Relating research and professional development: Assessment task design

Abstract

In Australia, the presence of a large-scale, nationally funded, English language program for migrants has provided a unique site for research into a range of aspects of language learning. This paper explores the contribution that participation in research can make to teacher professional development, using as an example a large-scale investigation of the competency-based assessment practices which are used in the Australian Adult Migrant English Program (AMEP). The project illustrates the ways in which research can involve teachers in investigative and reflective activities which enhance their understanding of assessment issues while at the same time offering concrete ‘findings’ which can inform teaching and assessment practice.

Introduction

With the increasing emphasis that is now given to accountability and outcome reporting in language learning programs, assessment issues have assumed a much higher profile in language learning programs. It is thus becoming important for teachers to have a quite detailed and in-depth understanding of the principles and practice of assessment. However, as a number of researchers have noted in the context of both general education and language education, teacher preparation in assessment is not always sufficient to enable them to handle the responsibilities associated with developing and administering their own assessments with confidence (Brindley 1997; Doherty et al 1996; Fager et al 1997).

One way of addressing this issue is to provide formal specialist language testing and assessment courses into teacher training at both the pre-service and post-training level (Bailey and Brown 1996). However, the content of such courses is frequently focused on technical issues of measurement and thus may not always be perceived by teachers as relevant to what they do in the classroom (Whittington 1999: 20). As Brindley (in press) argues, what is needed is an approach that:

❖ Links preparation in assessment closely to student behaviour in the classroom
❖ Capitalises on existing practices, focusing on what teachers already...
Encourages a research orientation to professional development through involving practitioners in research projects in their own workplace.

This paper reports on a research project based on these principles that set out to investigate aspects of the assessment system that is used within the Australian Adult Migrant English Program (AMEP). The project was designed and conducted with the dual aim of enhancing teachers’ skills and knowledge through involving them in the process of the research and ensuring that the product was of practical use in the classroom.

Assessment in the AMEP

The Certificates in Spoken and Written English (CSWE)

During the 1980s, in response to demands for more systematic learner pathways into further education and more transparent reporting of learning outcomes, a nationally accredited curriculum framework — the Certificates in Spoken and Written English (CSWE) — was adopted by the Adult Migrant English Program in Australia. The CSWE consists of certificates corresponding to four levels of English language proficiency and provides statements of learning outcomes in the form of language competencies in oral interaction, reading and writing (NSW AMES 1998). Each competency consists of elements, describing the skills and knowledge involved in the performance, the performance criteria against which the learner will be assessed and range statements that specify the key features of the text to be produced or understood and the conditions under which the performance will occur.

In the CSWE system, learners may obtain a Certificate at each level once they have successfully completed a specified set of competency requirements. Competency attainment is assessed on the basis of learners’ performance on tasks that are designed to elicit the behaviours described in the competency statements. A binary rating system is used, whereby the performance criteria for each competency are rated by teachers as achieved or not achieved. For any competency, the task used for assessment purposes may be selected from a range of tasks designed for the purposes of assessing that particular competency, or alternatively, the task may be developed by the teacher.

Assessment and reporting in the AMEP curriculum

The CSWE system of assessment is designed to allow teachers, students and other stakeholders to determine levels of competence in the language, and subsequently to describe competency gains following given periods of instruction. Thus, the system allows learners to assess their own
language gains in terms of their ability to perform particular activities, while teachers have a common framework within which they are able to assess and report individual learner performances (Feez, in press).

As Brindley (1998: 52) notes, outcomes-based systems such as the CSWE allow assessment and reporting to be closely tied to classroom learning:

... since there is a direct link between attainment targets, course objectives and learning activities, assessment is closely integrated with instruction: what is taught is directly related to what is assessed and (in theory at least) what is assessed is, in turn, linked to the outcomes that are reported.

For the learner, there are also considerable advantages to this type of achievement assessment. It does not take place under rigid test conditions, the assessment is generally performed in a familiar situation, with a familiar teacher, and the assessment reflects real-world activities in which the learner may reasonably expect to take part in their daily lives. Thus the pragmatic aspects of the assessment procedures are clear for the student.

The requirements of funding bodies, however, are rather different from those of teachers and learners and tend to be politically and economically motivated (Brindley 1998). Funding authorities now require increasingly more precise information on competency achievement in order to report on program outcomes for accountability purposes (Ross 2000). Since a learner’s access to future language learning opportunities may depend on the number of competencies he or she has achieved, it is important that this information is dependable. However, for this to happen, there needs to be an assurance of comparability across different assessment exercises, across different teachers’ ratings, and across the different tasks which may be used to assess a single competency.

**Teachers as assessment task developers**

CSWE assessment requires teachers to select or develop tasks with which to assess students on a range of competencies. They may need to assess a group of learners on up to eighteen competencies across different language skills, each competency requiring a different assessment task. However, since assessment tasks in the CSWE are closely related to the learning activities that are used in the classroom, teachers need to know not only how tasks work as teaching tools but also how they work as assessment instruments. In this regard, the following questions are critical:

❖ Are different tasks, used for the assessment of the same competency by different teachers, comparable in difficulty?

❖ To what extent, and how, do the different conditions under which the tasks are administered affect the outcome on the tasks?
To what extent can different characteristics inherent in the tasks affect the language outcome for the learner?

Do learners interact differently with different types of tasks?

What is the role of the interlocutor in the administration of the oral tasks?

To what degree do different teachers interact differently when rating different types of tasks?

In the context of the AMEP, the answers to these questions can assist teachers in a number of ways:

They can help to inform the decisions teachers must make about which task to use for which competency assessment.

They can develop teacher awareness of a range of issues which pertain to assessment practice in general, and assessment practice within the AMEP in particular.

They can provide guidelines to assist in the development of similar tasks for both assessment and classroom practice.

In summary, AMEP teachers require extensive knowledge about tasks and how they behave under different conditions if they are to use them for assessment purposes and provide dependable information that can be used to report program outcomes. In addition, it is important that investigation of the way these assessment tasks function within the AMEP takes place under operational conditions to ensure that the findings are relevant to the reality in which teachers work.

**Tasks and variability**

As we have seen, assessment procedures used throughout the AMEP involve the assessment of competencies through task-based activities. Tasks must therefore be designed with a view to providing an adequate sample which allows an appropriate assessment of the learner's performance. However, research into task performance in language assessment suggests that not only do tasks vary in terms of their difficulty, but that they may also elicit variable samples of language under the influence of a range of factors (Wigglesworth 2000). For example, the role taken by the interlocutor in an oral interview may affect both the quality and quantity of the language produced by learners (O'Loughlin 1997; Morton et al 1997). This variability is of particular concern with the CSWE competencies since tasks are not standardised across teaching locations (teachers may either choose from a set of pre-written tasks, or develop for themselves the tasks which they use to assess their students). Thus there is great potential for variability in the tasks used for similar assessment purposes.

These findings regarding task variability are consistent with emerging...
evidence from studies which have involved detailed analyses of the language of second language learners on a variety of tasks and task types. In a wide-ranging examination of task-based approaches to teaching and assessment, Skehan (1998) found that performance on tasks may be influenced by either the characteristics of the task (features inherent to the task) or the conditions of the task (related to how they are administered). Findings from a number of studies suggest that even quite small differences in the way the task is designed or delivered may affect both the quantity of linguistic output, and its quality in terms of fluency, accuracy and/or complexity (Foster and Skehan 1996; Robinson forthcoming, Robinson et al 1995; Skehan and Foster 1997; Wigglesworth 1997).

**Investigating CSWE assessment tasks**

Given the potential for variability in assessment task design and the need for teachers to be able to provide dependable information on learner performance, a research project was initiated which had two principal aims:

1. To investigate the effects of a range of variables on learner performance in CSWE assessment tasks, with a view to providing a research base which could assist teachers to develop their own tasks for CSWE assessment, as well as for the classroom.

2. To develop teacher awareness of issues in assessment task design through involvement in the project.

In pursuit of the first of these aims, four separate but related sub-projects were initiated, each focusing on one of the four skill areas of reading, writing, speaking and listening. These projects set out to examine how different competency-based tasks function when administered under normal assessment conditions. In particular, they were concerned with establishing the extent to which task difficulty could be affected by variations in task conditions or task characteristics. To address the second aim, the projects were designed so that teachers would be involved throughout in developing and administering the tasks to learners while at the same time reflecting upon the empirical results which emerged from the project. It was hoped that these results would help them to focus on the key variables that need to be considered when designing competency assessment tasks.

**Research procedures**

For this investigation, a range of assessment tasks was developed for each skill area. To avoid confounding effects, the tasks were developed so as to be identical in all but one variable. The variables chosen often reflected key aspects of the range statements corresponding to the competency in question which specify the conditions of administration and the features of the elicitation procedure or text to be used for assessment. The range
statements were targeted in this way, since it was important to establish whether the full range of factors affecting task outcomes was included. If the results of the investigations revealed significant variability in relation to some variables or suggested that other key variables were absent, this information could then be incorporated into the assessment tasks and, ultimately, into the competency specifications themselves. For example, in two of the oral competencies, the same task could be administered by either a native speaker or a non-native speaker. If it emerged from the investigation that having a different interlocutor affected learners’ performance, and hence their chances of achieving the competency in question, then this would need to be explicitly recognised in the design of assessment tasks in order for learners to have an equal chance of success.

Other key variables which were investigated related to the nature of the task itself. For example, for some of the oral and written tasks, structured prompts were provided (for example, indications of specific points to be addressed in a letter), while in others the structure was not provided. Another variable which was examined in relation to the writing tasks was the effect of pre-task classroom activities on the performance of the task: for example, was there a difference depending on whether there was a teacher-fronted discussion first, or a pair-work exercise? In the reading tasks, the variables related to either the type of question (for example, short answer responses versus true/false responses), or the length of time allowed to complete the exercise, or to the passage (for example, whether the passage was written in the first or third person). The investigation of listening tasks examined features of the input (for example, speed of delivery, text type, live versus recorded input) and of the assessment task itself (for example, item format).

**Teacher involvement in the research project**

The involvement of teachers in AMEP centres throughout Australia was a crucial element of these research projects. This involvement took a number of different forms. For the productive skills, oral and written, AMEP providers throughout Australia were asked to submit a number of exemplars of oral and written tasks for the selected competencies. These were used as a basis from which to develop the series of tasks employed in the project which were then returned to the AMEP centres where there were teachers who had volunteered to be involved in the project. These teachers acted as the interviewer for the oral tasks, administering the series of tasks to the learners, and recording the responses on tape. The written tasks were also administered by the teachers in their classes. In the case of the written tasks, the differences were often related to a pre-task activity. Thus, one teacher might conduct a teacher-fronted group activity related to the task before the learners wrote their responses to the task, whilst another teacher would organise the learners to do a pair-work activity related to the task prior to responding to the task. The teachers carried
out the activities in their classes, and then collected the written responses from the learners.

For the development of the reading and listening tasks, full day workshops were held for small groups of volunteer teachers in a number of states throughout Australia. These workshops introduced teachers to issues and problems involved in the development of reading and listening tasks and provided them with an opportunity to look critically at a range of assessment tasks. At the beginning of the workshop, teachers themselves responded to a reading or listening test. These exercises were designed to encourage teachers to reflect upon a variety of issues which relate to assessment tasks. Teachers were asked to comment on the following aspects of the test:

❖ Were the rubrics clear and adequate? Did they provide enough contextual support and relate well to the items?
❖ Were the texts appropriate for the target group? Were they authentic? Were they appropriate in length and complexity?
❖ Were the items too easy or too difficult? Were they appropriate to the text? Did the response depend on skills other than the one being tested? Were they culturally biased?
❖ Were the questions comprehensible? Was ambiguity avoided? Did they include difficult vocabulary?
❖ How salient were the answers in the text?

This activity was used to motivate further discussion and to develop and discuss the features of tasks which contribute to making a good reading or listening test, and to identify the basic faults which can be made in item design. From this point, the activities focused more specifically on the competency tasks, with a variety of sample tasks being examined and discussed. For example, teachers were asked to rank the tasks in order of relative difficulty, to reflect upon their reasons for their ranking, and to identify tasks which they felt would be appropriate for use as competency assessment tasks, with or without modification.

The second half of the workshops consisted of teachers working in pairs or small groups and selecting appropriate texts, adapting them if necessary, and then developing items which were appropriate to the text for assessment of the relevant CSWE competency. Thus the workshop activities allowed a variety of reading and listening assessment tasks to be developed, and these formed the basis of the tasks which were then piloted nationally.

Once the final set of tasks had been selected, the teachers administered them, and collected the learner responses. In the case of the productive competencies, another group of teachers rated the responses in order to identify any differences in the performances by the learners according
to the variables. Quantitative and qualitative analyses were carried out, which were conducted by the researchers involved in the project, and these allowed the effects of the variables on the tasks to be identified. Detailed discussion of the procedures used, and the results and conclusions of these investigations to date can be found in Brindley (2000) and Wigglesworth (in press).

**Professional development outcomes**

Findings from the various projects described here have been incorporated into a number of professional development workshops and presentations for teachers on assessment task design and evaluation (Brindley and Slatyer 1999). Workshops are designed to encourage teachers to evaluate and reflect critically on task differences which may result from the different characteristics of the task, or the conditions under which they are administered. For example, in workshops related to the oral tasks, teachers were provided with examples of all the tasks used in the project, and asked to discuss them in relation to their level of difficulty. They were then asked to rank the set of tasks within each competency in relation to their perceived difficulty, and to provide reasons for their ranking. Once the teachers had completed this activity, the variables which had been used were introduced and discussed. The task used for this activity is included in Appendix 1. In groups, teachers reflected upon how they considered the variables might influence the difficulty levels of the various tasks, and the language elicited by the tasks. Once this had been completed, the empirical results from the project itself were presented, then discussed in relation to the variables.

The next part of the workshop provided teachers with transcribed samples of discourse from a range of tasks. These examples provided a mechanism for discussing specific learner performances on different tasks, and for examining other factors which impact upon the task. Samples of data from two different learners doing the same competency task with different interviewers are provided in Appendix 2. Teachers were able to examine these data, and to identify reasons why the same task could elicit such different responses. From these activities teachers were able to observe the effect the interviewer can have on the task. As the examples in Appendix 2 show, the interactive nature of the task can change quite dramatically depending upon how the interviewer approaches the activity. Additional discourse examples allowed teachers to identify other features of interviewer performance which can affect the difficulty of the task, such as the complexity of the structures used, the amount of support provided to the learner, or the level of vocabulary used. The issues which emerged from these activities were then discussed in relation to task design, and consideration was given to the ways in which tasks could be constrained to reduce the variability.
Building on the materials used in these workshops, a set of professional development resources is being developed which will be used to complement existing training kits, such as that produced by Christie and Delaruelle (1997). These resources will include relevant professional readings from the research literature, along with a range of sample assessment tasks for different CSWE levels, accompanied by examples of learner performance that illustrate the different outcomes elicited by different kinds of assessment tasks. In addition, during 2000 a national group of teachers has been involved in ongoing development and trialling of assessment tasks through the kinds of workshops described above. When trialling has been completed and the tasks have been modified in the light of the results, a bank of fully piloted assessment tasks will be available to teachers which can be used in their present form or as models for task development.

**Teacher involvement in research-based projects**

Teachers involved in the projects have reported in positive terms on the value of their participation. For example, one teacher reported as follows:

... I’d like to say that I have developed in many ways during this project. Obviously I have become more proficient at designing and evaluating assessments. As I have increased my own awareness, I’ve had some opportunities to work with others and contribute to their gains in awareness. This has included working informally with colleagues and more formal moderation sessions.  

(Oettel 1999)

More generally, teachers can relate to these findings at the practical level because the findings of such studies feed directly into classroom learning and enhance their understanding both of assessment practice in general and CSWE assessment practice in particular. Critical scrutiny of assessment tasks, for example, can help to focus teachers on the types of language that are elicited by different types of tasks and draw their attention to those factors which may help or hinder learner performance, both in teaching and assessment contexts.

The enthusiastic response from teachers exemplified above and the interest from the research community in the issues discussed here (see, for example, Norris et al 2000), has been testament to the value of research which aims to involve teachers in investigating their own practices and using the results to help improve teaching and learning (cf also Burns, this issue).

**Conclusion**

The American Federation of Teachers, the National Council on Measurement and Education, and the National Education Association have identified seven standards for teacher development in the area of assessment. Two of these are that, first, teachers should be skilled in choosing assess-
ment methods appropriate for instructional decisions and second, that they should be skilled in developing assessment methods appropriate for instructions and decisions (American Federation of Teachers 1990).

Given the assessment and reporting requirements of the CSWE, it is essential for teachers working within the Australian AMEP to have the opportunity to acquire these skills. The type of research-based professional development described in this article constitutes an important and necessary first step in this direction.

References


Brindley, G 1997. ‘Assessment and the language teacher: Trends and transitions’. The Language Teacher, 21, 9: 37, 39


### APPENDIX 1

Consider the following variables and try to decide what effect, if any, they might have on the difficulty of a task they were assigned to:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Operationalisation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Structure</strong></td>
<td>five specific questions or pieces of information to collect rather than a general prompt</td>
</tr>
<tr>
<td><strong>Familiarity of activity</strong></td>
<td>an activity which it was considered would be familiar to the participants (that is, frequently carried out) rather than one which was not (that is, infrequently carried out)</td>
</tr>
<tr>
<td><strong>Complexity</strong></td>
<td>obtaining information about something related to self (less complex) versus obtaining information about something related to someone else (more complex)</td>
</tr>
<tr>
<td><strong>NS vs NNS</strong></td>
<td>interlocutor is a native speaker of English, or interlocutor is a non-native speaker of English. Non-native speaker may be another student.</td>
</tr>
<tr>
<td><strong>Planning</strong></td>
<td>five minutes is provided for preparation of the task, or no time is provided</td>
</tr>
</tbody>
</table>
APPENDIX 2

Example 1 [Certificate II, Competency 5: Can give spoken instructions]

I: Oh Vinh, can you help me, I don’t know how to use this card.
S: Yeah. Er, if er if you you have an automatic teller machine card
I: No
S: You want you want the er push the push out the money?
I: Yes I [I want to]
S: [in] Commonwealth Bank?
I: Yes
S: Er you um you can you /can/ um the start and start you can push the button, um, your PIN number
I: My PIN number
S: Your PIN number
I: Oh yes I [remember] my [PIN number] yes
S: [yeah] [PIN number] yeah [/?]?
I: [so I]
S: You put PIN number again
I: Yeah
S: Then you er you put OK [OK]
I: [yes] yes
S: /a/ and withdraw and push the withdraw button
I: Um where’s that withdraw, oh yes
S: On the on the left [on the left]
I: Yes [right OK]
S: And you push the saving or the mm /depend you/ [saving] [/??/ [saving] yeah
I: Oh [right ] er I er[ yeah I have a savings] account yes
S: You push the button saving
I: Savings account, yes
S: Yeah and then you push the, and number the money er
I: OK I want to take out fifty dollars
S: Yeah how many money and how
I: Fifty dollars
S: Yeah how much
I: So where do I put fifty dollars
S: Then yeah you push the er final you push OK you er you wait, you wait
I: OK OK yes and now what I wait
S: Yeah you wait then the machine, er, pay you the card er card and er the paper the write the mm the mo/the money of you
I: Oh yeah
S: And er you you put the money [laughs]
I: OK
S: You get money
I: Oh thank you, that’s easy
S: Yeah
I: Thank you very much, for your help /?/
S: Oh no problem
I: Bye

Example 2 [Certificate II, Competency 5: Can give spoken instructions]

S: Put into ... ATM card inside the machine. ... Press your /private/ number. ... Look at the machine er er look at the /?/. Choose choose what do you want first er, ... Receive money or your card. Finish.