out in her book *Man Made Language*, reading a text which states 'man is a warm blooded animal who suckles his young' is a complete nonsense.

While there may be some agreement on the language that should be avoided, conveyed in various statements on non-discriminatory language[3], it is often much more difficult to obtain agreement about the extent to which particular issues are addressed by using particular forms of language. Is it enough to avoid certain ways of using language and use other forms, such as gender neutral language? Whitehouse goes on to argue that:

> it has been common practice to do away with what is overtly masculine and replace it with what is neutral (*human* substituted for *man*, etc.). This is supposed to make the document gender neutral, but the effect is to still render women invisible and their contributions invisible.

Whitehouse's article suggests that more frequent references to women and the achievements of women can be part of a proactive affirmative action through language.

3.3   Is the representation of women still an issue?

Of course, not all readers will agree with Whitehouse that there is a need for this affirmative action for girls through language use. In recent years the argument has been raised that girls are doing better than boys and that equity programs have neglected boys. This argument has included assertions that the learning needs of boys have been neglected in the drive for gender equity.[4]

While it is not the purpose of this study to evaluate particular positions about differences in outcomes[5], the findings of this study should raise some important issues for those involved in developing assessment instruments for use in schools, a task that necessarily involves scrutiny of the language of those assessment instruments. The fact that the Queensland data since the late eighties show changing patterns of participation and outcomes by gender is a fact that can be considered quite separately from another question addressed by this study: What issues about language use are suggested by data from a scan of assessment instruments performed by state review panellists? Are there differences in the representation of males and females that still need to be addressed? The present study is as much about the representation of males as it is about the representation of females: the scan sheet used in the project (see appendix) was designed to collect data on the representation of both males and females.

---

[3] See in particular *A Fair Deal: Equity Guidelines for Developing and Reviewing Educational Resources*, Queensland Department of Education, 1991.
[4] See, for example, an article appearing in the *Courier-Mail* (26.11.1994) entitled 'Girls are smarter – survey'.
[5] A recent study (1995) by the Senior Secondary Assessment Board of South Australia, *Gender Equity in Senior Secondary School Assessment*, examines gender differences in outcomes in considerable depth.

3.4   Why should we consider subject resources?

Another key issue to do with the representation of gender raised by this study is subject resources — references to texts by men or women, including extracts from texts as well as titles of books and articles and so on by men or women. Work programs and assessment instruments used in Queensland schools may not reflect male and female authorship equally for a broad range of reasons, including the fewer number of female authors in some fields.

Of course, such questions about differences in the gender of those who produce the books students are required to read in a course of study really means thinking about other questions to do with *who* produces the knowledge that students must learn. Questions like 'who produces the books we read for English?' or 'who produces the knowledge we learn in mathematics?' are a necessary part of considering 'does the presence of a gender imbalance in who produces the books and knowledge students are required to learn advantage or disadvantage particular groups of students?'. The Board's syllabus evaluation process offers a way in which this critical evaluation can take place. However, teachers can also consider these questions when designing and writing assessment instruments.

The issue of how many and what references to texts by women or men one can find in assessment instruments and teaching materials can be part of broader interest in how the construction of knowledge may favour particular groups of students, and present barriers to other students showing what they know and can do. For example, we might ask 'is the absence of texts and theories by women part of the construction of subject knowledge in ways that exclude the experiences and achievements of women?'.

This question and others like it inevitably confront particular values in the community. With reference to the English context, Gipps and Murphy (1994) note:

> recent changes in GCSE practice to broaden the range of texts used to include female writers and those from ethnic minorities, as well as writing which is more accessible to pupils, has been highly contentious because it brings up the issue of what is 'valuable' to learn. (p. 26)

Again, the purpose of this study is not to debate the merits of these and other positions in the equity debate. However, in relation to this issue this study offers interesting information about the extent to which a group of professional teachers found assessment instruments refer to texts by women or men, information that should be useful to those wishing to explore these questions further.

3.5   What about using references to human figures?

This question is not so much about differences between how we refer to males and females, as differences between subjects in the extent to which human figures *per se* are referred to. Scrutiny of the enrolment figures for Board and Board-registered subjects might provide some support for the perception that there is an apparent tendency of females to choose subjects for study that are 'more human' or more oriented to analysis of social and cultural issues, rather than subjects in which the knowledge to be acquired has no or very little reference to human figures and social relationships. This apparent tendency raises questions about why this should be the case and if the absence of references to human figures in some subjects might be a barrier to the participation of some students. It also raises questions about whether the presence or absence of references to human issues in

assessment instruments in any one subject is something we should simply accept. Clearly, the presence or absence of references to human figures and issues in assessment instruments raises important questions about students' experience of a subject. In this study, differences between subjects in the total number of references to gender (male *and* female) found by state review panellists in the sample of assessment instruments examined should offer a point of departure for further discussion of this set of questions.

3.6 How might language use in assessment influence role-modelling for male and female students?

The absence of references to women, and moreover particular kinds of references to women, might well have an important bearing on how female students develop a positive self-image and gain a sense of control over social realities. For example, one question might be whether the portrayal of men rather than women as achieving professionals in assessment instruments in a particular subject should be seen as not promoting equal access to education because it does not promote girls' positive self-image.

Why should we be interested in whether assessment instruments present both girls and boys as professionals in a subject? (An example might be a description of a scientist conducting a laboratory experiment in Biological Science.) The importance of attitudes and the fact that girls and boys develop different attitudes to certain disciplines is something teachers are well aware of. Attitudes often affect confidence, motivation and performance, as well as students' perceptions of the relevance of certain subjects for their future lives. One concern in the past has been that females more often than males have negative attitudes to mathematics and science. If we think of our 'attitudes' as meanings we construct *in language*, the value of critically evaluating the language in which gender is constructed in each subject seems indicated. Of course, the way in which gender is represented in teaching and resource materials for any subject cannot *explain* the attitudes of students in those subjects, but it can suggest something about *just one of many possible influences on students' attitudes to a subject that teachers can explore*.

There are many other language issues to do with the equity, quality and validity of assessment items; however, this study aims to offer a possible starting point for discussion and teacher scrutiny of assessment instruments. As noted above, the relationship of particular ways of representing gender in language to educational outcomes, and gender differences in performance in specific educational contexts, is not explored by this study. The fact of gender differences in educational outcomes, and differences in the participation of different groups in education, cannot be easily explained by referring to apparent bias in the representation of gender in assessment instruments or to any other single factor. However, the fact that there are observable gender differences in outcomes might make us look at a whole range of factors, including teaching materials at primary, secondary and tertiary level, to see what the appropriate response might be at each level.

What did a group of experienced teachers find in a sample of assessment instruments drawn from schools across Queensland when they were asked to read this material for the presence of some of the above kinds of references to gender?

4 Method

4.1 What, in a sentence, is the study method?

The study method can be summarised thus: a group of state review panellists scanned sets of assessment instruments from school subject-groups across Queensland and counted the number of particular kinds of references to gender using the scan sheet provided.

4.2 What is the scale of the study?

Sixty-five state review panellists participated in the study as scanners of assessment instruments. Ten meetings were held over twenty-six hours yielding 359 scan sheets. Ten subjects were included in the study: English, Geography, Economics, Mathematics I, Mathematics in Society, Chemistry, Biological Science, Accounting, Theatre, and Health and Physical Education (HPE). These subjects have the highest enrolments in each of the eight learning areas of the National Profiles (excluding languages). One hundred and eighteen sets of school assessment instruments were viewed by scanners: these materials were collected as part of the 1994 BSSSS study of 1993 folios, *Random Sampling of Student Folios: A Pilot Study*. That is, the 118 sets of school assessment instruments were a random sample of assessment instruments drawn from schools across Queensland (there are 310 schools in Queensland — 95 schools were included in the study). The study used twelve school sets of assessment instruments in each subject except Accounting and English, where eleven school sets were used. This fact should itself curb any tendency to use the data to make sweeping generalisations about language use across Queensland in any of these subjects.

4.3 What are the limitations of the study?

The findings of a study that involves interpretive decisions made by scanners about the meaning of language representing gender in a limited number of assessment instruments must be carefully weighed. That does not mean the data are too inconclusive to offer a basis for discussion of language use in assessment instruments.

The results of this study must be weighed by an awareness that the data are based on readings of some assessment instruments by one group of educators. There are many other sources of information about the representation of gender in schools that may be relevant to gender equity but are not considered in this study. A more complicated way of saying this is to note that many factors that are thought to influence equitable outcomes and participation are not accessible through a procedure for reading and interpreting the language used in assessment instruments, and that those meanings panellists can 'read off' assessment instruments are not conclusive (they are imperfectly present or not completely accessible for scrutiny). That is, there are many factors not considered in this study method that are extremely difficult to study, such as unwritten behavioural texts to do with aspects of a school environment encouraging girls' and boys' active participation and achievements. In short, assessment instruments change, are mediated in various ways, and do not tell the 'whole story' about the representation of gender. If scanners found some instances of language that they considered not to be gender inclusive, they did not at the same time have access to other teaching instruments, strategies and practices.

For these reasons alone, broad judgments about how gender is being represented in one Queensland classroom or Queensland senior schools generally cannot be made on the basis of this dataset. Nor would it be correct to use this data to characterise teaching practices in any one subject across the State: the small sample used in this study offers no basis for such generalisations. We can be on much surer ground if we treat this data as information that teachers may use *together with other information* to reflect upon issues of gender representation in their teaching practices.

4.4 Who was involved in the study?

Each school set of assessment instruments was looked at about three times — by at least one expert in the subject and at least one non-expert in the subject. The term 'subject expert' refers to a state review panellist who is a panellist in the subject in which that individual acted as a scanner. 'Subject non-expert' refers to a state review panellist who scanned assessment instruments for subjects in which that individual is not a panellist. Thirty state review panellists took part in the study as 'subject experts' while thirty-five panellists took part as 'non-experts'.

The strategy of asking both subject experts and non-experts to view the same set of assessment instruments from a subject-group was used to explore the possibility of difference between these two kinds of scanners. The relevant questions here are 'is there any tendency for experts or non-experts to count more references to gender under the various categories?' and 'what are the implications of any such differences?'.

The study was also designed to obtain a gender balance where possible. Table 1 shows the number of panellists (whether expert or non-expert) and their gender for each subject (non-experts acted as scanners in more than one subject). Of course, the study also raises questions about gender differences in the visibility of particular kinds of references to gender. For example, did female experts tend to count more of one or another kind of reference to gender than male non-experts?

Table 1:   Scanners by gender, 'expertness' and subject

Subject	Subject expert			Non-subject expert		
	F	**M**	**Total**	**F**	**M**	**Total**
Accounting	4	–	4	5	6	11
Biological Science	3	–	3	9	5	14
Chemistry	1	1	2	7	6	13
Economics	1	2	3	6	8	14
English	2	1	3	4	6	10
Geography	–	2	2	8	5	13
HPE	–	2	2	10	5	15
Mathematics I	2	2	4	7	7	14
Mathematics in Society	2	2	4	4	10	14
Theatre	3	–	3	9	5	14

It should be noted that the state review panellists who participated in the study were located in South-East Queensland, in the first instance Brisbane, and to a lesser extent the Sunshine Coast, Gold Coast and Toowoomba. Meetings with state review panellists for the purpose of the scanning exercise were held in the Office of the Board in Brisbane.

4.5   What did participants in the study do?

The meetings took the form of a presentation of about twenty minutes' duration by a Board research officer, after which time state review panellists scanned photocopied sets of assessment instruments using a scan sheet — an A3 sheet with fifteen categories (see appendix). These scan sheets were used to collect the data for this study. A scan sheet records the number of references to gender per category found by a state review panellist in a set of assessment instruments for a particular school subject-group. Scanning was not just a simple 'tick and flick' task; it required interpretive judgments by state review panellists about whether the words on the page in assessment items could reasonably be counted under a particular category on the scan sheet. An example of a category on the scan sheet is:

3.   references to 'he' or men or 'she' or women when gender inclusive language
('she' and 'he' or preferably no reference to gender) could have been
substituted without affecting the substantive nature of the assessment item

'he' or men	Total		'she' or women	Total

Table 2 shows the number of scan sheets completed by experts and non-experts for each subject.

Table 2:   Number of scan sheets completed by subject

Subject	Number of scan sheets completed by:	
	subject experts	non-subject experts
Accounting	11	22
Biological Science	12	24
Chemistry	12	24
Economics	12	24
English	11	22
Geography	12	24
HPE	12	25
Mathematics I	12	28
Mathematics in Society	12	24
Theatre	12	24

4.6 How did participants approach the task?

4.6.1 What information was given to scanners?

It is precisely because scanning involved interpretive judgments that there is a need to report the ways in which the task was 'framed' or explained in information about the task given to scanners. That is, the ways in which the task was explained to scanners does have a bearing on how the results obtained should be understood because such 'framing' of the task almost certainly influenced the interpretive judgments made by these scanners.

A standardised presentation was delivered by Board research officers at the commencement of each meeting to ensure that each scanner involved in the study was given the same information about the nature of the task they were about to undertake. Scanners were:

- advised that the role of Board researchers did not include offering information about the appropriateness or correctness of their decisions about the meaning of language in assessment instruments

- asked to make judgments that can be reasonably supported by pointing to specific evidence of the way language is used in assessment instruments — observable or 'point-at-able' features of language

- advised that they were to focus on the meaning the language in assessment instruments had *for them*, and that they were not required to make judgments about the appropriateness of particular teaching practices, the intentions of teachers producing these assessment instruments, or other aspects of the school environment

- asked to interpret the categories to include or exclude references to gender, taking the broadest interpretation of a category they felt they could reasonably take

- advised that for Part One of the scan sheet the same reference to gender could be counted in different categories, but the same reference to gender could not be counted more than once under a single category; similarly, for Part Two of the scan sheet the same assessment instrument could be counted in different categories, but the same assessment instrument could not be counted more than once under a single category

- advised that some sets of assessment instruments may yield only a few references that seem relevant to particular categories on the scan sheet while others may yield many references that can be counted under many categories

- advised that their decision not to count a reference in any category on the scan sheet offers information that is as useful as their decision to count a reference; accordingly, returning blank scan sheets that represent decisions not to include any items whatsoever is as important and valuable to the project as returning scan sheets that record many references under the categories provided

- advised that, for the purposes of counting references to gender under the scan sheet categories, they were not to count every word denoting gender mechanically; rather, scanners were asked to make on-balance decisions about whether particular words referring to gender could reasonably be 'clumped' and counted as one reference because, for example, they appear on the same page as a series of connected sentences with an overall meaning that seems relevant to a particular category (i.e. they were asked to count a gender reference in an assessment instrument as any word or group of words that represents a *discrete example* of a category on the scan sheet)

- given specific examples from different subjects of how the categories might be interpreted in a way that emphasised this was only one possible interpretation of those categories.

In short, scanners were encouraged to make their own decisions about the meaning of references to gender in assessment instruments, rather than recording their perceptions of the views of other teachers, students, Board researchers, or the Queensland community. The emphasis was upon inviting scanners to supply their perceptions about the appropriateness of the fifteen categories on the scan sheet for the assessment instruments they viewed. Scanners were encouraged to feel free to return blank scan sheets if they felt they could not reasonably count any references to gender in assessment instruments under the fifteen categories on the scan sheet. Therefore, it is reasonable to suppose that the data supplied by state review panellists represent, for the most part, their own judgments about the language of assessment instruments.
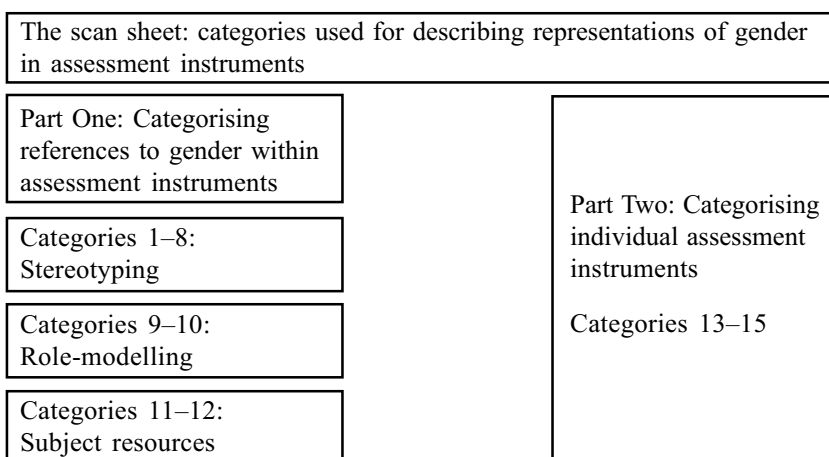
4.6.2 What did scanners use to count references to gender in assessment instruments?

The scan sheet used by scanners contained fifteen categories for describing references to gender. It needs to be emphasised that the scan sheet was *not designed to include every possible reference to gender* that might be found in assessment instruments, but rather *particular kinds of references* that have an obvious relevance to current and relatively accessible debates about the representation of gender in language. Accordingly, the return of a blank scan sheet did not necessarily mean that a scanner had found no references to gender *whatsoever* in a particular set of assessment instruments. Rather, the return of a blank scan sheet signified that a scanner could not find any references to gender that could be included under the *particular categories listed on the scan sheet*. Notwithstanding this, the scan sheet categories were sufficiently broad that it is reasonable to suppose they would cover *most* references to human figures in assessment instruments. The categories in the scan sheet were derived from:

- the educational equity statement provided by the Minister for Education for incorporation in Board syllabuses
- statements on language use and equity drawn from different educational contexts across Australia (tertiary and secondary school assessment)
- educational literature such as that referred to in the background section of this report.

The appendix to this study provides a copy of the scan sheet with fifteen categories that can be schematised as shown in figure 1.

Figure 1: The structure of the scan sheet

The scan sheet: categories used for describing representations of gender in assessment instruments

Part One: Categorising references to gender within assessment instruments	Part Two: Categorising individual assessment instruments
Categories 1–8: Stereotyping	Categories 13–15
Categories 9–10: Role-modelling	
Categories 11–12: Subject resources	

Scanners were asked to place a stroke in the relevant box when they felt material in assessment items could be counted under that category. The categories on the scan sheet involve judgments about the presence of particular ways of representing gender in assessment instruments viewed by scanners. Some of these judgments are more complex than others. For example, category 12 required scanners to count the number of:

12. references to texts by men or women (include extracts from texts as well as titles of books and articles etc. by men or women)

men	Total		women	Total

However, category 5 required scanners to make interpretive judgments about what a gender reference in a negative context might be:

5. references to ... in a negative context (as having a low status, as being victims etc.)

men	Total		women	Total

The scan sheet was designed to allow scanners to count references to both men and women in each category, whether these be references to gender in active or passive roles or whether these be references to gender in the context of descriptions of achievements or achieving individuals, or any other category on the scan sheet. That is, the scan sheet was designed to collect differences in the extent to which males and females are represented in terms of the fifteen categories. Under each category, the box for recording references to men appears before the box for recording references to women: this was quite intentional, to ensure that no errors in scoring were introduced by the reverse expectation.

The structure of the scan sheet, which was presented to scanners as a single A3 sheet, requires some further elaboration. Figure 1 schematises the scan sheet; it also shows how categories on the scan sheet were grouped to assist scanners to conceptualise related categories. Categories 1 to 12 are labelled 'Part One: Categorising references to gender

within assessment instruments'. That is, Part One of the scan sheet concerns images of gender in the language of assessment instruments, whether single words, phrases, sentences, or paragraphs: scanners were asked to categorise the language *within* assessment instruments. Categories 1 to 8 are grouped under the heading of 'stereotyping': that is, images of gender that could be loosely grouped under the idea of gender stereotyping. Categories 9 to 10 are grouped under the heading of 'role-modelling', while categories 11 to 12 are grouped under the heading of 'subject resources'.

Categories 13 to 15 are labelled 'Part Two: Categorising individual assessment instruments' and entailed making overall judgments about assessment instruments in a school set. That is, scanners had to make on-balance judgments about whether the three categories in Part Two were relevant or not to each of the assessment instruments in a school set, indicating with a stroke in the box provided if the category seemed true of a particular assessment instrument *considered as a whole*.

The background discussion in this report suggests something of the rationale for including particular categories and should be useful to a reading of the scan sheet.

However, the categories in Part Two should be briefly discussed here to signal their purpose; for example, category 13:

13. assessment instruments in which a student's comprehension or processing of the meaning of one or more assessment items *might* be influenced by not being male or *might* be influenced by not being female

not being male	Total	not being female	Total

This category was designed to gather a specific kind of judgment about individual assessment items. The scanner was not being asked to make judgments about whether particular assessment instruments were likely or certain to disadvantage either male or female students. The question was designed to assess something else — scanners' judgments about gender differences in comprehension based on their scrutiny of the language of assessment instruments, whether it includes overt references to gender or not. Scanners giving a positive response to category 13 may have judged that, for example, gender differences in comprehension of a particular assessment instrument might exist because of references to cultural contexts that males, rather than females, would be familiar with (an English comprehension examination question based on reading an excerpt about the game of football might be one example). Category 13 asked readers to make inferences about whether or not student understanding or processing of an item, however that is understood, might be influenced by not being male or not being female.

Category 14 in Part Two was also designed to add to the information supplied by scanner tallies under other categories.

This category required scanners to make assessments about whether there were any:

14. assessment instruments in which an opportunity for students to consider the experiences and achievements of women or men could have been included without affecting the substantive nature of the assessment instrument

... men could have been included	Total

...women could have been included	Total

That is, category 14 was designed to gather judgments about the extent to which scanners thought assessment instruments could have been altered to include representation of the experiences and achievements of women or men without affecting the validity of the assessment items — what they were designed to measure. Responses to this category might indicate educators' opinions of the possibility or opportunity for considering the contributions and experiences of women and girls within the scope of what assessment instruments have apparently been designed to assess.

A related category, category 15, asked panellists to count:

15. assessment instruments in which an opportunity for students to consider *gender issues* relevant to the subject could have been included without affecting the substantive nature of the assessment instrument.

... gender issues could have been included	Total

5 The data

5.1 How should the data be interpreted?

5.1.1 How were results obtained from the data?

If correct use of the dataset in any study involves an awareness of its limitations, it also involves consideration of ways of treating the data to produce results. The dataset obtained in this study, which used around three different scanners for the same set of assessment instruments, presents particular challenges. That is, there is a need to obtain *one* result for each set of assessment instruments in a subject from a school that accurately reflects the data from the *three* scanners of that set. An average is not appropriate because the data from one scanner may well swamp the results of two other scanners who are in fairly close agreement. The solution to this problem was to use the median (or middle) value for the three or four scans of each school set of assessment instruments in a subject. That is, if we took the average of three scanners' scores — three, three and fifty — we would have a value 'nineteen' which did not represent the views of any scanner, whereas if we took the median we would have a value 'three' which represents the views of two scanners. In this way one value for each set of assessment instruments was derived from the three scan sheets returned by different scanners of that set.

A further feature of the data should be highlighted. There is insufficient commonality between schools and subject packages to examine any differences by school, or by school type. That is, given that the study covers ten subjects, and that the assessment instruments are drawn randomly from different schools, the combination of subject data from each school is not sufficiently similar to allow meaningful comparisons. The study does not allow comparisons of data for a particular combination of subjects to be made between different schools, and therefore different kinds of schools (such as government or non-government schools, or single-sex versus coeducational schools).

In asking questions about differences in the references counted by female or male and expert or non-expert scanners, we might be interested in knowing what the data indicate about the visibility of references to gender for these different groups of users of assessment instruments. The difficulty in answering these questions lies in the fact that for any set of assessment instruments for a subject from a school the scanners may not have divided neatly into the various categories so that a comparison cannot be made between males and females, as well as experts and non-experts for *every* set. Accordingly, a method for comparing scanners' judgments of the same object — the same set of assessment instruments — was devised.

5.1.2 What can explain differences in scanners' scores?

Use of the data should also be informed by an awareness of issues of reliability in this study. As indicated in the discussion of the data that follows, there were a very few instances where one, two or three scanners returned a completely blank scan sheet while at the same time another scanner of the same set of assessment instruments did not. Broadly speaking, there were very few school sets of assessment instruments that varied in these terms, although when we examine individual categories the situation is quite different,

leading us to ask the general question 'how can it be that one scanner sees many gender references while another sees none?'.

The presence of scanner variation is not surprising — scanners were making interpretive decisions about the meaning of language. Studies of reliability generally indicate that the more complex the decision the less reliability there will be across different scorers. It is possible that some scanners reading the same set of assessment instruments were performing the task with quite different understandings of the categories. Variation may be a function of the fact that scanners reading the same assessment material were performing the task with quite different belief systems about the appropriateness of language used to represent gender, just as in the broader community there are differences in attitudes to, and use of, the language in which gender is constructed.

In some cases variation may also be a function of difficulties in understanding or applying the categories, or unintentional errors in the actual counting of references. It needs to be emphasised that the task asked of scanners also involved making decisions about what represents a fresh instance of each particular category in often long and detailed documents — assessment instruments.

However, closer scrutiny of specific cases where scanners returned very different scan sheets was undertaken in three cases of relatively great variation to further explore possible reasons for this variation. These three cases involved three sets of assessment instruments — one for Maths in Society and two for Biological Science. Close reading of the scan sheets and the sets of assessment instruments involved and comparison of these sets with others involved in the study suggest that the more complex the assessment instrument in terms of the range and number of references to gender, the more difficult the task of using the scan sheet and correspondingly the more individual variation between scanners.

With reference to specific categories, the three cases examined suggest that:

- category 1 [neutral] (which asked scanners to count the number of references that seemed gender neutral as opposed to references that suggested gender stereotypes of traits or attributes) may have presented particular problems for some scanners; e.g. some scanners simply counted *every* reference to gender while other scanners counted only well-known examples of gender neutral language, such as the words 'he/she' used repeatedly to refer to a scientist conducting an experiment in a short-answer question in Biological Science.

- some scanners also appear to have counted *every* reference to gender in assessment instruments under category 3 ('references to "he" or men or "she" or women when gender inclusive language ["she and he" or preferably no reference to gender] could have been substituted without affecting the substantive nature of the assessment item'); e.g. a scanner counted all references to gender in a description of reproductive processes in Biological Science.

- some scanners appear to have taken a much broader interpretation of the meaning of category 4 than other scanners ('references to "he", "man" or "mankind" or other language referring to men is used or "she" or "women" or other language referring to women is used when describing matters central to the subject itself'); i.e. some scanners took a very broad interpretation of what might be central to Maths in Society or Biological Science.

- some variation seems to be a function of greater willingness to count the same references under more than one category; for example, counting references to men 'in the context of descriptions of skills, activities or games relevant to the subject'
(9 – men) as being also in some sense references to the 'achievements of men, high achieving men, or men as students or professionals in the subject' (10 – men).

5.2   What are the results of the study?

5.2.1   What about scanners who did not find any gender references?

Any consideration of the data obtained by this study needs to be informed by the fact that *for most categories most scanners returned scan sheets with no observations recorded. That is, for most categories most scanners indicated no gender references of the kind listed on the scan sheet were found in assessment instruments.* This result may indicate the general absence of human figures from assessment instruments, or a reluctance by some scanners to record such references, or that these references were not visible to scanners. Given that a reasonable interpretation of the categories on the scan sheet would cover most references to human figures in assessment instruments, the study raises some interesting questions about the absence of references to human figures generally (not just differences in references to males and females).

Having said this, other questions remain: for how many sets of assessment instruments did scanners find the categories on the scan sheet were not relevant to the assessment instruments and decide not to count any references? That is, of the 118 school subject-group sets of assessment instruments included in the study, for how many sets were entirely blank scan sheets returned by all three scanners? For how many sets did two scanners return an entirely blank scan sheet? For how many sets did one scanner return an entirely blank scan sheet? Further, we need to ask 'what is the number of scan sheets returned with zero references *by category*?'.

In fact, of the 118 school sets of assessment instruments included in the study, in only one case did three scanners return an entirely blank scan sheet (no references counted under all fifteen categories) and this was for a set of assessment instruments that had four scanners (the fourth scanner did not return a blank scan sheet). For only two sets did two of the three scanners return an entirely blank scan sheet. For only five sets did one scanner return an entirely blank scan sheet.

Figure 2 is a bar graph that gives information about *the number of scan sheets with zero in that category*. It demonstrates the point made above — that for nearly all categories most scan sheets were returned with zero references found. This is itself an important finding, with the most likely explanation lying in the extent of references to human figures *per se* in assessment in most subjects included in the study. However, we can see from this graph that subcategory 11 [women] (references to key theories by women) was the category in which the highest number of scan sheets were returned with a zero result. Subcategory 3 [men] (references to men when gender inclusive language should have been used) as well as subcategory 10 [men] (references to the achievements of men) were subcategories in which the lowest number of scan sheets were returned with zero references found. Those findings are themselves useful indicators of areas of language use that might need to be the focus of discussion by writers of assessment instruments.

Figure 2:   Total number of scan sheets = 359
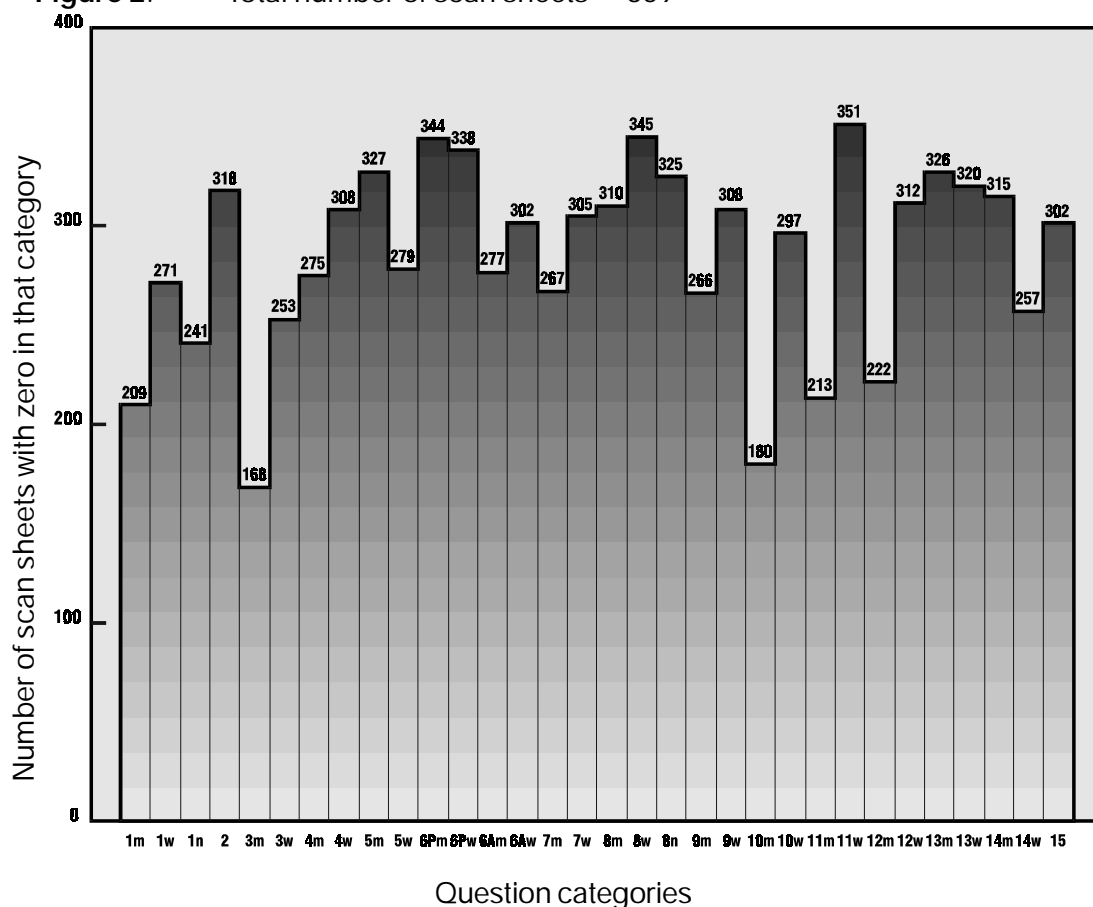
5.2.2 What about the data from scanners who did find gender references?

As shown above, for most categories most scanners indicated that no gender references of the kind listed on the scan sheet were found in assessment instruments. However, the data produced by scanners who *did* find some of these references to gender offer answers to three questions:

- is there a difference in the total number of references to females and males by subject?
- is there a difference in the number of references to males and females in particular kinds of categories?
- is there a difference in the number of references to gender recorded by female or male scanners, expert or non-expert scanners?

The answers to these questions can be used to reflect upon:

- assessment materials; and
- issues to do with the visibility of references to gender for different users of assessment instruments (male–female, expert–non-expert).

5.2.3 What differences in references to females and males were found?

As noted above, in order to find a way of exploring differences in the number of references to females and males found by scanners, the median of scanner responses was recorded for

each category. The median figure represents what could be described as the 'consensus' opinion about a set of assessment instruments for a school subject-group: it has the advantage of providing a single value for each set of assessment instruments from the scores of three scanners. Of course, the median is only a crude value, but it offers a way of asking questions about differences between *sets of assessment instruments* based on what might be called the typical scanner for any one school subject-group set.

Figure 3 reports the data for subjects when we take the sum of these single values — medians — for categories 1, and 3 to 12. Care should be taken with this graph: it does not offer a basis for hasty conclusions about subject differences, even if we restrict our conclusions to the particular sample of assessment instruments included in the study. Common sense tells us that, the more words used in an assessment instrument, the higher the likelihood of references to both the male and female gender. Some subjects seem to be characterised by assessment instruments with a greater number of words. We would not expect all subjects to be equal when we take into account the differences in the extent to which assessment in some subjects seems to require more words than assessment in other subjects. For example, a Biological Science multiple-choice test may contain a great deal of stimulus material whereas an Economics essay test may contain very little.

For these reasons, we should pay attention to the ratio of references of males to females obtained for the different samples of assessment instruments from each subject. For example, the graph shows a high number of references to males in the sample of English folios along with a high number of references to females in comparison with other subjects: the subject English may well involve more references to human figures *per se*. At the same time, the subject Chemistry involves fewer references to males and females but, in comparison with English, there seem to be proportionately more references to males than females. As indicated earlier, caution should be exercised in making generalisations about subjects on the basis of this sample size. While in the Mathematics I sample the ratio of references to males and females is roughly equal, the very small number of references in total means that we should be careful about concluding that there is no imbalance in the representation of gender in this subject.
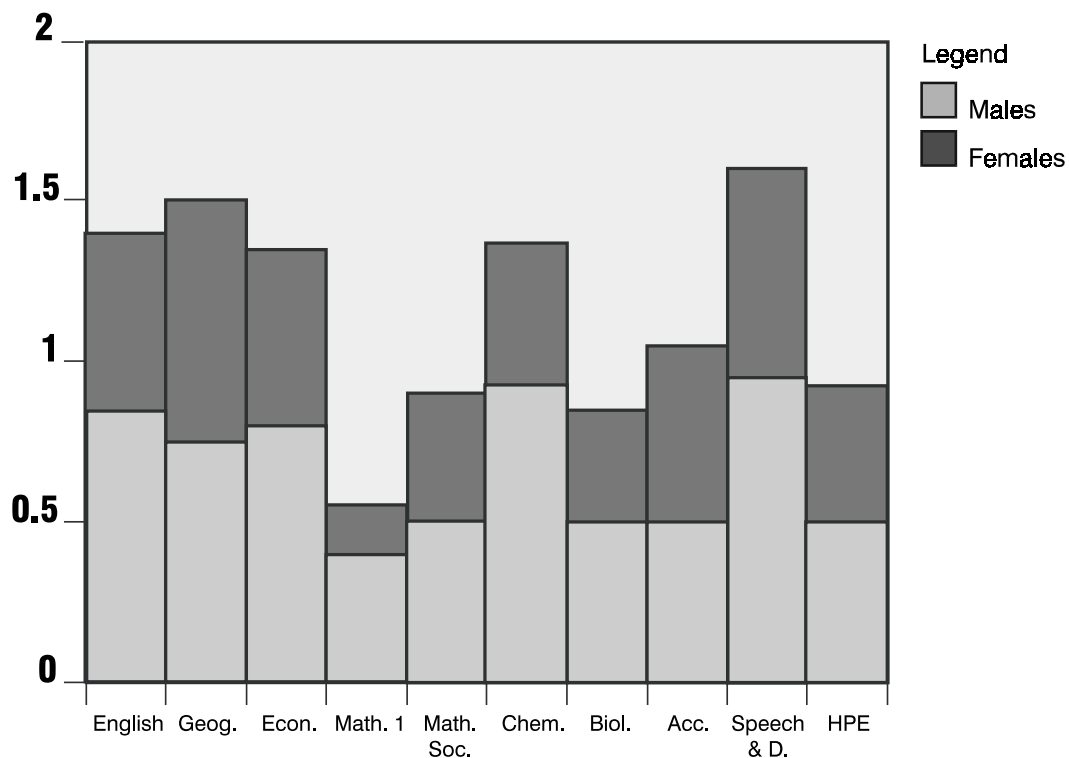
Figure 3:   References to gender
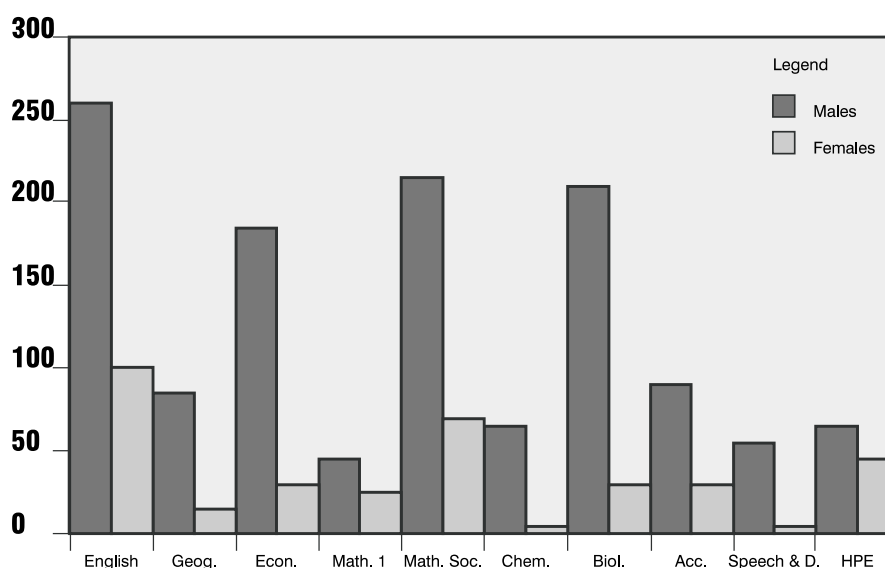
Another way of looking at differences between references to females and males is to look at particular *kinds* of categories, perhaps by grouping categories that seem to 'belong' together. One way of grouping these categories is to look at gender differences in positive versus negative references to gender. It seems reasonable to see references to males under categories 6 [active], 7, 8, 9, 10, 11, 12 as being positive, rather than negative. For example, it seems more likely that a reference to men as property owners or financial players (e.g. names of companies such as Jack Casey and Sons Pty Ltd or Jill Casey Pty Ltd) is more likely to be positive. Similarly, scanners who counted references to females under categories 6 [active], 7, 8, 9, 10, 11, 12 were more likely to be observing positive, rather than negative, references to females. There is no certainty that this was always the case but it allows us to look at the data from another perspective.

Figure 4 shows the responses of scanners by subject as a proportion of total responses for the categories (e.g. positive references to males expressed as a proportion for English is the number of positive references to males for English divided by the total number of references to males). The graph shows that, for *every sample* of assessment instruments *from each subject* included in this study, not only are there more references to males, they are more often 'positive references'. The number of references to males is proportionally greater in the 'positive' categories.

A further way of grouping categories can be considered here. Given that work programs are developed by schools from the subject syllabus, can we identify positive references that probably did not originate in the syllabus? While it is true that it is difficult to identify with certainty all references to gender where a 'choice' existed for those producing assessment instruments — a choice whether or not to include a particular item — a number of observations about this issue may be made. To begin with, we can be sure that the nature

**Figure 4**:　　　Positive references (proportion)

and incidence of references to gender in assessment instruments cannot be simply or wholly explained by pointing to the syllabus.

For example, the observation might be made, upon reading the syllabus for the subject Accounting, that a reference to a company name such as 'Bob Morrison and Sons' in an Accounting test is a reference that is not derived from the syllabus: it is a reference that might originate in the work program. Other references in another subject to Albert Einstein's theory of relativity, might be identified as references that are a function of the construction of knowledge given in the syllabus, references that must be used to develop the work program.

Figure 5 shows gender differences by subject for positive references where it might be argued some choice exists. Category 11 (references to key theories by men or women) is not included, as it could be argued that this is a category that records forms of language that writers of assessment instruments had no choice about. Again, the data suggest relatively more positive references to men than to women.

Some readers of figure 5 might argue that, for example, with respect to category 7, if a subject syllabus requires study of the workings of real financial institutions in society, then of course references to those institutions in assessment instruments will reflect the fact that men, more than women, have a greater role in them. This argument might involve the assertion that it is not the function of assessment instruments to be agents of change in society, and that it is inevitable that they reflect society. Other readers might feel that

Figure 5: Positive references (where a choice exists)
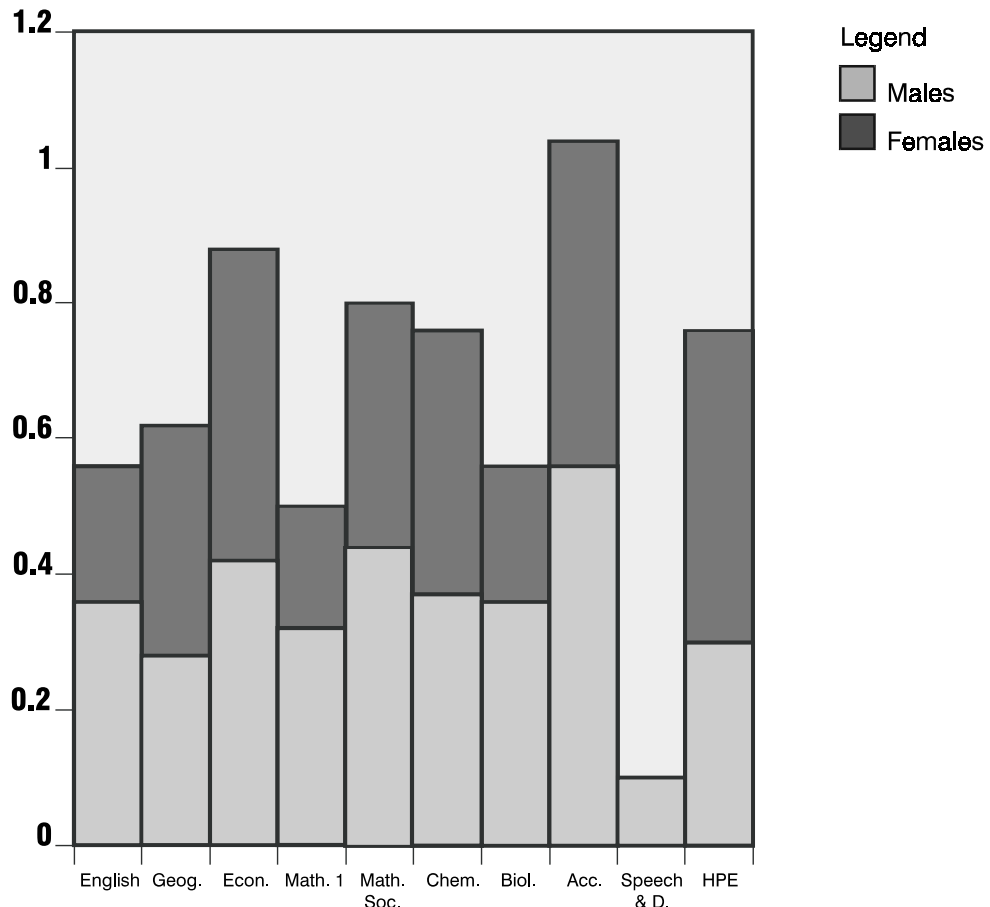
assessment instruments should reflect good practice and highlight inequality in contemporary society. These readers might argue that what is at stake is better understood as a choice about *which* social reality we will reflect in assessment instruments. Further, they might assert that good choices about how we represent gender in assessment are not necessarily based on the assumption that males and females behave in identical ways.[6]

5.2.4 What differences are there between the data from different groups of scanners?

Each set of assessment instruments (for a subject-group at a school) was seen three times by scanners who might be female or male, or who might be experts or non-experts. It would be interesting to know which scanners were more (or less) likely to find more references in the same assessment materials (regardless of what kind of references these might be).

Of every possible pair of scanners looking at the same package, 51 per cent of the time the scanner finding more references was male — a figure that matches expectations given the uneven allocation of scanners to the same assessment materials. This suggests that there is no significant difference between male and female scanners in terms of how likely they were to find more gender references *per se*. On the other hand, if there were no difference between experts and non-experts we would (given the allocation of scanners) expect the scanner recording more references to be an expert 42 per cent of the time. The observed proportion was 51 per cent: therefore, we can conclude that in this study experts are more likely to find more references than are non-experts.

This analysis can be extended to cover four categories of scanners: female experts, male experts, female non-experts, male non-experts. Table 3 shows the expected and observed proportions of the category of scanner finding the higher number of references for all items and for category 6.

The difference for category 6 is marked: female, particularly female non-expert scanners, are much more likely to record a greater number of references. That is, female scanners, particularly non-experts, appear much more likely to find references to men and women in active and passive roles. This suggests that there may well be important differences in the visibility of particular kinds of gender references for different readers of assessment instruments. The significance of this finding might be that it indicates the value of asking different teachers to read assessment instruments to see if the language of these instruments has any equity, validity and quality implications that may not be apparent to any one reader.

Table 3:   Expected and observed proportions of scanners finding the highest number of references

	Female experts	Male experts	Female non-experts	Male non-experts
Expected proportion	20.1%	22.1%	28.8%	28.9%
Observed all items	24.8%	25.7%	24.6%	24.9%
Observed category 6	26%	18%	40%	15%

[6] Analysis of the robustness of these findings for positive references to gender shows that the proportions of male–female references do not change significantly when we distinguish positive references that probably did not originate in the syllabus from other positive references.

6 What are the implications of these results for writers of assessment instruments?

The findings discussed above are summarised in the first section of this report. A final question remains: 'What implications do these findings have for writers of assessment instruments who wish to scrutinise these instruments for language issues that may affect assessment validity and quality?'. While the answer depends on the particular contexts in which any writer of assessment instruments is working, this study does suggest the value of scrutinising assessment instruments with particular questions in mind.

The data show that for most categories of language use most scanners did not record any observations; i.e. more often than not, no gender references of the kind scanners were asked to identify were found in assessment instruments. This may suggest a general absence of human figures in assessment instruments or a reluctance to record these references on the part of scanners (or both), or that these references were not visible to many scanners. This result indicates the value of scrutinising assessment instruments to see if it is appropriate to include more references to human figures.

Where scanners found references of the kind listed on the scan sheet, the data show:
* scanners found more references to men when gender inclusive language could have been used, as well as references to the achievements of men:
  > for writers of assessment instruments this result indicates the value of scrutinising language to see if it reflects *contemporary standards of gender inclusive language*
* for every sample of assessment instruments from each subject included in this study, not only are there more references to males, they are more often 'positive references':
  > for writers of assessment instruments this result indicates the value of scrutinising assessment instruments to see if they are *sufficiently equal in terms of the extent to which they contain positive or affirming references to men and women*
* there are gender differences in the number of positive references to males/females that do not appear to be prescribed by the syllabus (i.e. work program writers appear to have had a 'choice' about inclusion of these references); again, the data suggest relatively more of these kinds of positive references to men than women:
  > for writers of assessment instruments this result indicates the value of scrutinising assessment instruments to see if they suggest *good choices about ways of representing gender* that are informed by the educational equity statement provided by the Minister for Education for inclusion in Board syllabuses
* in relation to the visibility of references to gender, subject experts appear more likely to find more references to gender than non-experts; there are no significant differences between male and female scanners in terms of the *total* number of references to gender they found; however, female scanners, particularly non-experts, appear much more likely to find certain *kinds* of references to gender:
  > for writers of assessment instruments these results indicate the value of *asking different teachers to read assessment instruments* to see if the language has any equity, validity and quality implications that may not be apparent to any one reader.

The important point that needs to be made about the above issues is that they are issues emerging out of real assessment instruments used in Queensland schools that were scanned by a broad range of experienced Queensland teachers.

7 Recommended reading

Courtemanche, L.M. & Einbinder-Miller, R. 1984, *Guidelines Respecting the Elimination of Sexist Language in Departmental Communications*, Department of Justice, Ottawa.

Curriculum Development Centre 1976, *Avoiding Sex Bias: Guidelines for the Avoidance of Sex Bias in Educational Materials and Media*, CDC, Canberra.

Department of Administrative Services 1988, *Guidelines for Non-discriminatory Language*, Canberra.

Department of Education, Queensland 1991, *A Fair Deal: Equity Guidelines for Developing and Reviewing Educational Resources*, Brisbane.

Gipps, Caroline & Murphy, Patricia 1994, *A Fair Test? Assessment Achievement and Equity*, Open University Press, Buckingham.

Kett, M. & Underwood, V. 1978, *How to Avoid Sexism: A Guide for Writers, Editors and Publishers*, Ragan, Chicago.

Maggio, R. 1987, *The Non-sexist Wordfinder: A Dictionary of Gender-free Usage*, Beacon Press, Boston.

National Council of Teachers of English 1975, *Guidelines for Non-sexist Use of Language in NCTE Publications*, NCTE, Urbana, Illinois.

Nilsen, A.P., Bosmajian, H., Gershuny, H. Lee & Stanles, J.P. (eds) 1977, *Sexism and Language*, National Council of Teachers of English, Urbana, Illinois.

Pauwels, Anne 1991, *Non-discriminatory Language*, AGPS, Canberra.

Scottish Examination Board, Research Bulletin 3, 1995.

Senior Secondary Assessment Board of South Australia 1995, Gender equity in senior secondary school assessment, unpublished study.

Sorrels, B.D. 1983, *The Non-sexist Communicator: Solving the Problems of Gender and Awkwardness in Modern English*, Prentice Hall, Englewood Cliffs, New Jersey.

Spender, Dale 1985, *Man Made Language*, 2nd edn, Routledge & Kegan Paul, London.

Stobart, G. et al. 1992, 'Gender bias in examinations: how equal are the opportunities?', *British Educational Research Journal*, vol. 18, no. 3.

Whitehouse, H. 1992, 'Research into gender equity and assessment: a personal viewpoint', *Opinion*, vol. 21, no. 4, pp. 11–13.

Appendix: Copy of scan sheet

(The original A3 scan sheet is reproduced here in A4 format.)

Name _____   For assessment instruments:

      identify school: _____   identify subject: _____

Part One: Categorising references to gender within assessment instruments

Stereotyping

1. references to ... which suggest stereotypes of traits or attributes, or are neutral

men	Total	women	Total	neutral	Total

2. references to men and women which seem to 'reverse' or write back against gender stereotypes of traits or attributes

'reverse'	Total

3. references to 'he' or men or 'she' or women when gender inclusive language ('she' and 'he' or preferably no reference to gender) could have been substituted without affecting the substantive nature of the assessment item

'he' or men	Total

'she' or women	Total

4. references to 'he', 'man' or 'mankind' or other language referring to men is used or 'she' or 'women' other language referring to women is used when describing matters central to the subject itself

'he', 'man' or 'mankind'	Total

'she' or women	Total

5. references to ... in a negative context (as having a low status, as being victims etc.)

men	Total	women	Total

6. references to ... represented as having a passive role/active role

	men	Total		women	Total
Passive					
Active					

7. references to ... as property owners or financial players (e.g. names of companies such as Jack Casey and Sons Pty Ltd or Jill Casey Pty Ltd)

men	Total	women	Total

8. references to male-dominated or female-dominated or gender neutral social institutions, clubs or organisations

male-dominated	Total	female-dominated	Total	neutral	Total

Role-modelling

9. references to ... in the context of descriptions of skills, activities or games relevant to the subject

men	Total		women	Total

10. references to the achievements of men, high achieving men, or men as students or professionals or the achievements of women, high achieving women, or women as students or professionals in the subject

achievements of men ...	Total		achievements of women	Total

Subject Resources

11. references to key theories by men or women

men	Total		women	Total

12. references to texts by men or women (include extracts from texts as well as titles of books and articles etc. by men or women)

men	Total		women	Total

Part Two: Categorising individual assessment instruments

Number of assessment instruments in this set   Total

	Total

13. assessment instruments in which a student's comprehension or processing of the meaning of one or more assessment items *might* be influenced by not being male or *might* be influenced by not being female

not being male	Total		not being female	Total

14. assessment instruments in which an opportunity for students to consider the experiences and achievements of women or men could have been included without affecting the substantive nature of the assessment instrument

... men could have been included	Total		... women could have been included	Total

15. assessment instruments in which an opportunity for students to consider *gender issues* relevant to the subject could have been included without affecting the substantive nature of the assessment instrument.

... gender issues could have been included	Total