

My Approach to Designing Language Tests

D. E. Ingram

Professorial Fellow,
Faculty of Education, University of Melbourne,
Hawthorn Campus, 442 Auburn Road,
Hawthorn, Victoria, 3122,
Australia
©2006

My Approach to Designing Language Tests

D. E. Ingram

Professorial Fellow,
Faculty of Education, University of Melbourne,
Hawthorn Campus, 442 Auburn Road,
Hawthorn, Victoria, 3122,
Australia
©2006

Contents

- I My Involvement in Language Testing
- II My Approach to Language Testing
- III Factors influencing the Nature and Design of a Language Test
- IV Background to the Thai Test
- V Basic Principles in the Approach to Test Design
- VI Phases of the Thai Project
 - Phase 1: Preparation
 - Phase 2: Workshop 1
 - Phase 3: Writing the Pilot Test
 - Phase 4: Trialling the Pilot
 - Phase 5: Workshop 2
 - Phase 6: On-going Management of the ‘Live’ tests
- VII Key Features of the Test Development Project
 - VII.1 Needs
 - VII.2 Specifications
 - VII.3 Scale
 - VII.4 Training of the People involved in the Test
 - VII.5 Washback Effect
- VIII The Design of Other Tests
- IX How I would go about Designing an ESL Test for Use in Business or industry
- X Conclusion

My Approach to Designing Language Tests

D. E. Ingram

Professorial Fellow,
Faculty of Education, University of Melbourne,
Hawthorn Campus, 442 Auburn Road,
Hawthorn, Victoria, 3122, Australia
©2006

I MY INVOLVEMENT IN LANGUAGE TESTING

- Who I am: an applied linguist whose interest is in second language education, especially language policy, language education planning, curriculum design, methodology, language testing especially the measurement of language proficiency.
- I have a B.A., and A.Ed. from the University of Queensland, a Master of Arts and Ph.D. in applied linguistics from the University of Essex in England.
- I taught for 14 years in Primary and Secondary Schools in Queensland and Cambodia, then got involved with teacher education in 1972 at what is now the Mt Gravatt Campus of Griffith University in Brisbane, ending up as head of the teacher education program in what is now called the Charles Darwin University in the Northern Territory in Australia, before returning to Brisbane to found eventually three language centres. I was Professor of Applied Linguistics and Director of the Centre for Applied Linguistics and Languages in Griffith University and accepted a position as Professor and Executive Dean of the School of Applied Language Studies in Melbourne University Private in 2004. I have recently also been confirmed as a Professorial Fellow of the University of Melbourne.
- In 2003, my work in language testing, especially in language proficiency assessment and other aspects of applied linguistics was recognised by the Australian Honours system, when I was made a Member of the Order of Australia, the citation indicating that this honour was given for "*service to education through the development of language policy, through assessment procedures for evaluation of proficiency, and through research and teaching*".
- The first articles I published nearly 40 years ago were about language testing.
- In those early years, I regularly acted as an examiner and Chief Examiner for the public examination system (the university entrance examination) in my home State of Queensland.

- Since then I have written dozens of articles in journals and books about language testing, especially the assessment of proficiency.
- I have been invited to give papers at major national and international conferences around the world, especially in Australia, Britain, China, Finland, Hong Kong, Japan, Korea, the Netherlands, New Zealand, Nigeria, Singapore, Sweden, Thailand, and the United States and have run numerous training programs on language testing and other aspects of language teaching policy or methodology in various countries.
- From 1978 to 1983, I was an adviser to the Australian Immigration Department on the development of new ESL courses for immigrants to Australia. At the same time I was advising the Queensland Board of Secondary School Studies on the development of courses in language other than English. In both of these contexts, I realised there was need for a means of assessing the learners' language proficiency in a practical way and so I set out to develop what is now called the *International Second Language Proficiency Ratings* or ISLPR. That research has become a major pre-occupation now for nearly 30 years. My colleague, Elaine Wylie, and I have now developed many versions of the ISLPR to measure and state proficiency in English and many other languages. The scale is used to measure and state general proficiency and proficiency in special purpose contexts such as English for business, Indonesian for language teachers, and so on.
- In 1987 – 88, I was sent as the Australian representative on a joint Australian-British project based at the University of Lancaster in Britain to develop the International English Language Testing System (IELTS Test). When the test was released in 1989, I was appointed Chief Examiner (Australia) and remained in that role for ten years to 1998. After that I joined the Board of Directors of IELTS Australia for another four years.
- In the early 1990s, I was an adviser to a large project to develop an ESL test for the Australian Immigration Department to assess the English proficiency of applicants for migration to Australia. When that test was released, the Centre that I was directing at that time won the contract for the academic administration of the test for all the time it was used up to 1996.
- In 1999 – 2000, I was contracted to assist the Faculty of Arts in Chulalongkorn University, the most prestigious university in Thailand, to develop a university entry test in Thai writing. This was a particularly successful exercise and followed a model approach to test development, the same broad approach as I apply to all test development. It is especially suited to developing tests for the specific needs of industry or business in Korea or anywhere else.

- In 2002, I assisted another organisation based in Brisbane (now called the International Universities Consortium) to develop a test of English for use with young Korean children learning English in Years 3 to 10 of the Korean education system.
- My approach to test development is to find ways to ensure that it meets the immediate practical needs of whomver the test is being developed for, especially, for instance, business or industry.

II MY APPROACH TO LANGUAGE TESTING

- I see myself as an applied linguist with an interest in language testing, not as a language tester.
- This is important because it determines how I go about developing language tests.
- I claim rather to be an applied linguist with a broad interest in second or foreign language education as it relates to the teaching of ESL and other languages and language testing is just part of that interest. I have done a lot of work in language testing, including developing some very significant language assessment systems used around the world, but I am not a language tester. I like to see language testing as an applied linguistic field where the focus is on developing tests that reflect the real nature of language as it is used in the real world.
- I do not claim any expertise in statistics. I use statistics to the extent that statistics can help me in what I am doing. When I need to use statistics, I involve experts in statistics to assist. However, I see statistics as a useful tool and not what I, as an applied linguist, should be about.
- For me, the important thing in writing and validating language tests is to ensure that they reflect the nature of language. Most “language testers” seem to see language testing statistically with their main aim being to produce tests that can be shown statistically to be valid and reliable according to a statistical interpretation of what validity and reliability mean. However, it seems to me to be quite irrelevant that statistics show a test to be valid and reliable if the way you are testing does not truly reflect the nature of language.
- If our aim is to say something about how students use English or their practical ability in using English, it is essential that the test accurately reflect the nature of language and how it is used. Statistical notions of validity and reliability are of secondary importance to the nature of language and the purpose of the test. This profoundly influences the way tests are designed, what itemtypes, for example, are acceptable and how you interpret test results.

- One outcome of my approach to test development is that we can develop tests that accurately reflect the needs of the business, company or industry that wants a test of the practical English language skills of their employees.

III FACTORS INFLUENCING THE NATURE AND DESIGN OF A LANGUAGE TEST

The nature and design of a language test should be determined by:

- the nature of language and its use
- the purpose of the test
- the context in which it is to be administered
- the use to be made of the results and by whom
- the potential washback effect of the test on the teaching and learning of the language.

In brief, it should be needs-based, with "needs" being interpreted in terms of:

- the language it needs to test
- the needs of the persons wanting the test
- the candidates that need to be tested
- the contexts in which the test needs to be administered
- the purposes the test needs to fulfil
- the contexts in which the language is to be used and in which the language tasks that learners have to carry out occur, and the contexts for which the test is required
- the need for test security, what it should cover and how it is to be managed
- the nature of the washback effect the test needs to have.

In designing a test, there are several stages that one needs to go through:

- identifying the needs for the tests (including all those listed above and possibly others)

- developing specifications for the test
- determining how the test results are to be expressed and communicated
- trialling the test
- determining how, if required, the test is to be regenerated (e.g., a university entry test needs to be administered every year or every semester depending on how often students are accepted into university); (a test for a large company might be administered several times a year but can probably be kept strictly secure and so won't need to be regenerated so often.)
- determining training for the people involved in the test development, the test's regeneration, and the test administration
- determining on-going management procedures (including test security).

Broadly this approach was used in all the tests that I indicated earlier I have been involved with but here I will outline how I went about developing a test for the Faculty of Arts in Chulalongkorn University in Thailand. This is a particularly good model of how one should go about developing a good, valid and reliable test with results that are readily understood or interpreted.

IV BACKGROUND TO THE THAI TEST

In early 1999, the Faculty of Arts at Chulalongkorn University in Bangkok, Thailand, approached the Centre for Applied Linguistics and Languages (CALL) at Griffith University, Brisbane, Australia (where I was Director) to work with a project team to develop a test of Thai writing aimed at students leaving Secondary School and wishing to enter Chulalongkorn University.

The test was intended initially for students entering the Faculty of Arts but it was hoped that it might eventually be used much more widely across Thailand.

A very important aim of the project was seen as improving the writing skills of Thai students, especially those wishing to enter university. There was a strong belief that the use of multiple choice tests in the national university entrance examination had had a very harmful washback effect on the teaching of Thai language in the secondary schools. It was believed that the use of multiple choice tests had led schools to concentrate too much on training students to take such tests with the result that they neglected more extended writing and students came to university unable to write essays, assignments or anything that required extended writing. So the test had to be designed to have a favourable washback effect on the schools, i.e., it had to be designed

so as to encourage good teaching practices as schools prepared students to take the university entrance examination. In addition, of course, the test needed to test the students' ability to write in Thai, to identify the difficulties that some of them had and the remedial action required, but it also needed to be designed for quick and reliable marking because of the very large numbers of students who would take the test.

In brief, the new test had to be designed with the following basic characteristics (not in any priority order):

- It had to be rapidly and reliably marked.
- It had to identify students who would, potentially, have problems in coping with University studies as a result of deficiencies in their ability to write in Thai.
- It had to focus on the students' ability to write connected and extended texts in Thai.
- It had to have a favourable washback effect on the teaching of Thai in Secondary Schools, in particular, by encouraging the schools to pay attention to the quality of Thai students' writing, to their ability to write extended texts, and to their ability to write for academic purposes.
- To prompt appropriate remedial teaching, the test was to identify and draw the schools' attention to any areas of weakness currently seen in students' ability to write in the context and purposes of university studies. Since university life demands the ability to carry out a wide range of language tasks, the intention was to encourage students to develop ability in both the semi-formal and formal Thai styles.

To develop the test, the Faculty of Arts established a project team oversighted by the Dean and advised by myself. My role was as a facilitator, with the test remaining at all times the product of the project team and the Chulalongkorn Faculty of Arts.

V BASIC PRINCIPLES IN THE APPROACH TO TEST DESIGN

In designing the test, certain fundamental principles were adopted:

- The test was to be needs-based, i.e., it was to focus on the skills that were required in order to cope with the writing of Thai during students' university life and, within that, it was to identify any areas of weakness that could be remedied through appropriate washback into the schools. For this purpose, questionnaires were designed

and interviews held to elicit from the “end-users” (the university lecturers for whom the students write) both the types of writing that they expect students to do and the strengths and weaknesses that they perceive.

- The test was to be as realistic as possible, focussing, within the constraints imposed by the test situation, on forms of writing actually undertaken by Thai university students. This led to the use of itemtypes that prompted, as far as the test situation would allow, those forms of writing actually undertaken by Thai students in the course of their university life both within their formal studies and in their wider activities.
- Results had to be expressed in ways that would be meaningful in terms of the candidates’ writing abilities and in terms of the nature of their strengths and weaknesses. This led to the use of a rating scale that described the students’ performance and, in particular, the adequacy of that performance for university purposes.
- The test had to be able to be replicated rapidly and economically, giving consistent, valid and reliable results. This led to the development of clear and rigorous test specifications that would guide future regeneration of the test.
- Finally but not least, the test had to have a beneficial washback effect on the schools in terms of syllabus design and teaching methods.

VI PHASES OF THE THAI PROJECT

The work on the development of the test was undertaken by the project team working under my direction as facilitator. Much of the guidance was given by email though I also held two face-to-face workshops with the project team.

Because some of the project team members had had little formal training in language testing, the project had the dual role of developing the test and training the project team members in language testing.

The project, as originally designed, evolved in six broad phases:

Phase 1: Preparation

The first step was to find out exactly the needs for the test and the purpose of the test. Using a questionnaire that I designed with them, project team members interviewed key stakeholders (Chulalongkorn staff mainly in the Faculty of Arts but also in other Faculties) about what they believed the test should cover. Specifically, information was sought about staff expectations of students’ writing, the sorts of tasks that students in various disciplines have to carry out, the deficiencies that

staff had noted, and any other information that the staffmembers or the project team considered should be taken into account in devising the test. Samples of writing of Thai students were also obtained and team members were asked to grade six examples according to their adequacy, and to provide a short descriptive statement of each. They were also asked to obtain as many examples as possible of tests of writing (both in Thai and other languages) as input to the test design considerations.

Phase 2: Workshop 1

The project team, under my guidance, met for four days to draw up test specifications, using the information gathered from key stakeholders in Phase 1. This workshop was also designed to provide basic training in language testing for the team members. It commenced with an overview of the nature of language and of writing and it reviewed the purposes and types of assessment, giving particular attention to the notion of proficiency and how it is most effectively assessed. The nature and role of test specifications were discussed, as were alternative approaches to expressing test results, in particular the nature and use of rating scales.

Drawing on their own experience and the information obtained through the Phase 1 survey of stakeholders, the project team determined the broad areas to be included in the specifications. Issues discussed included the needs of the students; the purposes of the test; the itemtypes that could be used; the nature, purpose, strengths and weaknesses of particular itemtypes; how the results were to be expressed and reported; and the duration that would be appropriate for such a test considering the nature of the test, issues of fatigue, and the cost and reliability of marking. Examples of Writing tests were examined and evaluated and the team discussed how the test might be administered, scored and interpreted. Team members brought in some tests for consideration and several major international tests were reviewed.

By the end of the workshop, a draft set of Specifications for the test had been developed and evaluated in preparing exemplar items and items that could be used in the trials of the test. In addition, some alternative formats for a rating scale were discussed and a draft scale sketched. A scale was chosen partly for assessment purposes and partly to enable the test results to be expressed in ways that told end-users about the students' actual writing skills.

Phase 3: Writing the pilot test

Following the first workshop, the project team prepared a pilot or trial test. I also re-worked the draft Specifications and finalised the draft of the scale. Throughout all of this activity, there was continual communication and feedback both ways between the project team and the facilitators.

Phase 4: Trialling the pilot

The test was trialled with some 250 secondary school students. It was administered and marked by members of the project team, who also used the exercise to provide feedback on success of the format of the test and its practical effectiveness to the next workshop.

Phase 5: Workshop 2

A second workshop was held to review progress on the test (especially in light of the experience gained during the test's development and its initial administration), to evaluate the trials, to review the Specifications and the scale, and to consider the training required for future assessors and itemwriters. This workshop adopted a 'train-the-trainer' model so that the participants would be able, in future, to train other people to work on the test as itemwriters or assessors. The workshop also considered what was required for the on-going management of the test, including test security.

Phase 6: On-going Management of the 'Live' tests

The final phase of a test development project is the test's on-going maintenance, administration and regeneration of the test. This is in the hands of the project team under the Faculty of Arts which owns the test. Part of this on-going management requires the continual review of the Specifications, the training of assessors and itemwriters, the writing and trialling of new versions of the test, and the monitoring of the performance of assessors to ensure consistency of judgements between assessors and across test versions. I have occasionally kept in touch with the people running the test, e.g., during my recent lecture visit to Chulalongkorn and have provided advice on any problems or new initiatives.

VII KEY FEATURES OF THE TEST DEVELOPMENT PROJECT

There are several important aspects of this project which I would urge be considered in any major test development project, especially if the aim is to meet the specific needs of business or industry or of a large enterprise:

- **The needs-based nature of the test:** it is highly desirable that any test be closely related to the needs, both the needs of the candidates, what the test has to measure, the needs of the end-users (or the people who use the results of the test), and the needs of the people who want the test
- **Specifications:** Detailed specifications provide the blueprint for the development of the test and for its regeneration.

- **The emphasis on practical skills:** whether it is for university entry purposes or for vocational purposes or simply to know how well the candidates can use the language, it is important that the test identify the practical skills that the candidates have and not only their formal knowledge about the language.
- **The level of the knowledge and skills of the assessors and itemwriters** that could reasonably be expected: if you want a high quality test, it is essential that the people who develop the test and who later administer it, be well trained in the best approaches to language testing. I would strongly recommend an approach under which people familiar with the situation are preparing the test but under guidance from someone expert and experienced in language testing but also willing to listen to the needs of the end-users or people commissioning the test.
- **The washback effect of the test:** Any test that serves an important purpose such as university entry, employment in a company, or salary bonuses is going to influence how the language is taught in schools and other training institutions and so one should always take that washback effect into consideration in designing a test.

VII.1 Needs: In order to identify the needs that the test should meet, prior to the first workshop, we prepared a questionnaire for the project team to present to key stakeholders (university lecturers in particular) in order to elicit what they saw as the nature of the writing that Thai university students undertake during their course and the deficiencies that they see in their students' writing skills. In developing a test for use, for example, in an enterprise in Korea, I would want to identify clearly the needs the test is to meet, the skills the company wants its employees to have, the tasks they have to carry out, and so on.

The questionnaire we used in the Thai project sought the views of key stakeholders on

- the adequacy of the Thai writing skills of students entering university,
- the strengths and weaknesses that the informants observe in the students' writing,
- the importance of certain key features of writing including mechanical features (such as handwriting, spelling, grammar and self-editing), discourse or text features (including paragraphing, overall text structure, use of linking devices, and coherence of argument and ideas), and formal features of academic writing (such as referencing and a bibliography),

- the writing tasks, text types and written functions that they require of their students,
- the use they expect students to make of reading materials and other sources of ideas and information,
- their support for the introduction of a writing test for university entry purposes, and
- other information or advice they thought the project team should consider.

VII.2 Specifications: The development of systematic specifications for a test is of great importance for at least three reasons:

1. The test specifications provide the parameters over which the test is constructed and help to ensure that subsequent versions of the test are similar in format, focus and demand.
2. Specifications are important in ensuring the quality of the test and public accountability, features of vital importance in a “high stakes” test that is used for purposes that could determine the paths of the candidates’ lives.
3. If the test is to have a washback effect on the education system, then the format of the test must be publicly known: it must be evident what the test is trying to measure and what features of the language are considered to be important and will receive focus in the test.

The wording and format of the Specifications for the Thai test were the result of the needs survey, much discussion by the project team and facilitators, and the results of the pilot test. Every test is different because it serves different needs and purposes but a typical set of Specifications would have headings such as these:

- Nature and Purpose of the Test
- Length of the Test
- Test Structure
- Itemtypes and Tasks
- Text types
- Topics
- Language Operations or Functions
- Texts and Other Stimulus Materials: Their Sources and Characteristics
- Proficiency Levels
- Difficulty of Task
- Rubrics

- Audience
- Authenticity
- Test Bias
- Assessment/Scoring
- Certification
- Presentation by Itemwriters
- Quality Assurance
- Short Checklist for Itemwriters

The specifications for the itemtypes and the test as a whole should be carefully designed to ensure realistic language activities and the Specifications should be developed with this aim always in mind.

VII.3 Scale: The nature of the items in the test are important in focussing on the practical ability of the students. The way the results are expressed is also vital in focussing on their practical ability. In the Thai test, the results were to be given in terms of a rating scale that focussed on the ability of the students to cope with university studies and also focussed on those key features of the language that were identified in the needs survey as of particular significance in determining the suitability of the students' writing skills for university purposes. The advantage of using a descriptive scale with which to express the results is that a short description of what students can do in the language is more meaningful than a number such as 70% or a word like "Distinction" or "Pass".

Scales differ greatly in complexity from simple statements of language behaviour to very long and complex descriptions of the learners' language (as in the ISLPR). In the Thai case, we developed a simple six level scale describing five features of the written language that the people in the needs survey thought were important in university writing. These features were:

- "Task Fulfilment",
- "Argumentation and Discourse",
- "Formal Features",
- "Word Choice", and
- "Register".
- The sixth sub-scale is the "Overall Readiness Scale".

However, the features to be described in, say, a scale for the purposes of a Korean manufacturing company will undoubtedly be different depending on the needs and the purpose of the test.

VII.4 Training of the People involved in the Test: An important aspect of the approach to test development adopted in the Thai test was the training on language testing that the people involved in the test received, both the initial project team and the people subsequently involved.

VII.5 Washback Effect: As already noted, the Thai test was designed to have a favourable washback effect on the teaching of Thai in the Secondary Schools in Thailand. The development of careful Specifications, the use of itemtypes that focussed on realistic language skills, and expressing the results in a readily interpretable behavioural scale that highlighted important aspects of writing were all important in producing that favourable washback effect.

VIII THE DESIGN OF OTHER TESTS

A similar approach was adopted in designing the other tests with which I have been involved, including IELTS, ACCESS and ISLPR.

IX HOW I WOULD GO ABOUT DESIGNING AN ESL TEST FOR BUSINESS OR INDUSTRY IN KOREA

- First, it is essential that the test meet the specific needs of the client for whom we are developing the test.
- Hence, I would strongly recommend an approach like that I followed in Thailand where the project team is familiar with the needs of Korea and the company and who are also given training in good language test development.
- I would recommend a short 2 or 3 day training program for the project team so that they are familiar with the best approaches to test development and have the opportunity to consider the specific needs of the enterprise or industry.
- A good test must meet the needs of the people and the education system. Hence, I would ask the project team to identify clearly the needs the test is to meet. For this, I would want them to review whatever policy documents are available and, using a questionnaire that I would develop with them, to get information from relevant people about the needs of the test and what they want to see tested. “Relevant people” will include policy-makers, teachers, enterprise staff and other “end-users”. Most of all, I would want information on the specific tasks that the employees to be tested have to carry out in English so that the test to be developed would give specific information to the employers about the candidates’ skills and their linguistic ability to carry out the tasks required of them in English. Developing a suitable questionnaire and identifying the source of other information on the needs of the test would make up part of the initial workshop.
- When the needs survey had been completed, I would recommend a second workshop in which the project team under my guidance developed the Specifications for the test and drafted a trial test. In this workshop, we would also consider how the results were to be

expressed and develop whatever instrument was needed for that purpose. A scale is a useful tool for this purpose but there are other ways of expressing results as well and these would have to be considered.

- The trial test would then be trialled, the trial observed by the project team, and the results analysed.
- A final third workshop would be held to consider the results of the trial, adjust the Specifications, and finalise the preparation of the test. The future management of the test and its regeneration for use whenever university entry examinations are held would also be considered and a management plan developed for recommendation to the appropriate authorities.

This approach has a number of important advantages over, for example, an approach in which some foreign test organisation comes in and offers its existing tests for use in the country:

1. The test should be based strictly on the needs of the end-users and whoever commissions the test.
2. It should be developed in close consultation with the end-users (e.g., the Korean company): no one can know better than the people in the company what the needs are and no outside body should seek to impose a test developed for other purposes on Korea or on the company. The role of the testing expert in this activity would be to contribute experience in developing tests for many different purposes and for use in many different countries, to guide the project team in the best possible approach to language test development.
3. Very importantly, the test would be designed to have a favourable washback effect on the teaching of ESL in schools, in university or in the company's own programmes.
4. The project would not only produce a world class test of ESL, it would also provide training for the project team in test development. In fact, by the end of the project, the project team would have gone through a "train the trainer" program so that they can themselves go on to train other people in test development or specifically in developing and administering the new ESL test for university entry purposes.

X CONCLUSION

As I said at the outset, my approach to language testing is very much an applied linguistics approach that takes language testing beyond heavy reliance on statistical procedures to take strong account of the nature of the language that the candidate is meant to be using, the purpose of the test, the needs of the candidates and the end-users, and the interpretability of the test

results. In this approach, consideration is also given to the washback effect that any high stakes test is likely to have on the language education system whether in the schools or in the private classes of the enterprise that has commissioned the test. This approach to test development is one that has been tried with considerable success in other test development projects that I have managed.