**Submission to the Senate Legal and Constitutional Affairs Committee in relation to an Inquiry into the Australian Citizenship Amendment (Citizenship Testing) Bill 2007**

Prepared by Professor Tim McNamara,

Professor of Applied Linguistics and former Director, Language Testing Research Centre, School of Languages and Linguistics, The University of Melbourne.

## 0. Summary

This submission is a revised version of an earlier submission in response to the Discussion Paper released by the Government last year prior to the introduction of the current proposed legislation. This submission mainly addresses a specific aspect of the proposal for a new citizenship procedure, the incorporation of a formal *language* test to replace the existing informal language test as part of the citizenship test (the writer is a leading international authority on language testing). It argues that the Discussion Paper which provides the rationale for the legislation shows little understanding of issues of validity in language tests. It considers the following issues: how we can make any proposed language test fair; what any test might contain; what level should be required in order to 'pass'; the likely impact of the test; and the values implicit in the test. It also draws attention to the existing wealth of resources in language testing research available in Australia. It stresses the following problematic features of the proposal: the naïve belief that a change in test format favouring an objectively scored, technologically delivered test will represent a fairer test; the requirement of a higher level of proficiency than appears to be intended in the wording of existing legislation, a level moreover requiring levels of literacy and communicative skill which many Australian citizens who are native speakers of English do not possess; and the way in which the proposed language test represents a test of cultural values as much as of language proficiency. It also argues that any test of knowledge of Australian society and values that is administered solely in English would represent a *de facto* language test and would be subject to the same considerations and objections as are made in the light of suggestions for a formal language test. It would moreover appear to contradict the part of the legislation

requiring applicants for citizenship to possess 'a *basic* knowledge of the English language'. Australia has a sorry history of the misuse of language tests, and there is a significant likelihood that the current proposal, if implemented, would constitute a further such episode.

## 1. Introduction and scope of submission

This submission addresses a particular aspect of the proposed legislation, the proposal to introduce a *formal language* test as part of a revised citizenship procedure. Existing Australian citizenship procedures have long involved an *informal* language test, in line with long-standing practice in a number of countries, including the United States and (until recently) the United Kingdom. The proposed innovation is that there should be a *formal* language test, again following recent developments in certain countries abroad (the United Kingdom, Germany, The Netherlands, and others). The Discussion Paper describes these developments in the following terms:

> *§55. Both the UK and the Netherlands have gone a step further and have formal, consistent and objective language testing in place. The level of language proficiency required is defined in specific linguistic terms, with test arrangements making use of professional educational language testing expertise.*

This submission will address the assumptions in this statement, and the following questions:

> *Question 13*
> *Should prospective Australian citizens be formally tested for their level of English? If so, would it be necessary to test oral, written, reading and listening skills?*

> *Question 3*
> *What level of English is required to participate as an Australian citizen?*

## 2. Language tests and validity

Current validity theory in language testing offers a comprehensive framework for considering the development and implementation of such a test (Bachman, 1990; Messick, 1989; Kane, 2001; McNamara, 2006; McNamara & Roever, 2006). It stresses that the development of a language test involves much more than technical considerations in the production and implementation of an efficient instrument, but requires test developers to address not only questions of the fairness of the test, but also of its intended use, the values it embodies, and its impact both on educational systems and on stakeholders, including test takers and the community at large. Each of these issues will now be considered in relation to the proposals for a formal language test in the Discussion Paper.

## 3. How can we make any proposed language test fair?

All language tests, both formal and informal, are easily prone to unfairness; the risk of unfairness in informal language tests is simply more obvious. It is an important goal of language testing research to understand the risk and sources of unfairness in formal language tests, and to attempt to control and reduce them, while recognizing that they can never be eliminated entirely.

Thus, the association in the Discussion Paper between formal language tests and 'objectivity' is unwarranted, as in the reference in §55 to formal language tests as 'consistent' and 'objective', and the characterization of the current arrangements as 'more subjective than a formal test would be' (§52). While it is true that formal tests, if they are properly managed, can be fairer than informal assessments, they are not automatically fair, as they necessarily involve subjective judgements about test content and test format on the part of test developers. Further, language tests in particular are very likely to contain subjectively assessed components (speaking and writing ability) which rely on human judgement. In turn, all such subjectively assessed aspects of assessment can also be made acceptably fair if they are managed properly. In the Olympic Games, for example, it is accepted that some sports (diving,

gymnastics, and so on) must be judged subjectively, and measures are adopted to maximize the fairness of these judgements.

The Discussion Paper appears to recommend the 'objective' testing formats used in some European citizenship tests:

> *§20. The UK and the Netherlands ... have both made use of available technology in developing and administering their tests. For example, prospective citizens of the UK... do so on-line at approved testing centres. Test questions are randomly generated by computer from a large bank of questions and, as the questions are multiple choice, test results are able to be 'marked' electronically with results being readily available.*

> *§70. A computer-based test could be administered consistently and objectively throughout the country.*

However, 'objective' tests such as those involving multiple choice formats can be very unfair if they don't allow us to draw the right conclusions about a person's ability to communicate. While multiple choice testing might be suitable for *knowledge* tests, it has limited potential for performance tests of *language*, particularly of the skills involved in communication (speaking and writing).

Similarly, formal tests are *not* made more 'objective' if they are delivered on computer. In fact, for language tests, computer delivery can restrict unnecessarily the kinds of communicative tasks which might be needed for a valid test of communicative language skills. The technologically based test of speaking used in the Netherlands is the subject of intense critical discussion within the language testing community, and serious questions have been raised about its validity as a test of speaking, and particularly in this context.

A further consideration under test fairness is the need for accommodations. Test accommodations are based on notions of equity or equal playing field, that is, that there are certain factors beyond people's control that mean that they may be unfairly disadvantaged. Age may be one, because very young and very old may lack the

cognitive abilities of mature adults, but there are a host of others, such as sight, hearing and other disabilities. (Significantly in this context, some tests accommodate for limited English proficiency, for example in tests of subject matter knowledge at school for the children of recent immigrants.) The Discussion Paper makes limited reference to accommodations in terms of possible exemptions from the test on the grounds of age. There is a growing literature on language testing accommodations. A list of references to this literature is contained at the end of this submission.

## 4. What should the test contain?

The goal of language tests such as the one proposed is to make accurate predictions about how a person will cope in situations in which they have to communicate in the language concerned. This is achieved by establishing what the relevant situations are, and creating test tasks to elicit a sample of performance on the basis of which such predictions can be made. Obviously, real world situations differ enormously in their communicative demands – a person may be able to shop successfully, or write a simple note to the teacher, but have limited ability to write a job application, or to follow a lecture. This is true for first language speakers as much as it is for second language speakers: the communicative demands of many work and educational settings are beyond the communicative capacities of many native born Australian citizens who are first language speakers of English. A lot of people are tongue tied when it comes to giving a speech, even in a friendly environment, for example; others have great difficulty explaining work procedures to other people, even when they are familiar with the procedures. Many literacy tasks are well beyond the capacity of many native English speaking Australian citizens, as assessments of literacy levels in the community have consistently shown.

Two principles should guide us in thinking about the relevant situations (the 'target language use situations') that the test aims to make predictions about. The first is that they should reflect the intention of the legislation; the other is that they should only involve tasks that all Australian born citizens who are native English speakers could handle, as educational level and literacy level are not a bar to citizenship for this latter group.

The legislation makes a simple but plain statement of the communicative requirements of citizens: it requires of an intending citizen that 'the person possesses a basic knowledge of the English language' (Australian Citizenship Act 1948 - Sect 13). The Discussion Paper tells us that

> *Policy provides that the applicants may be assessed as meeting the English language requirements if they are able to speak and understand English sufficiently to respond in simple language during the citizenship interview. Applicants must be able to answer questions in simple English concerning personal particulars (such as, how long have you lived in Australia? What are your children's names?') (§12)*

This suggests strongly that a minimum threshold of communicative skill and intelligibility is intended, so that the applicant should understand and be understood while talking about a range of immediate personal matters and needs. The Discussion Paper, however, appears to suggest a very different interpretation of the legislation. In §25 it states that 'it would be important that people have a level of English which allows them to participate through education and employment', and that the proposed citizenship test would 'provide a mechanism through which we can be assured that new citizens have sufficient English… to maximise the employment and other economic opportunities which benefit the individual and Australia' (§33). §39 states that 'It is important that people who are wishing to become Australian citizens have the language skills which enable them to fully participate in society and to communicate with fellow Australians'.

This goes well beyond the wording of the legislation, and current policy. Clarification is needed of the nature of basic language proficiency as distinct from other kinds of basic educational achievement being targeted, as care is needed to ensure that language tests are tests of language, and not of other qualities not specified in the legislation. Obviously many Australian born citizens who are native speakers of English lack the educational levels to 'maximize their employment and other economic opportunities which benefit the individual and Australia' – but this is no bar to citizenship. The Birrell *et al*. study referred to in §31 as 'confirming the direct link

between English ability and employment outcomes' refers in the main to skilled occupations.

An urgent task, therefore, is to clarify the communicative situations to be targeted by any test, as a practical interpretation of the requirement of the legislation that intending citizens possess 'a basic knowledge of the English language'. These situations should not include any that many native born Australian citizens who are first language speakers of English could not meet.

This would then allow us to address the issue of 'What should be in the test?' The Discussion Paper lists some alternatives:

> *§65. Some options to consider for inclusion in a formal test include:*
> - *a written English component*
> - *an oral English component*
> - *a separate listening and reading component*
> - *an English language test based on an educationally defined level of competency*

A crucial issue is the clarification of the question of a literacy requirement for citizenship. As there is no literacy requirement for citizenship and accompanying rights (e.g. voting) for Australian-born native speakers of English, it seems unreasonable to insist on this in a citizenship test, and current practice does not require this. The use of literacy tests as a restriction on the right to vote has a sorry history in different parts of the world, for example in the United States, where such practices were used against English-speaking Irish immigrants in the mid-19th century, against African-Americans in the Southern states following emancipation, and against European immigrants around the time of the First World War (Leibowitz, 1969; Wiley, 2005). This does not mean that literacy components could not be included for literate applicants, but that exceptions or alternative arrangements would need to be made for applicants without literacy skills. There is a precedent for such alternatives in Germany, where *Test Deutsch alfa*, a test of spoken German which does not assume literacy skills, was developed following the introduction of

citizenship legislation requiring evidence of proficiency in German (Ruebeling, 2002). The possibility of using a written test to stand in for a broader test of communicative skill, as in the United Kingdom practice, may be acceptable for those with a good level of communicative skill and a balanced proficiency in each of speaking, reading, listening and writing. But this should not be done for the cases which count most, that is, the hardest to decide cases, those on the borderline of acceptability. Here, a more direct investigation of communicative skill would need to be conducted, as appears to be the Canadian practice through the institution of Citizenship Judges.

The final point raised in §65 refers to the possibility of defining the test target in terms of educationally defined levels of competency. The definition of relevant levels of competence has been attempted in proficiency scales of various kinds over the last 50 years. One relevant attempt is in the International Second Language Proficiency Ratings (ISLPR: Wylie & Ingram, 1995/99), formerly the Australian Second Language Proficiency Ratings (ASLPR: Ingram & Wylie, 1979), originally drafted to guide migrant language education in Australia, and drawing on well-known American precedents. Numerous other scales have been developed since, including the Common European Framework of Reference (Council of Europe, 2001). The Certificates in Spoken and Written English developed within the Adult Migrant Education Program (Hagan *et al*., 1993) represent another attempt to define proficiency at such levels. The lowest levels of such scales would appear to most clearly fit the wording of the legislation, and reflect current practice. These would therefore be the levels which should be targeted by any language tests which are developed. It should be noted too that all such scales, while they have been developed and are in use, are not themselves without problems, and their validity and relevance as statements of target performance cannot simply be assumed on the grounds of their availability and convenience.

In general, if testing is going to go forward, there must be a commitment to engaging the best available expertise (see below) to develop and continuously research and revise the assessment system. In other words, in the interests of fairness, the government must not assume that this test development is a once only - and once and for all - task.

**5. What level should be required in order to 'pass'?**

We have already noted the discrepancy between the wording of the legislation, supported by current practice, and the terms in which competency is discussed in the Discussion Paper. Apart from this basic issue of the general level of the test, there is the separate issue of how to establish a cut-score (the level at which a person has shown enough communicative ability to be eligible for citizenship), a major issue in language test development. Unfortunately, the Discussion Paper conceptualises the cut-score in terms of percentage marks on the test, as for example in its discussion of 'pass marks' in the testing arrangements in place in Canada, the Netherlands and the UK. But percentage marks are in themselves meaningless, as they depend on the degree of difficulty of the test; there is no universal norm of percentage marks in terms of which satisfactory performance can be established. Accepted methods of determining cut scores need to be adopted: these usually involve getting qualified judges to follow set procedures for careful examination of the content of test questions and tasks, to see which level of performance on which tasks should determine the passing level. Allowance should also be made for the inevitable uncertainty of the test, that is, its margin of error in estimating the ability of candidates.

**6. What is the likely impact of any such test?**

Current validity theory has reached a broad consensus for a number of years now that the uses of tests need to be considered as part of a validity argument. This not only involves gathering empirical evidence in support of the determination of a cut score, but consideration of the consequences and impact of the introduction of such a test. Who would be unfairly disadvantaged by it? What consequences would it have for educational provision, and on the wider society? What are the consequences of exclusion from citizenship of those who fail the test? What factors are most predictive of failure, and what responsibility does the test have to respond to them?

The Discussion Paper makes assumptions that the test would simply have a positive impact. For example we find it claimed in §24 that 'A formal citizenship test… would provide a real incentive to learn English', and in §39 that 'an emphasis on English language in a formal test could provide a greater incentive for people to learn and improve upon their English language skills'.

In fact, there are strong grounds for questioning such claims. The area of language testing research known as 'washback' (studies of the educational and social impacts of language tests) (Alderson & Wall, 1993, 1996; Cheng *et al.*, 2004), addresses this issue. The extensive literature on language test washback shows how complex and unpredictable, and frequently undesirable, the consequences of the introduction of language tests can be: washback can be positive, or negative, or both. For example, while tests can be used as diagnostic tools in language education, that is, to assist the learning and teaching process (an aspect of language testing which is notably absent from the proposal in the Discussion Paper), tests often have a narrowing effect on teaching, particularly when the format of the test is restricted to 'objective' question types: teaching becomes test strategy preparation while more serious learning goals are put on hold.

We have much to learn from the experience of other countries, where the implementation of language tests for citizenship has begun to be studied. It is unfortunate that the Discussion Paper cites international examples only when they appear to set useful precedents for the proposed new policy, without citing what has been learned in those contexts about the consequences of their introduction. Work such as Piller (2001) and Ruebeling (2002), and the work of British researchers on the implementation of citizenship policy through language tests in Europe (e.g. Blackledge, 2005, 2006; Hogan-Brun, 2005; Mar-Molinero & Stevenson, 2005) are beginning to address this issue. Evidence is emerging that the beneficial social and educational effects of citizenship language testing cannot be assumed: for example, in the United Kingdom, continuing barriers to learning, the reasons why people may not get to the level required to pass the test, and the groups most at risk, have recently been identified (Cooke, 2006). Further such research is required before confident claims about the impact of the proposed tests can be made.

The existence of any language test will mean that Australia has a responsibility to provide language courses for intending applicants to enable them to reach the level desired. Alternatives to formal testing, such as attendance at English language courses and satisfactory course completion, may have beneficial impact, for example in leading to the greater provision of English language courses. The British experience is that while there is a strong demand for such courses, there are unfortunately long waiting lists for them. Proper provision for such courses would need to be made if they are to represent a realistic alternative.

## 7. Language tests and values

Current views of validity in language testing argue convincingly that all language tests imply values, and that test validation must make explicit and provide defensible arguments for the values in the test. The Discussion Paper is clear and explicit about the values embodied in the parts of the citizenship test to do with knowledge of Australia, and raises the question of the commitment to shared values in the pledge of citizenship. However, it is far less explicit about the function of the formal *language* test as an implicit test of cultural values, and this is one of the most controversial aspects of the test. Language is a primary vehicle of culture and cultural values, particularly in this context. The arguments in favour of the test are framed in terms of pragmatic yields for the individual concerned (education and employment) and for the wider society as a result. This results however in the anomalies that have already been pointed out, that such uses of the language go well beyond the requirements of the legislation. Instead, it makes more sense to interpret the demand for higher levels of proficiency than are envisaged in the legislation in terms of the cultural meaning that knowledge of the language is meant to imply. While the Government may wish to make an argument for a language test in terms of cultural values, it should do so explicitly and will have to defend it in the face of the inevitable opposition that it will raise, as it introduces a new and extremely controversial dimension into the debate about Australian identity and the role of language within that.

It is important that a language test not be a veil for a de facto cultural values test. Language testing has a long and notorious history of acting in this way, as a cover for

other policies and values, as the literacy examples cited above show. Australia in particular has a very bad reputation in this regard; a Dictation Test was used as the main implementation mechanism for the White Australia Policy. Language testing professionals working in Australia are very conscious of this precedent, and of the need to avoid the use of language tests to mask contentious values positions under the guise of 'objective' tests supported by "state-of-the-art' technologies'.

## 8. A 'Knowledge of Australia' test as a language test

The proposed legislation refers to the need for an applicant for citizenship to establish the following two things: that he/she 'possesses a basic knowledge of the English language' (this remains unchanged from the previous legislation) and 'has an adequate knowledge of Australia and of the responsibilities and privileges of Australian citizenship'. It is important to realize that what has been said above about a formal language test in terms of its level ('a basic knowledge of the English language') and the need to avoid a literacy requirement also applies to any test of knowledge of Australia and of the responsibilities and privileges of Australian citizenship'. That is, if such a test were administered exclusively in English, and as a written test, it would represent a de facto language test at a far higher level than is required in the legislation and would also represent a literacy test. In other words, it would be open precisely to the objections that have been raised elsewhere in this submission. The confusion between the two kinds of test is actually reflected throughout the Discussion Paper. There seems to be an implication that the knowledge of Australia test might in fact be the only test, being used as a de facto way of substantiating the language requirement; but this would seem to *contradict* the requirement of the legislation that a person possess 'a basic knowledge of the English language', and as well introduce a literacy requirement that cannot easily be defended.

## 9. Australia's current level of professional expertise in language testing

Australia is fortunate in having a high level of 'professional educational language testing expertise'; a number of leading international figures in language testing

research are based in Australia, and Australia's leadership in this regard is widely recognized. Language testing is a well-established research field within applied linguistics. Australia's capacity to carry out research and development in language testing was associated initially with the introduction of oral proficiency testing as a key feature of the Adult Migrant English Program (AMEP) in the late 1970s, and was subsequently strengthened by the establishment in the 1980s of the AMEP Research centre  (National Centre for English Language Teaching and Research) at Macquarie University, and the establishment under the National Policy on Languages in 1990 of the Language Testing Research Centre at The University of Melbourne.  This very expertise, however, constitutes the basis for questioning several aspects of the language testing proposals as outlined in the Discussion Paper.

## 10. Conclusion

There must be serious doubts as to the validity of the proposed test, in terms of many of the aspects of validity envisaged in current theories of language test validity. While certain of the technical aspects of the development and introduction of a valid test could be addressed if a formal *language* test were introduced, the functioning of a test of knowledge of Australia and of the responsibilities and privileges of Australian citizenship as a *de facto* language test is extremely problematic, as it seems to contradict another part of the legislation and to introduce a literacy requirement for citizenship which it is impossible to defend.  It is extremely hard to support the legislation as it is currently proposed.

## 10. References

Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics, 14*, 115-29.

Alderson, J. C., & Wall, D. (Eds). (1996). *Washback*. [Special issue]. *Language Testing 13*(3).

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.

Blackledge, A. (2005).  *Discourse and Power in a Multilingual World*.  Amsterdam:

John Benjamins.

Blackledge, A. (2006). The racialisation of language in British political discourse. *Critical Discourse Studies* 3,1: 61-79.

Cheng, L., Watanabe, Y., & Curtis, A. (2004). *Washback in language testing: Research contexts and methods*. Mahwah, NJ: Erlbaum.

Cooke, M. (2006). 'When I wake up I dream of electricity': The lives, aspirations and 'needs' of Adult ESOL learners. *Linguistics and Education* 17: 56-73.

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching and assessment*. Cambridge: Cambridge University Press.

Hagan, P., Hood, S., Jackson, E., Jones, M., Joyce, H., & Manidis, M. (1993). *Certificate in spoken and written English*. Sydney: NSW Adult Migrant English Service and the National Centre for English Language Teaching and Research, Macquarie University.

Hogan-Brun, G. (2005). The Baltic republics and language ideological debates surrounding EU accession. In G. Hogan-Brun (guest editor).Language and Social Processes in the Baltic Republics Surrounding EU Accession, *Journal of Multilingual and Multicultural Development* 26/5, Special Issue.

Ingram, D. E., & Wylie, E. (1979). Australian Second Language Proficiency Ratings (ASLPR). In *Adult Migrant Education Program Teachers Manual*. Canberra: Department of Immigration and Ethnic Affairs.

Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement*, *38*(4), 319-342.

Leibowitz, A. H. (1969). English literacy: Legal sanction for discrimination. *Notre Dame Lawyer 45*(1), 7-66.

McNamara, T. (2006). Validity in language testing: The challenge of Sam Messick's legacy. *Language Assessment Quarterly*, *3*(1), 31-51.

McNamara, T. & C. Roever (2006). *Language Testing: The Social Dimension*. Malden, MA & Oxford: Blackwell.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed.) (pp. 13-103). New York: American Council on Education & Macmillan.

Mar-Molinero, C. & P. Stevenson (eds) (2005). Language Ideologies, Policies and Practices: Language and the Future of Europe. Basingstoke: Palgrave Macmillan.

Ruebeling, H. (2002). 'Test Deutsch alfa' – a German as a foreign language test for

people who cannot read and write.  Paper presented at the Language Assessment Ethics Conference, Pasadena, CA, May. Retrieved 16th November, 2006 from http://www.studentgroups.ucla.edu/scalar/Scalar5/LAEC_program.html.

Wiley, T. G. (2005). *Literacy and language diversity in the United States*. 2nd edition. Washington, DC: Center for Applied Linguistics.

Wylie, E. & Ingram, D.E. (1995/99). *International Second Language Proficiency Ratings (ISLPR): General proficiency version for English.* Brisbane: Centre for Applied Linguistics and Languages, Griffith University.

## 11. References on test accommodations

Abedi, J (2001) *Language Accommodation for large scale assessment in science: Assessing English Language Learners* (Project 2.4 Accommodation). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing, University of California, Los Angeles.

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, (1974,1985 and 1999). *Standards for educational and psychological testing*.  Washington, DC: Author

Khalifa, H (2005) Are test taker characteristics accounted for in main suite reading papers? *Cambridge ESOL Research Notes* 21, 7-10.

O'Sullivan, B (2000) *Towards a Model of Performance in Oral Language Testing*. Unpublished Ph.D. Dissertation, University of Reading.

Sireci, S G, Scarpati, S, & Li, S (forthcoming) *Test Accommodations for Students With Disabilities: An Analysis of the Interaction Hypothesis*.  Center for Educational Research Report No. 522. Amherst, MA:  School of Education, University of Massachusetts Amherst.  Revised version of paper to be published in *Review of Educational Research*.

University of Cambridge ESOL Examinations (2006) *Handbook for Centres*. Cambridge: Cambridge ESOL