

# Towards More Authenticity in Language Testing

**D. E. Ingram**

Professor of Applied Linguistics and Director,  
Centre for Applied Linguistics and Languages,  
Griffith University,  
Brisbane, Queensland, 4111,  
Australia.

Email: [D.Ingram@griffith.edu.au](mailto:D.Ingram@griffith.edu.au)

©2003

Paper to the AFMLTA National Conference 2003, *Languages Babble, Babel and Beyond*, Hilton Hotel, Brisbane, 10 – 12 July, 2003.

# Towards More Authenticity in Language Testing

**D. E. Ingram**

Professor of Applied Linguistics and Director,  
Centre for Applied Linguistics and Languages,  
Griffith University,  
Brisbane, Queensland, 4111,  
Australia.

Email: [D.Ingram@griffith.edu.au](mailto:D.Ingram@griffith.edu.au)

©2003

## CONTENTS

Abstract

The Presenter

I INTRODUCTION

II THE GAP LEFT BY MOST TESTS

III ATTEMPTS TO BRIDGE THE GAP

IV THE *INTERNATIONAL SECOND LANGUAGE  
PROFICIENCY RATINGS* (ISLPR)

V THE ISLPR GLOBAL

VI CONCLUSION

APPENDIX ONE: The *International Second Language  
Proficiency Ratings* (ISLPR)

APPENDIX TWO: Excerpts from the *International Second  
Language Proficiency Ratings* (ISLPR)

REFERENCES

# Towards more Authenticity in Language Testing

## D. E. Ingram

Professor of Applied Linguistics and Director,  
Centre for Applied Linguistics and Languages,  
Griffith University, Brisbane,  
Queensland, 4111, Australia.  
Email: [D.Ingram@griffith.edu.au](mailto:D.Ingram@griffith.edu.au)

©2003

## ABSTRACT

The history of language testing (especially of attempts to measure practical language ability) is, to a large extent, the history of attempts to bridge the gap between tests and real-life language use: it is the history of progress towards more authenticity in language testing. The inherent irony in most tests is that one tests one thing, in one context, in order to say something about something else in a different context: the gap to real language performance is inevitably considerable. This paper will briefly discuss the progress in language testing (especially proficiency assessment) towards more authenticity and refer to the variables that make the gap between language testing and real-life use of language difficult to bridge. The paper will finally discuss a new approach to proficiency assessment which is both globally applicable and much closer to focussing on real-life, authentic language performance than previous approaches.

## THE PRESENTER

David Ingram holds a Chair in Applied Linguistics at Griffith University, Brisbane, Australia, where he is also Director of the Centre for Applied Linguistics and Languages. He was President of the Australian Federation of Modern Language Teachers Associations for 14 years to 1996, Vice-President of the World Federation for six years, and, from 1992 to 1996, a member of the Australian Language and Literacy Council, the principal advisory body on language policy to the Federal Minister for Education. He has been an Adjunct Fellow of the National Foreign Language Center, Washington DC, since 1995. He has written extensively in applied linguistics. A recent book is *Language Centres: Their Roles, Functions and Management* [Amsterdam: John Benjamins Publishing Company, 2001]. In the Queen's Birthday Honours list in June 2003, he was appointed a member of the Order of Australia "For services to education through the development of language policy, through assessment procedures for evaluation of proficiency, and through research and teaching."

# Towards More Authenticity in Language Testing

**D. E. Ingram**

Professor of Applied Linguistics and Director,  
Centre for Applied Linguistics and Languages,  
Griffith University,  
Brisbane, Queensland, 4111,  
Australia.

## I INTRODUCTION

The history of language testing is, at least in part, a history of attempts to bridge the gap between tests and real-life use of language. The three central goals of language teaching are, in the present writer's view, the development of language proficiency, the development of cultural knowledge and understanding, and the fostering of more positive cross-cultural attitudes [cf. Ingram 2000/2001, 2000, 2000a; Ingram, O'Neill and Townley-O'Neill 1999]. There are many purposes for which one tests and many different approaches to language testing to try to achieve those different purposes but, in this paper, the focus is on the measurement of language proficiency or the ability to use language for practical purposes. In fact, the very definition of language proficiency is fraught with difficulty.

The present writer has discussed alternative approaches to defining and measuring proficiency in other papers [e.g., Ingram 2000, 1985]. However controversial academic definitions of language proficiency might be, the practical reality is that "proficiency" is an everyday, intuitive concept and there are many practical situations where it is useful to know how well or how effectively someone can use a language for practical purposes. The language tester's task is to develop instruments that let us do that and to state the results in ways that are meaningful for those practical purposes, in other words, in the context of proficiency assessment, the aim is to develop tests which will inform us about the candidates' ability to use the language in real-life situations.

This paper will consider how various tests in widespread use attempt to bridge the gap to real-life language use and then will briefly outline attempts that are being made to increase still further the authenticity of one of those assessment procedures, that based on the *International Second Language Proficiency Ratings*, which already has probably a higher level of authenticity than most other tests.

## II THE GAP LEFT BY MOST TESTS

The inherent irony of language testing (indeed, of most academic testing) is that one tests one thing generally in order to say something about something else, one assesses one component of a skill or one aspect of knowledge of a field in order to say how much of the skill or the field the student has

mastered, or one tests in one context in order to say something about a person's ability in other contexts. So teacher education courses test students' knowledge of educational theory, methodology or psychology to see whether they are likely to be capable of teaching effectively in the classroom and maintaining a beneficial learning programme for the students over an extended period; such tests are at best minimally supported by observation of the students' teaching ability in limited periods of classroom practice. By testing candidates' language knowledge or their ability to apply that knowledge in specified language tasks in tests in the formal context of the testing room, we assume that the results will give us information on the candidates' ability to use the language in other contexts, not least in real life. Yet we know as teachers and as testers that there is often a large gap between students' ability to perform in tests and their ability to use the language in everyday real-life situations: the gap between the language tests and real-life language experience is rarely bridged.

There are many ways by which to classify language tests. These have been outlined elsewhere [e.g., Ingram 2000b, 1985] and, here, reference will be made to just one classification, which illustrates clearly the problem of bridging the gap between language tests, on the one hand, and how learners might perform in real life, on the other. The classification of language tests into indirect, semi-direct, direct tests (and beyond) reflects historical changes in our understanding of the nature of language and language learning; in language teaching, it reflects the movement from more formal to more communication-oriented methodology; and, in language testing, it reflects a growing interest in and progression towards more authenticity in language use.

When language learning was seen as a process of learning grammatical rules and vocabulary and "rewriting" from one language to another, language proficiency was measured by tests of grammatical knowledge and translation: the gap between such tests and the way in which the language is used in everyday communication was considerable. In the days of behaviourist psychology and structuralist linguistics, language was seen as patterns learned by stimulus-response habit formation and tests focussed analytically on individual patterns or "discrete points" using, typically, multiple choice tests of knowledge of elements of the language. In proficiency assessment, such tests are commonly known as *indirect tests* since they essentially test one thing, characteristically knowledge of grammar or vocabulary, in order to say something about something else, in this case, proficiency or the learner's practical language ability. Indirect tests are characteristically analytic and focus on discrete-points of language knowledge or behaviour with the assumption being made that, if learners have mastered or internalised those discrete points, they will be able to perform similarly when using the language as a whole. With indirect tests, the test results are related to some notion of proficiency usually by psychometric or norm-referencing procedures in which the results are distributed over a normal distribution curve and cut-off points are identified for different proficiency levels. The major limitations of indirect tests arise from two facts: first, language performance and hence language proficiency are more than the sum of a multitude of discrete bits and part of

the skill of language use involves being able to put all the pieces together and comprehend them when received together. Second, interpretation of the results on indirect tests is fraught with difficulty: a score such as 4 out of 7, 80% or 525 says nothing about the level of the learners' practical skills or what they can do unless such scores can be related either to performance scales in which actual language behaviour is described or to the learners' subsequent performance in some activity involving the use of the language, e.g., success in an academic programme or the ability to carry out some vocational task. Thus, the difficulty of interpreting the results of indirect tests further increases the gap to real-life language ability and lowers the level of authenticity of the test.

Subsequently in the history of language testing, when the complex, integrated and redundant nature of language was noted, language tests emerged that used the principle of redundancy, deleting items by various means and assessing the extent to which learners could replace them using the redundant features of the text to identify what was deleted. With regard to proficiency measurement, such tests are commonly known as *semi-direct tests*. They are a step nearer to real use of language since they are integrative and, though they also focus on discrete items, they integrate those items into a total language event (e.g., listening to an oral text or reading a written text) and they test knowledge of, or ability to use, the items in that total event. Typical semi-direct itemtypes include cloze, dictation, white noise and interlinear tests. Such tests resemble both indirect and direct tests in that they focus on discrete items, the results are processed and interpreted psychometrically but they occur in the context of a total language event, which puts them somewhat closer to real-life language performance, i.e., the language text is somewhat more authentic but the gap-filling task is limited in authenticity, i.e., the language behaviour remains remote from real-life. In addition, the outcome of the test, like that of indirect tests, is, generally, a numerical score with the same challenges for interpretation of the results and for authenticity as we noted for indirect tests. In other words, a considerable gap remains between semi-direct tests and real-life language performance.

In recent decades, methodology in both language teaching and language testing has focussed on the communicative nature of language while language tests, especially language proficiency assessment, have come to focus on the learners' ability to use language communicatively, using tests that range in form from those that focus on the discrete tasks that learners can carry out through to approaches that focus on the learners' total language behaviour as they use the language for normal communication purposes. The last approach has often included the use of scales that describe how language behaviour develops and are used either to explicate the results on other types of tests or are directly matched against the learners' observed language behaviour. Such tests, where the focus is directly on the learners' language behaviour, are known as *direct tests*.

Direct tests, even more than semi-direct tests, are "integrative" and focus on actual language behaviour. They are characteristically used to measure proficiency by having learners perform actual communication tasks while their

language behaviour is observed and rated against proficiency descriptors that form a scale. Scales may take many different forms, which the present writer has discussed in other papers [e.g., Ingram 2000], but the most authentic, such as the *International Second Language Proficiency Ratings* (ISLPR), describe the language behaviour (essentially the tasks and how they are carried out) that can be observed as the learner uses the language [see Ingram and Wylie 1979/1999]. There are, however, many contexts in which direct tests are difficult to use though many proponents of proficiency in language education (including the present writer) would assert that direct tests give the most accurate and readily interpretable results, i.e., they invoke the most authentic use of language and go further than indirect or semi-direct tests to bridging the gap between the tests and real-life language ability. Nevertheless, there are serious limitations on their widespread use which a current project, to be referred to later, seeks to overcome.

### **III ATTEMPTS TO BRIDGE THE GAP**

It is evident from this brief description of some approaches to language testing that the gap between language tests and real-life use of the language arises from at least three sources: first, the focus of the test is on elements within the language rather than on the whole language and real language performance; second, the results of the test are presented in such forms that it is difficult to interpret them in ways that inform the user about the learners' ability to use the language in real-life, authentic tasks; and, third, the contexts in which the language occurs during the test are very different from the contexts in which the language will be used in real life. Testers have characteristically directed their concern to various forms of validity and reliability but, for the most part, these have tended to focus around issues of test procedures, the design, construction and performance of itemtypes and items, while their relationship to real language performance or real language behaviour has been accorded less significance. There are undoubtedly good reasons why this has been so, not least the difficulty of observing learners in real-life situations, the difficulty for the tester in controlling the language in such situations, and the impracticality of observing large numbers of candidates using the language in real-life. The result is, again, that, in most language tests, we essentially test one thing in one context in order to say something about something else or, at best, about the same thing in a different context. Intuitively and popularly, we would be more convincing and the tests more immediately informative and interpretable if we could observe the candidate performing in a real-life situation, control the situation so that we make sure that we observe the full extent of the learners' ability, and so assess the adequacy of their performance in the actual situations (or samples of the actual situations) in which they will be using the language. However, to do so, has, until recently, been impracticable.

Consequently, testers have focussed on the sorts of tests that they have been able to manage in the testing room and have sought to relate either the items or the results to real life performance in a number of ways through various forms of so-called validity. This paper will not discuss validity in general but

consider some of the approaches used in major tests of English as a second or foreign language, especially approaches used in tests with which the present writer has been associated.

The *International English Language Testing System*, commonly known as the IELTS Test, was developed in 1987-88 by a joint Australian-British team, which included the present writer as the Australian representative. On the test's release, the present writer became Chief Examiner (Australia) for ten years, supervising the regeneration of the test in Australia. The test was designed to be administered *en masse*, anywhere in the world, with minimal control over the quality of the assessors. Consequently, though direct assessment approaches are used in Speaking and Writing, other parts of IELTS favour semi-direct and psychometric approaches. To try to bridge the gap to real-life use of the language in academic and training contexts, the test developers sought information from academics and trainers receiving overseas students on the sorts of tasks that students in academic and training programmes in English-speaking countries are commonly required to do and test specifications and hence itemtypes were developed to reflect those tasks as closely as possible, i.e., the itemtypes were chosen to be as authentic as possible by resembling the activities in which students engage in academic or training programmes. The results of the tests are also expressed in terms of simple performance scales whose descriptors are intended to inform the end-users about what learners can do in real-life use of the language.

Despite the efforts made in IELTS to bridge the gap between the activities undertaken in the testing room and real-life use of the language, the gap remains considerable. First, the contexts within which the language occurs in most language tests are unavoidably limited and lack the richness and distracting features of normal academic activity. The conversation that is held between the Speaking assessor and the candidate is unavoidably dominated by the assessor despite the efforts that have been made in various versions of IELTS to throw some onus onto the candidate, and the range of topics that are discussed and the relationships between the interlocutors are limited by the test situation. In addition, the level of the test is pre-determined and, even though the test is designed to cover a span of the proficiency scales rather than focus on a single level, it is inevitable that, for some candidates, the test will be too hard, for others it will be too easy, and, for some, the topics that happen to be chosen will be either very familiar or very unfamiliar: in all cases, the actual proficiency of the candidates as it would appear in real-life usage will probably not be accurately identified.

It is also noteworthy that IELTS was designed specifically to measure the English language skills of candidates intending to study in academic or training contexts in English-speaking countries and the test specifications and itemtypes were designed to reflect as closely as possible the sorts of language tasks encountered in such circumstances. Where the test is used for other purposes (e.g., as a measure of general language ability for immigration purposes, as a measure of proficiency in vocational contexts or, still worse, as a test of the English language ability of native speaking medical practitioners wishing to work in Britain), obviously the test becomes even



further removed from real-life and so the gap between the test and real-life is wide, i.e., authenticity is low. In addition, even though IELTS presents its results in terms of simple performance-related scales, the actual outcomes of the Listening and Reading tests are translated onto the scale with its performance descriptions, not by matching observed behaviour with the descriptors but by a statistical or distributional process, i.e., the sub-tests are statistically matched for difficulty with previous versions of the test and cut-off scores are assigned for each proficiency level in order to obtain the same distribution of results as has been established over the life of IELTS.

Nevertheless, IELTS narrows the gap to real-life language usage, i.e., to authenticity, considerably more than does TOEFL with its analytic approach to test design. In TOEFL, the actual itemtypes are further removed from real-life language use than are those in IELTS. TOEFL is analytic, focuses on small elements of the language or, at best, on small language tasks, rather than on the whole language as it is used in real-life. Even in Listening, the focus is on small exchanges that necessarily lack, for instance, the extended context and discourse structuring that occurs in real-life listening activities while the use that is made of texts tends to be very different in real-life from the sorts of responses made to small oral exchanges that occur in the TOEFL Listening test or even in the more extended IELTS Listening tasks. In addition, TOEFL's use of a numerical scale unrelated to behavioural or performance descriptions removes the test even further from real-life language use or the ability of end-users to interpret the results in terms of candidates' likely ability to use the language in real-life academic (or other) situations.

In order to bridge the gaps that these limitations in test design impose, other approaches are adopted. Commonly, the results on a new test are compared with the results on other tests whose relationship with real-life is supposedly known. Cut-off scores may also be allocated as a result of experience with candidates who have previously taken the test. So, with TOEFL, it has become known over the years that students require a certain score in order to cope with academic study and that score or those scores then acquire some vague (and largely unreliable) relationship to real-life abilities. This is, itself, at best an unreliable procedure but it is made still more unreliable and the test further removed from authenticity, by virtue of the fact that it has been demonstrated over the years that TOEFL test scores can be increased by practising the itemtypes that are used and, indeed, TOEFL cram schools where the focus is not on developing real-life language ability but on how to "do" TOEFL tests are "big business" in all countries where TOEFL is taken.

More systematically and formally, predictive validity studies may be undertaken to relate test results to real-life performance. However, predictive validity studies are notoriously difficult to structure and to analyse not least because, at best, language tests measure ability in language whereas performance in real-life situations invariably involve many other variables which, even in the best designed predictive validity studies, are impossible to control or to measure accurately. Consequently, it is difficult, if not impossible, to reliably relate language test results to real-life abilities through predictive studies no matter how adequate the test design might be. What

studies there are tend to illustrate the relatively low correlation between results on tests such as IELTS and TOEFL and subsequent success rates in academic study, for instance.

In an internal (confidential and, hence, unpublished) study at Griffith University, the success rate in first year of more than 2300 overseas students was related to their entry path via bridging programmes without a language proficiency test, and via TOEFL, IELTS and a direct assessment scheme, the ISLPR [Sefton and Wylie 2002]. Though the differing global availability of the tests and the consequent intrusion of cultural differences made interpretation of the results difficult, the success rate of students who entered via the ISLPR, an observational approach to proficiency assessment which attempts to focus more closely on real-life language performance, seemed to be somewhat higher than for students who entered via other pathways.

There have been a number of predictive validity studies of IELTS in its intended use to assess the adequacy of overseas students' English skills for university studies. These have taken many forms but the more adequate tend to move away from merely statistical correlations to try to establish, often through introspection, the contribution of language ability to the students' subsequent academic performance or failure and the extent to which they themselves feel that IELTS had identified their real ability. In a study by de Prada and McVeigh [2000], students were asked the extent to which they felt, now that they were in university programmes, that their English was adequate and the extent to which IELTS had accurately identified their language skills. No student expressed faith in IELTS as a measure of their English ability, most commented that they had believed that, if they scored 6.5 (the level many institutions set as a pre-requisite to graduate study), they would be able to cope with university study but they found that they couldn't write an essay, speak, or keep up with lectures. In retrospect, many students felt that they would have been better served to have focussed more on EAP Study (i.e., practising the sorts of tasks they would encounter in their academic programme) than to strive to reach the specified IELTS level. The study summarised the results in these words:

*... [students] are aware that there are major differences between items on the IELTS test and the academic requirements of a university course. ...*

*... they suggest further that an IELTS result alone does not indicate readiness for tertiary study [de Prada and McVeigh 2000: 152]*

Rosen examined the success of students who entered through a particular preparation programme at Monash University. In relation to IELTS, Rosen identified the limitations that exist with IELTS, despite the fact that it seems to bridge the gap to real-life language performance more adequately than TOEFL and other analytic tests, when she stated:

*There is no comparison between IELTS writing – 150 and 250 word tasks – and the reports, the literature reviews and the writing*

*assignments of [the Monash preparation programme]. ... in the IELTS writing there is no requirement of referencing, no requirement to read relevant discipline-specific academic texts, and no requirement to write a sustained, well-mapped and integrated academic paper. ... A student could reasonably do well in .. [the IELTS] tasks and still be quite unprepared for university study. Similarly, the IELTS speaking test is an interview – an excellent indicator of everyday proficiency and even ability to speak about one’s future plans – but lacking in the focussed academic requirements of tutorial and seminar participation and presentation. The IELTS Listening and Reading tests vary in range of topics and may be good indicators of general skill, but certainly cannot predict ability to follow lectures or discipline-specific reading material. Research has shown that students often do badly at a one-off test. Each test is different and a student may strike a topic which is familiar on one occasion and one which is very unfamiliar on another. [Rosen 1998: 191 – 192]*

Clearly, no matter how carefully they might have been designed and, like IELTS, no matter how thoroughly the designers might have attempted to match the itemtypes with the tasks to be performed in subsequent real-life language use situations, most tests have difficulty in predicting candidates’ subsequent performance, even when the focus of that evaluation is on the candidates’ ability to use the language. Some of the reasons for the failure of tests to bridge the gap have been mentioned previously. The reasons (all of which point to a lack of authenticity and a serious gap between the test and real-life language use) include *inter alia*:

- the impoverished contexts within which the language exists and operates in traditional approaches to test design, yet language is known to be heavily situation-dependent;
- the limited range of situations which it is possible to include in tests that are largely constrained by pencil-and-paper presentation and response modes;
- the disparity between the test situations and real-life authentic language situations;
- the pre-determined and limited content of tests that have been statistically “standardised” in order to ensure statistical validity and reliability: the content and language elicitation modes of such tests are commonly limited to those that can be controlled and adjusted in accordance with statistical requirements;
- the inability of pre-determined tests to match individual candidates’ needs, interests, experiences, proficiency levels, and other personal characteristics, i.e., the lack of adaptiveness of such tests, and hence their inability to accurately identify the particular skills or lack of them that individual candidates present; and

- the failure of most tests to present their results in ways that allow their ready interpretation in terms of candidates' real-life language ability with the result that, for most tests, there is a double gap between the test and real-life ability: first, the gap between the test items and real-life use of the language and, second, the gap that the end-user has to bridge between how the results are expressed (e.g., in a numerical score or an abbreviated behavioural description) and the language demands of real-life language use situations.

One of the reasons why language proficiency tests do not necessarily correlate with real-life performance and why predictive validity studies are generally unsatisfactory is that there are many other variables besides language proficiency that determine the success of a learner in using the language in real-life situations. One approach to overcoming the limitations of language assessment alone has been the emergence of the concept of competencies, specifically language competencies particularly for the workplace. Since the focus of this paper is on language proficiency assessment, competencies will not be discussed here though they have been in other papers [e.g. Ingram 2000 c].

#### **IV The *International Second Language Proficiency Ratings* (ISLPR)**

Largely in response to the limitations of language tests just discussed, the present writer and Elaine Wylie developed the *International Second Language Proficiency Ratings* (ISLPR) in the late 1970s, they have researched them ever since, they have developed versions for many different purposes, and they are now about to launch an approach to assessment using the ISLPR which enables it to be readily administered worldwide and, potentially, in situations that will enable candidates' language ability to be observed in virtual situations that approximate more nearly to real-life language use than occurs in any other approach to assessment. At this point where, it is hoped, the level of authenticity is about to be raised substantially, it is worth remembering the steps that had been taken previously to make ISLPR assessment as authentic as practicable.

Motivation to develop the ISLPR came from three sources, in all of which, finding ways to bridge the gap between language tests and evidence of real-life language ability were critical. First, during research in the mid-1970s [Ingram 1978], the present writer wanted to be able to specify the foreign language skills students brought to university after five years of secondary school foreign language study but he found that the results on matriculation examinations in Britain and Australia gave no indication of what students' practical language skills were, i.e., the gap between the examinations, the examiners' reports, and evidence of the learners' practical language skills was unbridgeable. Second, about the same time, the present writer was involved in developing new national ESL courses for recently arrived migrants in

Australia and the project team wanted to be able to specify the sorts of language abilities that learners had on entering or exiting from their ESL classes, they wanted to be able to specify the language skills that the new courses should aim to develop, and they wanted a broad framework within which to systematically develop a series of courses progressively graduated through proficiency levels. Third, the present writer was also involved in advising on the development and assessment of new “proficiency-based” foreign language programmes for Queensland secondary schools. As a result of these three necessities, the present writer, together with Elaine Wylie, developed the *Australian Second Language Proficiency Ratings (ASLPR)* subsequently re-named the *International Second Language Proficiency Ratings (ISLPR)* in recognition of the scale's growing use internationally [Ingram and Wylie 1979/1999]. In the ISLPR and its approach to assessing proficiency, attempts are made to bridge the gap between assessment and real-life language ability, and to elicit and measure real-life language performance. Language proficiency is considered, in the ISLPR approach to assessment, as encompassing the tasks that learners can carry out and how they are carried out and, as far as possible in a test situation, the focus of the assessment is on real-life tasks.

Since the ISLPR is well known in Australia and has, in any case, been described in detail in other papers, it won't be described again here; rather the focus of this paper is on what has been done to bridge the gap between language proficiency testing and real-life language use, i.e., the focus is on what steps have been taken to enhance the authenticity of the assessment procedure and the scale itself.

First, the ISLPR is a scale that essentially describes how a second or foreign language develops from zero to native-like proficiency. It seeks to capture the intuitions that speakers have about language ability but, through the descriptors, to constrain the intuitive judgements that people make in order to achieve some commonality in the judgements and in how the levels are stated.

Second, the ISLPR provides performance descriptions couched in terms of the practical tasks that learners can carry out and how they carry them out along the continuum from zero proficiency to native-like proficiency, i.e., its focus is on the actual tasks that learners can carry out and how they carry them out, not on artificially contrived “itemtypes” that generally bear little resemblance to what people do in real life with language.

Third, the ISLPR seeks to describe the way in which a language learned as a second or foreign language develops so that, not only are the tasks identified and used authentic but the developmental path described is as authentic as current research into language development can make it. Since the psycholinguistic evidence is that different languages develop over broadly the same developmental path, the scale is readily applicable to any language even though it was considered useful for illustrative purposes to provide a number of versions of the scale in other languages beside English.

Fourth, language is situation-dependent and pre-determined tests often break down because a learner has simply not experienced a particular situation or the test is set at either too high a proficiency level or too low a level for the learners being assessed. Consequently they are unable to demonstrate their actual ability and, again, the authenticity of the test suffers. The ISLPR approach, on the other hand, constitutes a highly adaptive test; because different languages develop similarly in different specific purpose registers, the basic scale and the basic assessment procedure can readily be applied to the assessment of any language, for any specified purpose, in other words, the ISLPR and its characteristic application constitute a highly adaptive approach to proficiency assessment.

Fifth, in many ways, the best informed assessors of peoples' language ability are the people themselves, i.e., since they know intimately what they can or cannot do in a language, their assessment of their own ability should be more authentic than judgements made by others so long as they have had the experience of using the language for real-life purposes and so long as they approach the task of self-assessment honestly. In recent years, several different versions of the ISLPR, of differing levels of complexity, have, therefore, been produced for self-assessment purposes.

Sixth, one effect of the long and varied research effort that has surrounded the ISLPR as different versions have been produced, as it has been used for many different purposes in many different contexts is that the authenticity of the descriptors has been continually re-examined, the basic scale and its various versions have been continually re-evaluated, the reliability of the assessment procedures have been tested, and the assumptions underlying the scale and the assessment procedures have been continually tested. This has been done both through formal evaluations and observationally in the course of extensive usage [see Ingram 1985a, Lee 1993].

Seventh, few scales have adopted such a long and detailed process for their development and on-going re-development and validation as has the ISLPR. Even the most elaborated scales are, at best, partial descriptions of how a language develops and their validity, i.e., the authenticity of their descriptors, depends heavily on the processes by which they have been developed. In the case of the ISLPR and its various versions, the original and on-going development process consists of the following:

1. A notion of proficiency related to the language tasks that learners can carry out at different proficiency levels and how they are carried out was adopted and evolved as the scale developed.
2. Drawing on the intuitions and experience of the authors and others (including the authors of other scales) to provide insight, Ingram and Wylie sketched descriptions of language behaviour and how it develops.
3. The initial descriptors were then tested, elaborated and refined in interviews with learners throughout the proficiency span. The aim of

these interviews was to elicit the features of the learners' language so as to evaluate whether the evolving descriptors were comprehensive, coherent and consistent, i.e. were they authentic descriptions of how a second or foreign language develops, the tasks that can be carried out at various proficiency levels, and how they are carried out. This process has continued over the years in the course of developing and using the different versions of the scale so that the latest versions and editions are the product of empirical studies involving many thousands of learners of English and other languages, including their use in specific purpose contexts.

4. At the same time, the emerging scale was compared with evidence from psycholinguistics to assess whether it was compatible with those general findings.
5. The scale has, several times, been formally trialled using adult and adolescent learners, especially of English but also of other languages [e.g., Ingram 1985a]. This formal trialling essentially assumed that, if the series of descriptors making up the scale really did reflect second or foreign language development, if they described features of the language that generally do co-occur, and if they were comprehensible and manageable, teachers trained to use the scale would be able to interpret the descriptors consistently and apply them reliably. This has always proved to be the case though the authors insist that the reliability of the assessment system depends heavily on the quality of the training of the assessors, their regular calibrating and re-calibrating of their interpretation of the scale, the regular monitoring of assessors' interview technique, and the moderation of their ratings.
6. Statistical processing has also been used to check the scale and the validity and reliability of the ratings [e.g., Ingram 1985a, Lee 1993]. In one study, Lee analysed the results of more than 300 ESL assessments of each of the four macroskills to establish whether the levels in the scale actually do represent a progression from zero to 5 along a common dimension, whether the four macroskills do form a reliable measurement variable, and whether the ordering provided the basis for construct validity. In summary, Lee concluded that both the ISLPR and the assessment procedure had a high degree of validity and reliability [Lee 1993].

The standard means by which the ISLPR is used to measure proficiency is in a face-to-face interview in which each learner's language is elicited and matched against the scale's behavioural descriptions. In this approach, the actual items used are less important than the fact that the learner's real language behaviour is elicited for observation and matching against the scale descriptors. Because the ISLPR focuses directly on the learners' language behaviour in practical use of the language and because the assessment

procedures seek to elicit such language, the gap between the assessment and its results, on the one hand, and real-life use of the language, on the other, is much less than in other approaches to language testing. In that sense, the language that occurs and is measured in the ISLPR and its assessment procedure is more authentic than in other approaches.

In principle, the ISLPR can be used to assess learners' language proficiency as they use the language in real life, in the course of normal language use whether in academic or vocational contexts or in everyday life, i.e., with maximum authenticity. However, the difficulties with such direct observation, are that it is very time consuming and quite impractical where a large number of learners are to be assessed. In addition, unless one spends a great deal of time in such observation, one can never be certain that the full extent of a learner's strengths or weaknesses has been observed. For these reasons, the ISLPR is normally applied in a face-to-face interview in which the interviewer sets out to elicit the candidate's maximum language ability and matches the observed language behaviour against the scale descriptors.

As noted earlier, one of the essential features of the ISLPR that contributes greatly to its authenticity is its adaptiveness. Unlike other tests, the assessment procedure does not rely on a pre-determined set of standardised items but rather on a trained assessor's ability to elicit the candidate's maximum language ability. For that purpose, assessors adjust items according to each candidate's proficiency level so as to ensure that the maximum ability is observed, with the tasks used being neither too easy for the candidate's level nor too difficult. In addition, since language is situation-dependent and familiarity or unfamiliarity with a situation strongly influences a person's ability to perform linguistically, assessors are able to vary situations and tasks so as to see whether it is the candidate's proficiency that is causing the problem or the particular situation or particular task. In addition, because the scale and its application do not rely on a pre-determined set of standardised items, its use can readily be adapted for application in a variety of vocational or academic situations and in a variety of language domains or genres with the only limitation being what tasks, domains or genres can be utilised in an interview setting. Thus, the ISLPR can readily be used for specific purpose assessment, especially for vocational proficiency assessment. In addition, since the psycholinguistic evidence is that all languages develop over similar paths, the ISLPR can be used with any language being learned as a second or foreign language [see Appendix One]. In other words, the ISLPR supports a highly adaptive assessment procedure which makes it better able than other approaches to adapt in order to match the needs of different learners, using the language in different domains and genres, at different proficiency levels, and with different vocational, academic or other usage interests.

## **V THE ISLPR GLOBAL**

Despite the many advantages of the ISLPR as a measure of practical language ability and its ready interpretability in terms of real-life language



ability, the ISLPR in its present application procedures has certain limitations. First, the time required for interviews may seem to be substantial, being, on average, 30 to 45 minutes to rate speaking, listening and reading, with writing being administered separately. However, in reality, when one considers the time taken to develop and standardise tests such as IELTS, the time to administer listening, reading and writing and then to mark the scripts, and the time taken for a speaking interview, the comparison with other tests is by no means unfavourable to ISLPR. Second, ISLPR is administered in a face-to-face interview and this limits its use on a global scale, for instance, to assess the proficiency of overseas students or immigrants wishing to travel to English-speaking countries. Third, whereas other tests depend on prior standardisation for their validity and reliability, the ISLPR assessment procedure relies on the use of trained assessors whose performance, desirably, can be monitored and moderated. Though other tests rely similarly on trained assessors for some components (for Speaking and Writing in the case of IELTS, for example), it is an even more important requirement for the ISLPR where all four macroskills are assessed by elicitation and direct observation of language behaviour and its matching against the scale descriptors. Consequently, its present administration procedure is not appropriate for use on a global scale when only minimal training and supervision of assessors are possible. Fourth, the range of situations in an interview room is very limited and is largely restricted to what can be achieved in a face-to-face conversation: this has serious implications for the authenticity of the language that can be elicited and for the gap that exists between the language elicited and rated and the situational demands of real-life language use.

To address these limitations, the authors of the ISLPR are about to release a new assessment scheme, currently known as ISLPR Global. ISLPR Global will continue to use the ISLPR scale as the basis for assessment, certification and interpretation of results but, instead of the assessment being conducted face-to-face in an interview room, it will be conducted on-line using appropriate hardware and adapted software. Initial trials have shown that available hardware and software are suitable with relatively minor adaptations and with very little modification to standard interview procedures. This new approach has many advantages, including these:

1. On-line administration enables the ISLPR to be used worldwide and so greatly increases the range of candidates who can take it.
2. The ISLPR Global will be administered from a central location or a small number of administration centres in different time zones together with very basic test centres in appropriate locations or in cooperating institutions. Hence there is no need for a network of elaborate administration centres employing trained staff such as IELTS requires.
3. Because ISLPR Global will be administered from a small number of centres using thoroughly trained professional staff, strict quality control can be applied to both the interviews and the ratings.

4. The technology that will be used will allow a high level of personal security to occur. In other “high stakes” tests such as IELTS or TOEFL, security (including candidate identification) is a perpetually difficult issue. With the ISLPR, item security is not an issue but, in addition, ISLPR Global will enable photographs, voice prints, or even eye scans to be built into the security and certification systems, thus ensuring that, when certificates are issued, the possibility of identification fraud occurring will be minimised.

In other words, ISLPR Global builds on the advantages of authenticity found in the current approach to the use of the ISLPR but goes further, narrows the gap between the test and real-life language use, and also makes the procedure accessible worldwide. In future, that gap will be narrowed still further since, once the ISLPR Global has been established on-line, virtual reality will be used to create virtual situations within which the assessor can observe the candidates’ real-life (or virtual real-life) use of the language. Initial trials have demonstrated that it is possible to create on the computer screen and to transmit globally scenarios in which candidates can respond as they would do in real-life and the assessor can observe that interaction across the web and rate it as if he or she is observing the candidates in the course of their normal use of the language in everyday life or in the particular vocational or academic context in which they wish to use the language. In other words, using the resources of modern technology, the internet and virtual reality, the gap between proficiency assessments and real life can be narrowed almost to non-existence. In fact, the ISLPR Global approach has advantages over real-life observation because the assessors will enter into the virtual scenarios to interact with the candidates as normal participants in the language situation but, in the dual role of assessor and participant, they will also have some control over what occurs to the extent that, as with the current interview approach with the ISLPR, they will be able to modify the interaction in order to match it to the candidate’s needs, proficiency level, and the other variables identified earlier.

An ISLPR Global prototype has been developed and trialled, the results are being evaluated, and more extensive trials are currently being conducted; the corporate structure within which it will be marketed is being established, and we anticipate that ISLPR Global will be commercially available in the course of 2003.

## **VI CONCLUSION**

This paper has focussed on the gap that exists between tests of language proficiency and real-life language performance and the attempts that have been made over time to move towards more authenticity in language proficiency assessment. To some extent, that gap will always exist because language is only a part of any activity and a person’s performance in that activity is influenced to a greater or lesser degree by other variables. Nevertheless, since language is situation-dependent, the more the tasks and

contexts in which the language is tested resemble those of real-life, the more accurately is the language test likely to predict how the candidate will cope, at least linguistically, with real-life activities. Along the fairly simplistic continuum of indirect, semi-direct and direct approaches to proficiency assessment, the ISLPR tries to focus, more strongly than most assessment procedures, on candidates' practical language ability and their performance in real-life language activities. The geographical limitations of the ISLPR imposed by the need for candidates to be able to access a centre for a face-to-face interview and the limitations imposed by the relationships and situations that can be reproduced in an interview room are further reduced by the ISLPR Global as we search for ways to bridge the gap between language tests and real-life language performance and move towards more authenticity in language testing.

## APPENDIX ONE

### The International Second Language Proficiency Ratings (ISLPR)

The International Second Language Proficiency Ratings (ISLPR) were initially developed by Elaine Wylie and D. E. Ingram in 1978 and first published in January 1979 as the Australian Second Language Proficiency Ratings (ASLPR). The basic scale is designed to measure general proficiency or practical language skills in real-life language contexts in second or foreign language learners. The scale consists of 12 levels from zero to native-like, numbered from zero to 5 as shown below. The scale is presented in three columns: the first column provides a “General Description of Language Behaviour” and is almost identical across all versions of the scale, the second provides “Examples of Language Behaviour” and is specific to the particular version of the scale, and the third is a “Comment” column that explains, gives definitions and draws attention to critical features of the descriptor or level.

The outcome of using the ISLPR for the assessment of a second or foreign language learners’ proficiency is a profile showing the rating for each macroskill separately, e.g., S:3, L:3+, R:2+, W:2. The levels in each of Speaking, Listening, Reading and Writing are identified with a number and a short descriptive title as follows:

0	Zero Proficiency	e.g., S:0, L:0, R:0, W:0
0+	Formulaic Proficiency	
1-	Minimum ‘Creative’ Proficiency	
1	Basic Transactional Proficiency	
1+	Transactional Proficiency	e.g., S:1+, L:1+, R:1+, W:1+
2	Basic Social Proficiency	
2+	Social Proficiency	
3	Basic ‘Vocational’ Proficiency	
3+	Basic ‘Vocational’ Proficiency Plus	
4	‘Vocational’ Proficiency	e.g., S:4, L:4, R:4, W:4
4+	Advanced ‘Vocational’ Proficiency	
5	Native-like Proficiency	

Ingram and Wylie have worked on the ISLPR virtually continuously since 1978. It has been formally trialled in a number of different contexts and has been applied and re-developed in a number of different versions listed below. It is now the most widely used instrument for the specification of proficiency levels in Australia, it is used in many places around the world, and it has significantly influenced proficiency scale development elsewhere (e.g., the *ACTFL Guidelines*).

The ISLPR currently exists in the following versions:

- *The Australian Second Language Proficiency Ratings - Master General Proficiency Version (English Examples)*. Brisbane: Centre

for Applied Linguistics and Languages, Griffith University, 1979/1995/1999. ISBN 0 86857 814 2. Co-author Elaine Wylie

- *The Australian Second Language Proficiency Ratings - General Proficiency Version for English*. Brisbane: Centre for Applied Linguistics and Languages, Griffith University, 1979/1985/1995/1999. ISBN 0 86857 815 0. Co-author Elaine Wylie
- *The Australian Second Language Proficiency Ratings - Version for Teachers of Indonesian*. Brisbane: Centre for Applied Linguistics and Languages, Griffith University, 1996. ISBN 0 86857 819 3. Co-authors Elaine Wylie and Geoff Woollams
- *The Australian Second Language Proficiency Ratings - General Proficiency Version for Indonesian*. Brisbane: Centre for Applied Linguistics and Languages, Griffith University, 1995. ISBN 0 86857 816 9. Co-authors Elaine Wylie and Geoff Woollams
- *The Australian Second Language Proficiency Ratings - Version for Second Language Teachers*. Brisbane: Centre for Applied Linguistics and Languages, Griffith University, 1995. ISBN 0 86857 817 7. Co-author Elaine Wylie
- *The Australian Second Language Proficiency Ratings - English for Business and Commerce Version*. Brisbane: Centre for Applied Linguistics and Languages, Griffith University, 1995. Co-authors Elaine Wylie and Hilda Maclean
- *The Australian Second Language Proficiency Ratings - English for Engineering Purposes Version*. Brisbane: Centre for Applied Linguistics and Languages, Griffith University, 1995. Co-authors Elaine Wylie and Laura Commins
- *The Australian Second Language Proficiency Ratings - English for Academic Purposes Version*. Brisbane: Centre for Applied Linguistics and Languages, Griffith University, 1995. Co-authors Elaine Wylie and Catherine Hudson
- *The Australian Second Language Proficiency Ratings - Version for Japanese*. Brisbane: Centre for Applied Linguistics and Languages, Griffith University, 1994. Co-authors Elaine Wylie and Peter Grainger
- *The Australian Second Language Proficiency Ratings - Version for French*. mimeograph. 1981. Co-authors Elaine Wylie and Edwige Coulin

- *The Australian Second Language Proficiency Ratings - Version for Italian*. mimeograph. 1981. Co-authors Elaine Wylie and Carlo Zincone
- Various self-assessment versions ranging from very short, simplified versions administered by telephone to computer-based versions, and versions used with language teachers.

The ISLPR was originally named the Australian Second Language Proficiency Ratings (ASLPR). In 1997, the ASLPR was re-named the International Second Language Proficiency Ratings (ISLPR) in reflection of its growing international usage and to emphasize its appropriateness to any country, to any language, and, not least, to any variety of English learned as a second or foreign language.

## REFERENCES

- Alderson, J. Charles and Arthur Hughes (eds.). 1981. *Issues in Language Testing, ELT Documents No. 111*. London: The British Council.
- Bachman, L. F. 1990. *Fundamental Considerations in Language Testing*. London: Oxford University Press.
- Bachman, L. F. and A. Palmer. 1996. *Language Testing in Practice*. London: Oxford University Press.
- Canale, M. 1988. "The measurement of communicative competence". In *Annual Review of Applied Linguistics*. No. 8: 67 – 84.
- Canale, M. And M. Swain. 1981. "A theoretical framework for communicative competence". In Palmer, Groot and Trosper 1981.
- Canale, M. And M. Swain. 1980. "Theoretical bases of communicative approaches to second language teaching and testing". In *Applied Linguistics*, Vol. 1: 1 = 47.
- Clark, John L.D. (ed.). 1978. *Direct Testing of Speaking Proficiency: Theory and Application*. Princeton, N.J.: Educational Testing Service.
- Coates, Sharon *et al.* 1995. *National Reporting System*. Canberra/Brisbane: Commonwealth of Australia and the Australian National Training Authority.
- Council for Cultural Cooperation, Council of Europe. 1973. *Systems Development in Adult Language Learning: A European Unit/Credit System for Modern Language Learning by Adults*. Strasbourg: Council for Cultural Cooperation, Council of Europe.
- Cummins, J. 1979. "Cognitive/academic language proficiency, linguistics independence, the optimum age question and some other matters." In *Working Papers in Bilingualism*, Vol 19: 97 – 205.
- Cummins, J. 1983. "Language proficiency and academic achievement". In Oller 1983: 108 - 126.
- de Prada, Kathi and Lynn McVeigh. 2000. "Struggle, survival and success: International students' beliefs about study in Australia". Paper to the Thirteenth Annual English Australia Education Conference, Fremantle, 2000. Reprinted in *Proceedings of the 13<sup>th</sup> Annual EA Education Conference, 2000*. Sydney: English Australia: 147 – 158.
- Education Committee, Council for Cultural Cooperation, Council of Europe. 1996. *Common European Framework of Reference for Language*

*Learning and Teaching, Draft 1 of a Framework Proposal*. Strasbourg: Council of Europe.

Hyltenstam, K. and M. Pienemann. 1985. *Modelling and Assessing Second Language Acquisition*. Clevedon: Multilingual Matters.

Ingram, D. E. 2000/2001. "The Enigma of Cross-Cultural Attitudes in Language Teaching, Parts 1 and 2". In *Babel*, Vol. 36, No. 2, Spring 2001: 12 - 18, and *Babel*, Vol. 36, No. 3, Autumn 2002: 17 - 22, 37 - 38.

Ingram, D. E. 2000. "Language proficiency assessment and its application to communicative language teaching". Paper to the Tohoku Chapter of the Japan Association of College English Teachers, Akita University, Akita, Japan, 19 August, 2000. Reprinted in *Akita English Studies (Akita eigo eibungaku): Trans-Equator Exchanges: A Collection of Academic Papers in Honour of Professor David Ingram*. Akita, Japan: Akita Association of English Studies: 2001: 21 – 59.

Ingram, D. E. 2000a. "The Relevance of EFL education to the fostering of more positive cross-cultural attitudes". Paper to Junior and Senior High School teachers of English, Akita University and Akita Prefectural Board of Education, Akita University, 21 August, 2000. Reprinted in *Akita English Studies (Akita eigo eibungaku): Trans-Equator Exchanges: A Collection of Academic Papers in Honour of Professor David Ingram*. Akita, Japan: Akita Association of English Studies: 2001: 60 – 99.

Ingram, D. E. 2000b. "Devising a test of academic writing for university entrants". Paper to the December 2000 Conference of the Applied Linguistics Association of Korea, Seoul, Korea, 9 December, 2000.

Ingram, D. E. 2000c. "Proficiency, Competencies and Assessment". Paper to the December 2000 Conference of the Applied Linguistics Association of Korea, *New Horizons in Discourse Competence*, University of Seoul, Seoul, Korea, 9 December, 2000.

Ingram, D. E. 1999. "Cross-Cultural Attitudes in Foreign Language Programmes". In *PASAA: A Journal of Language Teaching and Learning in Thailand*, Vol. 29, December 1999, pp. 1 – 32.

Ingram, D. E. 1999a. "Cross-Cultural Attitudes as a Goal of Language Teaching in the Global Context". Keynote paper to the RELC Seminar 1999, *Language in the Global Context: Implications for the Language Classroom*, Regional Language Centre, Singapore, 18 – 22 April, 1999 and to the Post-RELC Seminar, Chulalongkorn University, Thailand, 26 – 27 April, 1999. Reprinted in Ho Wah Kam and Ward 2000: 104 – 163. Reproduced in the *ERIC Clearinghouse on Languages and Linguistics* (1999), ED430 398/FL025 833.



- Ingram, D. E. 1985. "Assessing proficiency: an overview on some aspects of testing". In Hyltenstam and Pienemann 1985: 215 - 276.
- Ingram, D. E. 1985a. *Report on the Formal Trialling of the Australian [now International] Second Language Proficiency Ratings (ASLPR, now ISLPR)*. Canberra: Australian Government Publishing Service/Department of Immigration and Ethnic Affairs, 1984.
- Ingram, D. E. 1978. *An Applied Linguistic Study of Advanced Language Learning*. Thesis for the degree of Doctor of Philosophy, University of Essex, Colchester, England.
- Ingram, D. E. and Elaine Wylie. 1991. "Developing proficiency scales for communicative assessment". In *Language and Language Education: Working Papers of the National Languages Institute of Australia*, Vol. 1, No. 1: 31 - 60. [ERIC FL019252/ED342209]
- Ingram, D. E., Shirley O'Neill and Kerry Townley-O'Neill. 1999. "The Effect of Language Learning on Cross-Cultural Attitudes". Report of a project funded under the Griffith University Research Grants. Mimeograph, 1999.
- Ingram, D. E. and Elaine Wylie. 1979/1999. *The International Second Language Proficiency Ratings*. Brisbane: Centre for Applied Linguistics and Languages, Griffith University.<sup>1</sup>
- Languages Lead Body. 1992. "Language Standards: Draft Framework". In *Languages Lead Body News*, Issue Two, August 1992: 3.
- Languages Lead Body. 1992a. *National Standards for Modern Foreign Languages: Units of Competence and Assessment Guidance*. London: Languages Lead Body.
- Lee, Tony. 1993. "A Many-Faceted Rasch Analysis of ASLPR ratings". Nathan: Centre for Applied Linguistics and Languages, Griffith University. mimeograph.
- Little, William, H. W. Fowler, and J. Coulson, revised and edited by C. T. Onions. 1967. *The Shorter Oxford English Dictionary on Historical Principles, Third Edition*. Oxford: Clarendon Press.
- Messick, S. 1994. "The interplay of evidence and consequences in the validation of performance assessments". In *Educational Researcher*, Vol. 23, No. 2, March: 13 - 23.
- National Training Board. 1991. *National Competency Standards: Policy and Guidelines*. Canberra: The National Training Board.

---

<sup>1</sup> Note in Appendix One reference to the name change from *Australian Second Language Proficiency Ratings (ASLPR)* to *International Second Language Proficiency Ratings*.

- Oller, J. W. (ed.). 1983. *Issues in Language Testing Research*. Rowley, Mass.: Newbury House.
- O'Neill, Shirley. 2000. *Languages in Use in Tourism and Hospitality: Language Competencies for Tourism and Hospitality Training Packages 2000. Report to Tourism Training Australia*. Brisbane: Centre for Applied Linguistics and Languages. Mimeograph.
- Palmer, A. S., P. J. Groot, and G. A. Trosper (eds.). 1981. *The Construct Validation of Tests of Communicative Competence*. Washington DC: TESOL.
- Rosen, Ruth. 1998. "Procedures, papers and forms in ELICOS: Academic transition into university culture or, what are we teaching our students?". Paper to the Eleventh Education Conference of the ELICOS Australia Association, Melbourne, 1998. Reprinted in *Proceedings of the 11<sup>th</sup> Annual EA Education Conference, 1998*. Sydney: English Australia: 185 – 193.
- Sefton, Jane and Elaine Wylie. 2002. "Maximising the satisfaction of international students and minimising their attrition: Report on a Griffith University quality enhancement project". Confidential mimeograph.
- Vollmer, Helmut J. 1981. "Why are we interested in 'General Language Proficiency'?". In Alderson and Hughes 1981: 152 - 175.
- Wylie, Elaine and D. E. Ingram. 1995. *Minimum Skills/Competencies needed by LOTE Teachers, Part One: Language Proficiency: International Second Language Proficiency Ratings (ISLPR): Version for Second Language Teachers*.

## **APPENDIX TWO**

**Excerpts from the *International Second Language Proficiency Ratings (ISLPR)***

GENERAL DESCRIPTION OF LANGUAGE BEHAVIOUR	EXAMPLES OF LANGUAGE BEHAVIOUR VERSION FOR ENGLISH	COMMENT See also NOTES and GLOSSARY
<p><b>S:4 'VOCATIONAL' PROFICIENCY</b></p> <p><b>Able to perform very effectively in almost all situations pertinent to social and community life and everyday commerce and recreation, and generally in almost all situations pertinent to own 'vocational' fields.</b> The learner conveys his/her desired meaning in straightforward conversations, interviews, discussions and monologues with virtually the same fluency, precision and complexity, and to virtually the same depth as do native speakers of the same sociocultural variety. Usually needs no more support from the context to communicate than a native speaker does. The learner may for a short time in some situations produce language which is indistinguishable from that of native-speaking peers. In very complex texts, however, has less control of the argument than such peers do. Rhetorical structure in such texts may at times be non-standard, particularly in less familiar situation types. No grammatical structures are missing from the learner's repertoire; errors of grammar are fairly rare, and are often picked up in a monitoring process and corrected immediately. Errors never interfere with understanding, although there may be occasional lapses in the use of cohesive devices (typically when the referent is well separated) which may momentarily distract listeners. Vocabulary range is close to that of a similarly educated native speaker and allows for some stylistic variation for aesthetic purposes (e.g. for euphony). High- and medium-frequency colloquial and idiomatic forms are secure but some non- or misuse of other items occurs. Is secure in the use of borrowings (from other languages or other varieties) that are in high- and medium-frequency use in the speech of native-speaking peers. There may be an obvious 'foreign' accent, but this in no way impedes comprehension by a native speaker of the same or a similar variety. Has considerable sensitivity to register requirements. There are, however, occasional minor lapses in terms of appropriateness of expression (e.g. inappropriate influences of written text) and, particularly in less familiar situation types, in terms of what meanings may be (directly) expressed. Such lapses do not confuse interlocutors, and do not generally <i>per se</i> offend native-speaking peers.</p>	<p>Can take on <i>community</i> roles which are linguistically demanding (e.g. organising and leading a delegation of parents to lobby a parliamentary representative about a problem at school).</p> <p>Copes with all spoken aspects of '<i>vocational</i>' roles requiring specialised skills, except a very select few where the highest level of mastery of the linguistic and cultural systems is essential (e.g. top-level diplomatic negotiating or interpreting into English). In many educational systems, learners at this level are considered sufficiently proficient to be responsible for teaching English (and other areas of the curriculum) to native-speaking children.</p> <p>In <i>work</i> situations, can make a product presentation and respond in depth to technical questions. Can negotiate contracts and other important agreements. Can handle complicated complaints from a customer or colleague about a product or service. Can tailor language to an audience of a very different level of technical knowledge, sophistication or maturity (e.g. a teacher talking to young students).</p> <p>Can give a presentation at an <i>academic</i> conference or seminar, and respond in depth to questions from the audience, integrating references to handouts, visual aids or previous points made by self or others. Can convey own precise opinions in a seminar or symposium, and use modal forms effectively to temper comments about input from (native-speaking) peers.</p> <p>There are occasional localised errors in forms such as articles and prepositions, particularly when the learner is tired or under stress.</p> <p>In most straightforward situations in everyday life and own '<i>vocational</i>' field(s), can convey meaning accurately and fluently in informal consecutive interpreting from L1.</p>	<p>A key factor at this level is a high degree of mastery of the specialised language of learners' '<i>vocational</i>' field(s) (with a high degree of flexibility, permitting communication with lay people).</p> <p>The learner will perform 'very effectively' in 'almost all situations' pertinent to his/her '<i>vocational</i>' field(s) unless a major field happens to be one of the 'very select few' which feature tasks which demand the highest level of mastery of the linguistic and cultural systems (see the <i>EXAMPLES</i> column on this page).</p> <p>The range of straightforward everyday situation types in which learners can perform effectively is very close to that of native speaking peers; flexibility when confronted with new situation types is close to that of such native speakers.</p> <p>Any mismatch between what learners convey through their speaking in English (with accompanying non-verbal communication) and their intentions and self-image is rarely attributable to L2 developmental factors (but note reference to accent below).</p> <p>At this level, grammatical development is almost complete. Errors occur in complex texts and/or when the context is very distracting (for example, when significant extra-linguistic processing is required or the learner is very tired or emotional). Learners will, however, often hear and correct such errors. A high proportion of grammatical mistakes made are not systematic errors but the kinds of slips that native speakers make (see the <i>W:4 COMMENT</i> column). The lexicon is much greater than at the previous described level.</p> <p>Strength of accent will depend on individual factors (e.g. personality and musicality) and on the L1 and the age at which the learner was exposed to English. If the L1 phonological system is very different from that of English, and there was no significant exposure pre-puberty, it is likely that the accent will be fairly strong, although not, at this level, strong enough to interfere with understanding. Some learners at this level have an accent which is often associated by native speakers with learners of a different sociocultural variety.</p> <p>For comment on 'borrowings', see the <i>S:5 COMMENT</i> column.</p> <p>For comment on progress beyond this level, see the <i>L:4 COMMENT</i> column.</p>

GENERAL DESCRIPTION OF LANGUAGE BEHAVIOUR	EXAMPLES OF LANGUAGE BEHAVIOUR VERSION FOR ENGLISH	COMMENT
<p><b>L:3 BASIC 'VOCATIONAL' PROFICIENCY</b></p> <p><b>Able to perform effectively in a wide range of informal and formal situations pertinent to social and community life and everyday commerce and recreation, and in situations which are not linguistically demanding in own 'vocational' fields.</b> Provided the utterance rates are normal, the learner understands sufficiently well to participate with ease in most straightforward conversations with native speakers about everyday topics, and to participate in some semi-specialised (but not in-depth) interviews and discussions on 'vocational' topics relevant to own interests and experience. Provided the utterance rates and clarity of articulation are normal, and there is little acoustic interference, gets the gist of many conversations which s/he overhears (but does not participate in) and most radio and TV interviews on the topic types indicated. Generally follows simple straightforward monologues on such topic types (whether face-to-face or transmitted electronically) at normal rates of utterance, but has problems with complex or unfamiliar rhetorical structure. Even in texts with a relatively simple structure, may occasionally have problems if there are discourse relationships which are not clearly marked, which are marked by low-frequency cohesive devices, or in which a referent is well separated. May have problems with any particularly complex grammatical structures. Can often work out the meaning of unfamiliar words from the context, but has problems with relatively low-frequency idioms and unfamiliar borrowings and acronyms. May have problems with highly colloquial speech. May have to ask for explanation of references to unfamiliar cultural phenomena or institutions, and is likely to have significant problems with meanings associated with esoteric aspects of the culture. Has some ability to go beneath surface meaning but, when there is less support from the context, may fail to perceive the illocutionary force of less straightforward statements. Fails to perceive subtle nuances of meaning. Has significant sensitivity to register variation, although finer distinctions are not made. In straightforward situations, usually perceives the purposes, attitudes and moods of speakers. Genuine appreciation of stylistic variation for aesthetic purposes is very tentative, limited to obvious stylistic effects. Has a broad understanding of the varieties of English that are very closely related to the target variety, but often misses regional references and forms.</p>	<p>Can participate effectively in most face-to-face and phone conversations on everyday topics. Can readily take information by phone in most situations related to <i>social and community life and everyday commerce and recreation</i>. Can get the gist of radio and TV news stories on relatively abstract topics (e.g. a medical breakthrough or overseas political development) and of longer but simply-structured prepared monologues on such topics, provided they are straightforward and aimed at general audiences, and read at normal rates (approximately 180 w.p.m.). Can get the gist of straightforward radio and TV interviews on such topics, provided the speakers do not significantly and/or continually exceed 180 w.p.m. and the speech is coherent and in the target variety or a very closely related variety.</p> <p>In 'vocational' (e.g. work) situations, if utterance rates are normal, can generally follow briefings and participate in well managed meetings up to the point where complex or in-depth meanings are involved, particularly in specialised aspects of the register. When responding to complex utterances (e.g. with embedded propositions), may miss the interlocutor's point. In key situations, particularly when under stress (e.g. confrontations with angry colleagues or customers) misreading of important verbal and non-verbal signals can cause problems.</p> <p>In situations which involve registers remote from everyday language or from the language of own 'vocational' fields, and particularly when under stress (e.g. when interviewed by police or cross-examined in court) is likely to misunderstand key meanings realised by unfamiliar lexis or by structures such as modal or tensed verbs, embedded clauses or question tags.</p> <p>In <i>academic</i> (i.e. formal learning) situations, can partly follow lectures in new areas of learning if the lecturer gives comprehensive, effective 'advance organisers'. The type of concentration required tends to limit attention to the particular point being made at the expense of the flow of the argument, and effective note taking is difficult. Often fails, for example, to pick up links which are not clearly signalled to points previously made, and parentheses are likely to cause problems (both because of discourse relationships and 'throwaway' delivery). May fail to perceive the illocutionary force or personal relevance of instructions, warnings, or suggestions (e.g. about submitting drafts of assignments) which are delivered in other than the most straightforward forms. Has problems with complex interplays in seminars.</p> <p>In some uncomplicated straightforward situations in everyday life or own 'vocational' field(s), understands sufficiently accurately to undertake informal consecutive interpreting into L1.</p>	<p><i>See also NOTES and GLOSSARY</i></p> <p>Register flexibility and sensitivity feature at this level, although finer distinctions are not made. Familiarity with the topic is no longer an important factor determining the level of understanding of everyday texts; the extent of learners' English language repertoire is such that they can understand the propositional content of a high proportion of those in which the information is presented in a straightforward manner. As well, learners' level of understanding of how English-language texts are influenced by situational variables is such that they can usually perceive the speakers' purposes, attitudes and moods in straightforward texts.</p> <p>Learners at this level have serious problems, however, when information is presented fast and without discipline (as in the type of meeting referred to below) or there is significant acoustic interference.</p> <p>Limited familiarity with the target culture may still affect understanding at this level. Learners who have not been submersed in the culture for a significant period of time are likely to have problems with, for example, mass media items or academic lectures which assume significant cultural knowledge about domestic politics or other local institutions. Learners may have serious problems when there is allusion to 'peripheral' aspects of the culture or to phenomena distinctive of an unfamiliar subcultural group. Understanding is also likely to be seriously affected if learners are unfamiliar with a concept or tradition which is so fundamental to a particular group of native speakers that it is not overtly referred to. (See also the L:4 COMMENT column and reference to overheard conversations in the L:2 COMMENT column.)</p> <p>Learners at this level have some degree of mastery of the specialised language of their 'vocational' field(s). Limitations in 'vocational' (including academic) situations relate to the types of texts and interactions characteristic of the role, and to the speed with which learners can process texts. A particular 'vocational' role may place demands related to either or both of these. Ability to cope with discussions (particularly large-group discussions such as meetings and seminars) depends on, <i>inter alia</i>, the amount that participants interrupt each other, talk over the top of others and/or speed up in anticipation of being interrupted.</p> <p>For reference to societal roles that learners at this level can perform in the L<sub>2</sub>, see the S:3 COMMENT column.</p>

GENERAL DESCRIPTION OF LANGUAGE BEHAVIOUR	EXAMPLES OF LANGUAGE BEHAVIOUR VERSION FOR ENGLISH	COMMENT See also NOTES and GLOSSARY
<p><b>R:2 BASIC SOCIAL PROFICIENCY</b></p> <p><b>Able to satisfy basic social needs, and the requirements of routine situations pertinent to own everyday commerce and recreation and to linguistically undemanding ‘vocational’ fields.</b> The learner gets the sense of basic social correspondence on everyday topics from accommodative writers. Can obtain needed information from simply-structured routine correspondence pertinent to own consumption of goods and services and to a ‘vocational’ field which does not require high-level or specialised language skills. Understands a variety of high-frequency connectives, but extended lines of argument and complex or unfamiliar rhetorical structure are not followed. Has particular problems with any sections of texts where the information is dense (especially where there is little redundancy), or where important meanings are carried by complex or elliptical syntactic forms. In the routine situation types indicated, can overcome most difficulties related to vocabulary with a bilingual dictionary, but is likely to have problems with even relatively low-frequency idioms. If the situation is less routine but the register is familiar, can get some information with extensive use of a dictionary. Understands electronic or regular mail, provided in faxes the print-out is clear and in any handwritten texts the writing is neat and the style familiar. Provided there is strong support from the context, has some ability to go beneath the surface to understand implied meaning which is not subtle. Register sensitivity is limited. Generally understands only familiar varieties of English.</p> <p>Has some ability to meet recreational needs by ‘extensive’ reading in English. Can read for general interest and/or enjoyment simply-structured texts from the mass media, provided the genre and register are familiar (and generally the topic is not particularly abstract). Can enjoy popular ‘literary’ genres where there is significant support from the writing and/or editing process to make them accessible to native-speaking readers who would not otherwise be willing or able to read them, provided significant knowledge of the target culture is not assumed and the style is not unusual (e.g. dated). Any genuine appreciation of variation for aesthetic purposes is limited to the most obvious stylistic effects.</p>	<p>In <i>social</i> situations, can get the sense of simple personal letters on everyday topics. Even with a dictionary, however, will miss points if the writer has failed to accommodate the learner’s reading level and has, for example, made references to unfamiliar aspects of the culture or used difficult or unfamiliar forms such as elliptical structures, low-frequency colloquialisms and idioms, or low-frequency or idiosyncratic abbreviations. Understands handwriting only if the style is familiar and the writing is neat.</p> <p>In situations pertinent to own <i>consumption of goods and services</i>, can get the sense of simple routine correspondence (e.g. notice of increase in dues payable). With plenty of time and full contextual support, can follow sequential instructions (e.g. for a household appliance) written in non-specialised language and clearly presented (e.g. with labelled diagrams). Can understand the point of simple relevant community announcements (e.g. the recall of a product bought) but will need dictionary and/or other support for details. On first encounter with Social Security forms such as a <i>Claim for Parenting Allowance</i>, with some dictionary use can generally understand the instructions and associated information sufficiently well to know what information is required, but may need confirmation of some details.</p> <p>In ‘<i>vocational</i>’ (e.g. work) situations, can generally get the sense of simple routine bulletins without significant dictionary use. For simply-structured but less routine texts such as memos on a new venture and promotional material from a supplier, extensive use of a dictionary and/or other help will generally be needed for reasonable understanding.</p> <p>For <i>recreational</i> or general interest purposes, can get the gist of very simple news stories from a daily ‘tabloid’ or local weekly paper and of similar articles in popular magazines provided they are within familiar registers (and generally on topics which are not particularly abstract) and they have a familiar rhetorical structure and a significant amount of redundancy (e.g. single-thread human-interest stories about easily imageable events). Misses meanings realised by complex or unfamiliar grammatical forms or forms stripped of redundancies such as elliptical passives (e.g. headlines) or low-frequency idioms.</p> <p>Can read, with some use of a dictionary but a level of fluency sufficient to provide enjoyment, simple narratives such as novels based on screen-plays of films (particularly when the film has already been seen and/or stills from the film are incorporated) or popular modern novels simplified for the reluctant native-speaking reader.</p>	<p>At this level, learners can process texts which involve a variety of modifying devices at the sentence level, including subordinate clauses. They have sufficient mastery of embedding processes to be able to understand simple reported speech. They can understand continuous prose with simple propositional or textual relationships beyond the sentence level which are marked by high-frequency connectives.</p> <p>Communications from government and private institutions which can be understood by learners at this level are characterised by the use of ‘plain English’ (reflecting an awareness of the need to communicate better with all citizens, not necessarily L2 learners, by avoiding unnecessarily complex, jargon-ridden language).</p> <p>Learners confronted with a personal communication addressed to someone else are likely to find it inaccessible, because they are not privy to crucial underlying knowledge which the writer and reader share and do not need to verbalise.</p> <p>For an indication of the types of ‘everyday topics’ intended in the example of social correspondence given here, see the equivalent paragraph in the S:2 <i>EXAMPLES</i> column.</p> <p>Limited familiarity with the target culture significantly limits understanding at this level. Learners are likely to have problems with the language of the operations of ‘mainstream’ institutions which are unfamiliar to them. References to more ‘peripheral’ cultural institutions and phenomena are likely to be inaccessible. Unexpressed assumptions about unfamiliar aspects of the culture may cause total lack of understanding (see the L:3 and L:4 <i>COMMENT</i> columns).</p> <p>A genre or situation type may be familiar as a result of having been experienced in second language contexts or formal studies. In some cases, a genre may have a counterpart with very similar rhetorical structure and conventions in the learner’s L1.</p> <p>Where specialised situations have been experienced (e.g. at school or work) learners will have acquired elements of the language which are normally associated with more specialised aspects of the register, but will lack sensitivity in the use of such elements.</p>

GENERAL DESCRIPTION OF LANGUAGE BEHAVIOUR	EXAMPLES OF LANGUAGE BEHAVIOUR VERSION FOR ENGLISH	COMMENT See also NOTES and GLOSSARY
<p><b>W:1 BASIC TRANSACTIONAL PROFICIENCY</b></p> <p><b>Able to satisfy own basic everyday transactional needs.</b>            Provided there is support from the context, the learner can write very short series of original sentences directly related to own basic transactional needs or on very familiar topics. Can generally be understood in such situation types by sympathetic and/or experienced members of the general public. Uses a variety of functions, including giving and seeking factual information, suasion, and (very tentatively) asking about and expressing emotional attitudes. Information conveyed is usually very imprecise, however, because of the tentative state of grammatical development, with little or no use of modifiers. Uses only the most basic, high-frequency connectives; any extended discourse is largely a series of discrete sentences, relying on the context to provide coherence. Sentence 'frames' are complete, but typically are short and unelaborated. Grammatical errors may often cause or contribute to misunderstanding in less supportive contexts. Word order is strongly influenced by L<sub>1</sub> and there may be gaps in a sentence. Many basic cohesive devices are misused or omitted. The range of vocabulary demonstrated in actual purposeful language use is limited to that necessary to express basic needs and interests. Writes most words needed for the task types indicated with sufficient accuracy that they are recognisable; where individual words are not understood, the sentence meaning can usually be worked out from the context. Copies sentences related to basic transactional needs quite accurately. The influence of sociocultural factors from L<sub>1</sub> is strong. Register flexibility is extremely tentative.</p> <p>May use some items pertinent to specialised aspects of a register (e.g. in own 'vocational' field) if situations featuring such items have been sufficiently experienced.</p>	<p>Can write simple instructions related to transactions (e.g. delivery instructions or instructions to door-to-door vendors).</p> <p>With support from the context, can convey a simple message to a friend (e.g. an annotation on the advertisement for a function to suggest attending it together, or a note to advise the person of own inability to keep an appointment).</p> <p>Provided key vocabulary is familiar, or a bilingual dictionary or other supporting material can be drawn on, can write a short, very simple recount of a personal experience (e.g. on a postcard to a friend) or a report on a routine operation in the workplace (e.g. an annotation on a despatch or delivery docket about discrepancies).</p> <p>Can take down in dictation simple information (e.g. appointment details, including time of day, day and date; address and/or person to see).</p> <p>Can copy from written text quite accurately the sorts of information needed for basic transactional needs.</p> <p>Original sentences consist of little more than subject (phrase rarely exceeding three words), verb and object/complement/adverbial phrase (phrase rarely exceeding three words).</p> <p>Nouns and verbs are often uninflected. Modals are generally limited to <i>can</i>; <i>must</i>; <i>will</i> (signifying future). Negation may be signalled by the simple addition of <i>no</i> or of <i>not</i> without the auxiliary verb.</p> <p>Uses the most basic, high-frequency connectives (e.g. <i>and</i>; <i>but</i>; <i>so</i>).</p> <p>Register flexibility is limited to crude differentiations in basic courtesy forms such as greetings.</p>	<p>By this level, 'creative' language use is established. There are original collocations of words, which, while they may be very non-standard, have all the parts needed to be considered sentences (e.g. subject, verb, object). Learners can, therefore, express basic unpredictable needs within familiar situation types, those regularly encountered in, for example, shopping, commuting, work or school.</p> <p>There is, however, little or no use of modifying devices, those forms (e.g. verbs, adjectives, adverbs, phrases, clauses) that permit elaboration and qualification of the expression of ideas. Learners are, therefore, largely restricted to conveying 'universal' or stereotypic meanings and can not use English as a vehicle for expressing their own personality.</p> <p>Needs will differ according to whether the learner is living in an English-speaking environment or using the language in a 'foreign language' environment (e.g. with expatriate English speakers). Participation in an English language course creates its own 'needs'.</p> <p>'Experienced' in this context means 'used to communicating with non-native speakers'.</p> <p>'Suasion' covers a group of functions to do with 'getting things done'.</p> <p>Sentences at this level are 'simple' in the non-technical sense. Occasional sentences could technically be 'compound' in the sense of having two clauses in a simple co-ordinate relationship. Any attempt to produce original sentences with subordinate clauses often results in a confusion of sentence boundaries.</p> <p>For comment on the influence of sociocultural factors, see the W:0+ COMMENT column and the reference to using second person pronouns and persons' names in the S:1+ COMMENT column.</p> <p>The level of register sensitivity and flexibility is such that, if there is any use of items of specialised language learned as a result of, for example, work or school, learners are likely to have no awareness of the degree of technicality and the restricted applicability of this language.</p>