

**“Does it measure up?” Benchmarking the written examination of a university English pathway program**

**Kieran O'Loughlin,**

**Melbourne Graduate School of Education**

**The University of Melbourne**

**Abstract**

*In recent times there has been a good deal of controversy surrounding the exit standards in English for academic purposes (EAP) programs which provide international students for whom English is an additional language with an alternative pathway into Australian higher education. These pathways exempt students from meeting specified minimum levels of proficiency on a standardised test such as the International English Language Testing System (IELTS) or Test of English as a Foreign Language (TOEFL). Critics argue that students who have successfully completed these programs may not have attained the same level of proficiency as students who have achieved the minimum standard required on one of the standardised tests. This paper reports on a study which aimed to benchmark the final written examination in such an EAP program against the IELTS Academic Writing module so that achievement on the program could be referenced against a standardised proficiency test.*

**Background and Significance of the Study**

To date, most leading English-medium universities around the world have only admitted international ESL students who have reached set cut scores on recognized standardised English language proficiency tests, such as the International English Language Testing System (IELTS) and Test of English as a Foreign Language (TOEFL), as well

as satisfying their academic requirements. Recently, many of these universities have developed English pathway courses which allow students, upon successful completion, to commence their university studies without being required to achieve the requisite entry scores on one of these tests (see, for example, Leask, Ciccarelli and Benzie, 2003; Agosti, 2004; Cruickshank and Chen, 2005; Banerjee and Wall, 2006).

The introduction of such courses has been partly motivated by the commercial imperative for universities to attract more international students. In addition, there has been a growing recognition that standardised proficiency scores do not necessarily indicate the extent to which an individual is ready to undertake English-medium higher education (see, for example, Deakin, 1997; Turner, 2004). It appears that many students derive great benefit from completing an English language course *in situ* which aims to improve their level of English proficiency but which also equips them with the study skills and cultural knowledge they need to fulfil their academic potential once they commence their higher education studies. These institutions have used such courses to attract prospective students who have either failed to achieve the required level on a standardised test or simply prefer to avoid such stressful assessment procedures. However, these pathways courses have recently been the subject of media controversy with some commentators questioning the level of English proficiency required to successfully complete them (Birrell, 2006; Ziguas, 2007).

The problem is that there are no national or state standards for the summative assessments made on these courses. In the absence of such standards a possible alternative approach is to reference these assessments against a standardised proficiency test. This article reports on a study which sought to benchmark student assessment on the new University of Melbourne English Language Bridging Program (UMELBP) final examination against the International English Language Testing System (IELTS).

## Creating Standards

Standard setting, a process of creating cut scores which correspond to defined levels of achievement or proficiency on examinations, has become an important issue in educational testing and certification. Within language proficiency testing for tertiary study, cut scores reflect a student's "readiness to learn", or more specifically, the extent to which they have a sufficient command of the language to commence their studies.

While most educational institutions have cut scores, or minimum score requirements, for international proficiency tests, such as the IELTS and TOEFL, this is generally not the case for other more local measures of English ability, particularly pathway programs offered by the institutions themselves as well as their affiliated colleges and language centres. These include the English component of Foundation Studies (which provide a broad academic preparation for undergraduate degree study) and ELICOS (English Language Intensive Courses for Overseas Students) which are offered to prospective undergraduate and postgraduate students. Nor can the final assessments currently be benchmarked against any appropriate national or state standards. As Cizek and Bunch (2007, p8) suggest, "... explicit standards are likely to result in increased understanding by and trust on the part of the public".

In the absence of such standards, a default benchmark is the most widely-used and accepted proficiency test, the IELTS exam. A key question then is whether IELTS test scores and exit assessments on English pathway programs are equivalent. In other words, do they provide the same information about a test takers' readiness to study through the medium of English? In an important sense they do not: ELICOS and Foundation programs yield *specific* achievement assessments based on a particular course of study while a proficiency test such as the IELTS exam provides a measure of *general* achievement in the language unrelated to any particular course of

---

instruction. Nevertheless, it may still be possible to reference one kind of assessment against the other.

Several studies have examined different international language proficiency tests in order to examine their comparability. Two of these studies have focused on test score data using a third test variable to compare their respective tests. Geranpayeh (1994) compared the IELTS and TOEFL tests by using the test scores of aspiring Iranian PhD candidates who had taken the IELTS, TOEFL and the Ministry of Culture and Higher Education (MCHE) test, an Iranian test developed to screen students before they are allowed to take IELTS or TOEFL. While TOEFL and IELTS embody two very different approaches to proficiency testing, Geranpayeh (1994) argues the IELTS and TOEFL exam provide comparable information about an examinee's English language ability regardless of the differences in their format. In a similar vein, Mok et al. (1998), in their comparison of IELTS and ACCESS as a measurement of migrant applicants' English proficiency, used a third assessment, the Australian Second Language Proficiency Ratings (ASLPR) test but for very different reasons. Their rationale for this approach was that they did not want to limit their study to a small sample of individuals by administering two tests to the same group of examinees. Instead, they decided to find examinees who had taken either IELTS or ACCESS in addition to ASLPR. Using Rasch modelling they were able to estimate the equivalence of the IELTS and ACCESS scales.

Taking a more comprehensive approach, Bachman et al (1995) examined the comparability of the Cambridge First Certificate in English (FCE) and TOEFL by investigating the test tasks, the language abilities required to complete them, and the relationships between examinee performance on the two tests. They concluded that the two tests were comparable in terms of the abilities being assessed.

While these studies compared tests which are already established and recognised, there appears to be an absence of empirical research about the comparability of standardised proficiency tests and curriculum-based achievement tests.

### The University of Melbourne English Language Bridging Program

Since late 2006 the University of Melbourne English Language Bridging Program (UMELBP) has been offered to prospective undergraduate and postgraduate (excluding PhD) students across the university whose IELTS scores (or equivalent) fell just short of the required entry levels and whose faculty was willing to recognise the program for entry purposes. At the time this study was undertaken in 2007, a range of faculties at the university had chosen to recognise the UMELBP for entry to all or some of their courses including Economics & Commerce, Education, Architecture, Building & Planning, Arts, Engineering, Music, and Science.

The UMELBP is delivered as a 10-week (250 hours) full time course by Hawthorn-Melbourne, an English language centre affiliated with the University of Melbourne. In terms of students' English proficiency, the program caters for two groups of students: those with 1) an overall IELTS score of 6.5 where 7.0 is the required entry level for their course; and 2) an overall score of IELTS 6.0 where 6.5 is the required entry level for their course. In order to enter their chosen university course, students in the first group must achieve at least 70% and those in the second group must achieve at least 65% on the UMELBP. However, the equivalence between these UMELBP exit grades and their corresponding IELTS university entry scores (i.e., 70% = 7.0 and 65% = 6.5) has not yet been empirically established.

Formal assessment on the UMELBP is comprised of continuous classroom-based tests and other assessments (worth 70%) and a written examination (worth 30%). The aims of this study were to a) benchmark the results of the final UMELBP written examination

---

against the IELTS Academic Writing module; and b) in so doing to provide evidence that the University of Melbourne's English language entry standards are upheld by the UMELBP.

### The UMELBP final examination versus the IELTS Academic Writing tasks

Of the various assessments used throughout the UMELBP course, the final examination corresponds most closely to the IELTS Academic Writing module in so far as both assessment procedures are timed, impromptu tests requiring students to complete two broadly similar tasks: a descriptive task and an argumentative essay task. The UMELBP examination also provides a sound point of comparison with the IELTS exam because it is administered at the end of the course when students are most likely to be able to demonstrate significant gains in English proficiency. However, there are important differences between the IELTS Academic Writing sub-test and UMELBP test in terms of task design, as seen in Table 1 below.

	<b>Time</b>	<b>Words</b>	<b>Input</b>	<b>Task</b>
UMELBP Task A	60 min	Approx. 450 - 500	Introduction to lecture Reading text	Descriptive/ Synthesis Task
IELTS Writing Task 1	20 min	At least 150 words	Graphs, charts, diagrams or maps	Descriptive/ Information Transfer Task

**Table 1. Differences of task design between IELTS writing sub-test and UMELBP test.**

	Time	Words	Input	Task
UMELBP Task B	45 min	Approx. 300 words	Statement from reading passage	Argumentative/ Discursive Task
IELTS Writing Task 2	40 min	At least 250 words	Essay task question	Argumentative/ Discursive Task

**Table 1. (continued) Differences of task design between IELTS writing sub-test and UMELBP test.**

In the UMELBP final examination, the test takers first listen to the beginning of a short lecture. After listening to this micro-lecture, they read a passage that raises some of the issues mentioned in the lecture, but in a different order and from a different perspective. In UMELBP Task A, the test takers are required to write a text identifying and reporting those issues or ideas which are discussed in both the lecture and reading passage. In the UMELBP Task B, they are asked to write an argument based on a statement about the reading passage and support their stance using examples from the reading passage.

In the IELTS Writing Task 1, the examinees are required to organize, present, and compare data from graphs, charts, diagrams or maps provided in the test. The examinees provide a descriptive account of the phases, stages, or procedures within a process and describe the interrelationships or mechanisms within that process. In IELTS Writing Task 2, the examinees are required to present and justify an opinion, present a solution to a problem, or evaluate and challenge arguments or ideas, but without reference to any specific listening or reading text.

The differences between the UMELBP examination and the IELTS Academic Writing module can be explained by the UMELBP examination's explicit aim to closely resemble the tasks the students will subsequently be required to perform in their academic studies:

two of the major sources of information for university students are lectures and readings of textbooks and articles, and one of the main attributes of academic writing tasks is the student's ability to effectively refer to different sources, quote, cite and analyse them.

Although the tasks on the two tests are therefore different in certain respects, they target the same two macro-functions, description and argumentation. Moreover the two tests are designed to measure the same construct i.e., readiness to study in higher education, and more specifically, to write in an academic context where English is the medium of instruction.

### **Methodology**

Although the study was designed to be conducted in three distinct stages, the stages were conceptually interrelated as the findings from each stage were designed to feed into its subsequent stage. This approach necessitated minor changes and several unanticipated methodological decisions in the design as the study led from one stage into another.

#### *Stage 1*

Stage 1 of the UMELBP benchmarking project was carried out in a session held at the University of Melbourne. For the first stage of the study four retired accredited IELTS examiners (referred to as IEs in this paper) participated in the study. Retired IELTS examiners were recruited for the study as IELTS Australia does not permit current IELTS examiners to undertake IELTS assessments other than those they employ them to do. Their years of experience in this role varied between two and thirteen years. All of them had ceased working as IELTS examiners for at least a year and had not previously taught on the UMELBP program.

This stage began with a short re-familiarisation session in which the IELTS Examiners scored and discussed sample IELTS Academic Writing Task 1 and Task 2 responses from the official IELTS Sample

---

Materials (Cambridge ESOL, 2005) using the publicly available criteria and descriptors for this part of the test (IELTS, 2007). The two tasks are scored using four equally weighted criteria: for Task 1 the criteria are Task Achievement, Coherence and Cohesion, Lexical Resource, and Grammatical Range and Accuracy. The criteria for Task 2 are Task Response, Coherence and Cohesion, Lexical Resource, and Grammatical Range and Accuracy. The overall score for the Academic Writing module is obtained by applying a one-third weight to the band score from Task 1 and a two-thirds weight from Task 2 (O'Neill et al., 2007).

The IEs were then shown a recent version of the UMELBP examination. In this version students were required to listen to an audio-recording of a lecture and to read a written text about the destruction of rainforests. They were asked to make notes on the lecture and then read the other text. They were asked to complete two tasks: UMELBP Task A (worth 60%) which requires the examinees to compare and contrast the two texts and UMELBP Task B (worth 40%) which asks them to critically examine the point of view expressed in the reading passage.

The IEs independently rated seven scripts from this version of the examination which had been used at the end of a recent course. These scripts were chosen to represent a range of abilities on the basis of the scores previously assigned to them by two of the teachers at Hawthorn-Melbourne. The names of test takers were removed from the scripts and were identified in the marking and discussion solely by the candidates' ID numbers. This was important because the two teachers who had marked them previously were also participating in the current study as UMELBP assessors in Stages 2 and 3. The aim therefore was to reduce the possibility of their assessments being influenced by preconceived ideas about the candidate or memory of their examination performance.

The IEs were given the scripts one at a time for marking. They assigned IELTS scores (in whole numbers) to the candidate's response

for each task using the publicly available IELTS Writing Task 1 criteria and descriptors to score UMELBP Task A and the IELTS Writing Task 2 criteria and descriptors to mark UMELBP Task 2. They were also asked to assign an overall score (in increments of 0.5) for each task as an unweighted average of their scores on each of the four IELTS criteria used to rate the two tasks.

After marking candidate responses to both tasks, the scores from all IEs were displayed on an overhead projector by the researcher so that IEs could see the level of agreement existing between members of the group for each examination script. Where there were instances of low agreement, individual IEs were asked to reflect and comment on their assigned scores. This procedure was followed for all seven scripts. Consensus scores were then elicited for all criteria and the overall score on each task. This sometimes involved compromise amongst members of the group but it came as the result of careful discussion of the relevant script.

Finally, IEs were asked to agree on a final group score (in increments of 0.5) for each script overall taking into account the relative weighting of 60% for UMELBP Task A and 40% of UMELBP Task B. This varied from the IELTS Academic Writing module where, as previously noted, Task 2 carries twice as much weight as Task 1.

### *Stage 2*

Stage 2 of the study was conducted at Hawthorn Melbourne. The participants were four teachers from this centre (referred to as the HTs in this paper) who had either already taught or co-ordinated the UMELBP.

The participants were asked to mark the same seven UMELBP scripts (both Tasks A and B) used in Stage 1 with the UMELBP criteria and (See Table 2 and 3). Levels of performance for these criteria were accompanied by descriptors.

IELTS Task 1	Task Achievement	Coherence & Cohesion	Lexical Resource	Grammatical Range and Accuracy
UMELBP Task A	Comprehension: listening text, reading text, Synthesis of ideas	Structure	Language	

**Table 2. Comparison of initial scoring criteria between IELTS Task 1 and UMELBP Task A.**

At this stage for Task A (worth 60% of the total examination mark) Comprehension of the listening text carried 10%, Comprehension of the reading text 10%, Synthesis of ideas, 10%, Structure 10% and Language 20%

IELTS Task 2	Task Response	Coherence & Cohesion	Lexical Resource	Grammatical Range and Accuracy
UMELBP Task B	Analysis of arguments	Structure	Language	

**Table 3. Comparison of initial scoring criteria between IELTS Task 2 and UMELBP Task B.**

At this stage for Task B (worth 40% of the total examination mark) Analysis of arguments carried 20%, Structure 10% and Language 10%.

The HTs were asked to do the marking in two steps. First, they marked each of the scripts (UMELBP Tasks A and B). Each script was discussed after it had been marked and eventually an overall

consensus score for each script (the same process to the one described above for the IELTS examiners in Stage 1) was reached. The HTs were then informed about the group overall IELTS score assigned to the script by IEs. Further discussion then ensued about the comparability of the UMELBP examination and the IELTS Academic Writing Module in terms of tasks and scoring criteria.

Finally, the HTs undertook to revise the scoring criteria, levels of performance and descriptors and to use them in the next UMELBP examination prior to Stage 3 of the study.

### *Stage 3*

Stage 3 of the study was carried out at the Parkville Campus of the University of Melbourne. For the third stage of the study, the four IELTS examiners from Stage 1 and the four Hawthorn Teachers from Stage 2 all participated. The scripts used at this stage made up a sample of six scripts obtained from students who had just completed the same version of the UMELBP examination used in Stages 1 and 2. As noted above, the examination was now rated using revised scoring criteria, levels of performance and descriptors. Both the IEs and HTs carried out the marking of these scripts. The IEs marked the scripts using the IELTS writing criteria and descriptors used previously in Stage 1. The HTs assessed the scripts using the UMELBP criteria and descriptors revised at the end of Stage 2. The two groups of IEs and HTs were asked to have in-group discussions prior to discussing their assigned scores with the other group. This provided each group with the opportunity to reflect closely on their scores before comparing and discussing them with the second group. For the HTs it also enabled them to evaluate the effectiveness of the modifications made to the UMELBP tasks, criteria and descriptors.

## Results and Discussion

### Stage 1

In the discussions of their marking of the scripts several issues were raised by the IEs. These points dealt with the UMELBP examination and its tasks, as well as which IELTS writing criteria and descriptors were possibly less applicable to marking UMELBP scripts.

After some discussion amongst the IEs it was agreed that UMELBP Task A and B could be viewed as broadly similar to the IELTS Writing Task 1 and 2 notwithstanding the differences between them. The main point of perceived similarity was the functional demands made on test-takers: the first task in both tests required test-takers to write a detailed description of the stimulus material provided and the second task required them to write an argumentative essay.

There was strong debate about the degree to which it was appropriate to use the IELTS writing criteria Task Achievement (IELTS Writing Task 1) and Task Response (IELTS Writing Task 2) to mark the UMELBP examination. While the IEs agreed that the main issue in both instances was that the descriptors for these two IELTS criteria did not fully address the extent to which candidates effectively incorporated information from the listening and reading texts used in the UMELBP examination, especially Task Response for IELTS Task 2 where the task input is very limited. However, it was concluded that the extent to which candidates effectively did this could become part of the assessment of these two criteria in the UMELBP examination. This would include how well they are able to incorporate the material from the texts including author attribution as well as using direct quotations and paraphrasing.

The other contentious IELTS criterion was Lexical Resource as applied to both UMELBP Task A and UMELBP Task B. Several IEs remarked that this was a problematic criterion since candidates were able to use and perhaps over-rely on vocabulary found in the

---

listening and reading texts. After some discussion it was decided that students should be rewarded for using the vocabulary from the texts appropriately (but not excessively) in their responses. It was agreed that this was a matter of individual assessor judgment.

Through Stage 1 of the study, it became clear that the IELTS writing criteria and descriptors could be applied to UMELBP examination reasonably well, although the extent to which the tasks were comparable and therefore the degree to which the IELTS criteria and their descriptors could be used was continually debated during the moderation session. The scores assigned to the seven scripts in Stage One are shown in Table 4 below.

### **Stage 2**

Stage 2 of the study aimed firstly, to determine whether the IELTS scores could be matched with the UMELBP assessments of the same scripts made by the HTs and secondly, to specify the required modifications in the UMELBP test and criteria for a higher level of alignment with IELTS if necessary.

Table 4 shows the scores assigned to the seven scripts by the HTs together with those assigned by the IEs in Stage 1 for the same scripts.

Script No.	Stage 1 Marking by IELTS Examiners	Stage 2 Marking by Hawthorn Teachers
101	6.5	72%
102	5.0	50%
103	6.5	75%
104	5.0	53%

**Table 4. Comparison scores from Stage 1 and Stage 2 Marking**

---

Script No.	Stage 1 Marking by IELTS Examiners	Stage 2 Marking by Hawthorn Teachers
105	5.0	53%
106	7.0	76%
107	6.5	65%

---

**Table 4. (continued) Comparison scores from Stage 1 and Stage 2 Marking**

As can be seen in Table 4, the level of agreement between the scores from Stages 1 and 2 is reasonable although the HTs scores are higher for five of the scripts. The three key performance thresholds of -6.5/65%, +6.5/65% and +7.0/70% can be distinguished although scripts 101 and 103 are both problematic because of the lack of alignment in relation to these levels. Discussion after the marking of these particular scripts indicated that the tendency of the HTs to mark them higher than the IELTS score assigned by the IEs could be attributable to the criteria in the UMELBP examination being strongly weighted towards assessing the ability of the students to comprehend and make effective use of the content from the listening and reading texts. For both UMELBP Tasks A and B, 50% of the total mark was allocated to directly or indirectly assessing the use of this content. Scripts 101 and 103 in particular were considered by the group to be stronger in terms of "content" than "language use" (as measured by the three IELTS writing criteria of Cohesion and Coherence, Lexical Resources and Grammatical Accuracy in both tasks of the UMELBP examination). By comparison, the IELTS writing criteria are more heavily language focused and therefore scripts which were stronger in terms of content than language were considered likely to score higher when marked by the HTs using the UMELBP criteria. This issue highlighted the need to align the criteria of the IELTS and UMELBP examinations in the interests of establishing their comparability.

---

In the Stage 2 follow-up discussion on the use of the criteria and descriptors in the UMELBP examination, three issues were raised.

Firstly, it was suggested that the distinction between the criteria of Comprehension (of the Listening and Reading text) and Synthesis was not clear. The second issue concerned the descriptors for the criterion, Language. Two HTs believed that they found it difficult to distinguish between “few errors”, “some errors” and “many errors” as given in the UMELBP descriptors. A third issue was related to the difficulty of deciding on scores when more than one score has the same descriptor in each of the criteria. For example, both “5” and “6” for Synthesis in UMELBP Task A were described as “Issues or ideas partially identified with some development of the relationships.” The majority of HTs believed this affected the precision of their ratings.

There was a strong commitment amongst the HTs to retaining the integrative tasks used in the UMELBP examination. Accordingly, it was agreed that these should remain unchanged. Instead, it was decided that the UMELBP scoring criteria should be aligned more closely with the IELTS Academic Writing module in order to achieve greater comparability. However, the descriptors for Task Achievement (Task A) and Task Response (Task B) would need to be modified to make them more relevant to the UMELBP examination tasks.

Table 5 below shows the revised UMELBP scoring criteria against the IELTS Academic Writing criteria as agreed at the end of this session. The HTs agreed to develop the appropriate levels and descriptors for these criteria after the session.

UMELBP Task A	Task Achievement	Coherence and Cohesion	Lexical Resource	Grammatical Range and Accuracy
IELTS Writing Task 1	Task Achievement	Coherence and Cohesion	Lexical Resource	Grammatical Range and Accuracy
UMELBP Task B	Task Response	Coherence and Cohesion	Lexical Resource	Grammatical Range and Accuracy
IELTS Writing Task 2	Task Response	Coherence and Cohesion	Lexical Resource	Grammatical Range and Accuracy

**Table 5. Comparison of final scoring criteria between IELTS writing sub-test and UMELBP examination.**

For Task A in the UMELBP examination (still worth 60% of the overall mark) Task Achievement would carry 20%, Coherence and Cohesion, 20%, Lexical Resource 10%, and Grammatical Range and Accuracy 10%. For Task B (still worth 40% of the overall mark), Task Response, Coherence and Cohesion, Lexical Resource, and Grammatical Range and Accuracy would each carry 10%.

### Stage 3

In Stage 3 both the IEs and HTs rated six new scripts using the revised UMELBP criteria shown in Table 5 above. As shown in Table 6 below, there was an extremely close alignment between the two sets of scores apart from script 201, which was the first to be marked.

As in Stage 2 the two key performance thresholds of UMELBP 65% / IELTS 6.5 and UMELBP 70% / IELTS 7.0 can be clearly distinguished. However, the high discrepancy between the scores for script 201 was disconcerting. The HTs assigned a score of about 8% above the score assigned by IEs. As it was the first script to be marked it was agreed

---

that it should be re-marked after the scores for the other scripts had been discussed. The scores by the IEs did not change, but the HTs assigned lower scores to the scripts with the result that their score was much more closely aligned with the IEs' score. This adjustment to the scoring of script 201 appeared to be the direct result of the discussion of the other scripts, especially the use of the revised scoring criteria to rate them.

Script No	IELTS Examiners		Hawthorn Teachers	
	(1 <sup>st</sup> marking)	(2 <sup>nd</sup> marking)	(1 <sup>st</sup> marking)	(2 <sup>nd</sup> marking)
201	6.5	6.5	73%	67%
202	6.5		66%	
203	7.5		73.5%	
204	6.0		62%	
205	6.5		67%	
206	6.0		60%	

**Table 6. Scores from Stage 3 marking**

All of the participants agreed that the modifications made to the test tasks and the revisions made to the UMELBP criteria and descriptors were effective in improving the alignment between UMELBP and IELTS scores. They also expressed approval that this had been achieved without radically changing the UMELBP test. However, there were concerns about the clarity of the UMELBP descriptors, especially for Lexical Resources and Grammatical Accuracy. Further refinement of these descriptors was planned by the HTs.

### **Identifying key textual features**

The final step in Stage 3 was to identify features of texts at the two key levels of UMELBP 65% / IELTS 6.5 and UMELBP 70% / IELTS 7.0. To do this, pairs of IEs and HTs looked at scripts representing these two levels of performance.

For scripts at the level of UMELBP 65% / IELTS 6.5 the clear structuring of ideas in terms of introduction, use of signposting, paragraphing and a well-formed conclusion were considered important. In terms of language, it was agreed that the candidates' grammatical and lexical resources should allow for clearly expressed task responses with only minor errors. For the scripts at the UMELBP 70% / IELTS 7.0 level, stronger coherence and cohesion were expected resulting in a more natural flow and facilitating easy reading and communication. In addition, it was considered that these stronger scripts should demonstrate effective use of a wide range of sentence structures and vocabulary with only occasional errors. It was agreed that these features highlighted in the two texts could be used in further revising the descriptors.

Lastly, to improve the reliability of their marking, it was agreed by the HTs that the examination should be independently marked by an outside teacher not familiar with the students as well as the two class teachers engaged in teaching any one class.

### **Conclusion**

It is important to note that only a small number of raters and non-randomly sampled scripts were used in this study and that any conclusions can therefore only be tentatively drawn. This is a probable limitation of most locally-based benchmarking studies. Nevertheless, the findings in this study suggest that the UMELBP examination could be benchmarked against the IELTS Academic Writing module without compromising the integrative nature of the UMELBP test tasks. By bringing the scoring criteria of the two tests closer together, it was possible to retain the distinctive character of

the UMELBP examination which integrates listening, reading and writing skills in a manner resembling authentic academic work. Arguably, this is a potentially even more valid test than the IELTS exam which has separate tests for each of these skills.

The study also highlights the value of examining and, where necessary, modifying assessment practices in English pathway programs to achieve greater comparability with standardised proficiency tests. In the absence of common exit standards across these pathway programs, tests such as IELTS provide a benchmark against which the final student assessments in these programs can be evaluated. This kind of exercise may also raise the status of English pathway programs within and outside tertiary institutions. In the case of this study, the credibility of the UMELBP examination within the University of Melbourne was considerably strengthened as a result of the findings. Moreover, if similar studies were carried out in the context of other English pathway programs, the exit standards on these programs could potentially be referenced against each other. This seems a desirable goal given that pathway programs are now widely offered to prospective students by tertiary institutions in Australia and other countries as alternatives to further proficiency testing yet there is at present no means of comparing their exit assessments.

The benchmarking exercise reported in this study clearly needs to be followed up by further moderation sessions with Hawthorn teachers to ensure that the standards established in the study are maintained in subsequent UMELBP courses. This is particularly important given the rapid turnover of short-term contract staff working on a course which only runs periodically over the course of each year.

Finally, while this study has focused on the examination component of the UMELBP (worth 30% of the total course mark), there is a need to investigate the classroom assessment carried out by teachers during the course (worth 70% of the total course mark). This is the subject of a study currently being undertaken.

## Acknowledgments

I am grateful to Reza Tasviri and Melissa Deshayes for their assistance with the data collection and preparation of this article. I would also like to thank the raters who participated so professionally in this study.

## References

- Agosti, C. (2004). Six direct entry courses at Macquarie: Meeting students' various needs. *Proceedings of the 17<sup>th</sup> EA Educational Conference, October 28-30*, (pp. 28-30). Adelaide, Australia: English Australia.
- Bachman, L., Davidson, R., Ryan, K. & Choi, I. (1994). *An Investigation into the Comparability of Two Tests of English as a Foreign Language: The Cambridge-TOEFL Comparability Study*. Cambridge: Cambridge University Press.
- Banerjee, J. and Wall, D. (2006). Assessing and reporting performances on pre-sessional EAP courses: Developing a final assessment checklist and investigating its validity. *Journal of English for Academic Purposes*, 5(1), 50-69.
- Birrell, B. (2006). Implications of low English standards among overseas students in Australian universities. *People and Place*, 14(4), 53-65.
- Cambridge ESOL. (2005). *IELTS Sample Materials*. Cambridge: Cambridge University Press.
- Cizek, G.J. (1996). Standard-setting guidelines. *Educational Measurement*, 15(1), 13-21.

- 
- Cizek, G.J., & Bunch, M. (2007). *Standard Setting: A guide to establishing and evaluating performance standards on tests*. Thousand Oaks, CA: Sage Publications.
- Cruikshank, K. & Chen H. (2006). *Preparation and proficiency: co-requisites or opposing goals? A comparison of IELTS and internal EAP pathways to tertiary study in Australia*. Paper presented at the annual Applied Linguistics Association of Australia, Melbourne.
- Deakin, G. (1997). IELTS in context: Issues in EAP for overseas students. *EA Journal*, 15(2), 7-28.
- Geranpayeh, A. (1994). Are score comparisons across language proficiency test batteries justified? An IELTS-TOEFL comparability study. *Edinburgh Working Papers in Applied Linguistics*, 5, 50-65.
- IELTS. (2007). *Academic Writing Scoring Criteria and Descriptors*. Available from [http://www.ielts.org/researchers/score\\_processing\\_and\\_reporting.aspx](http://www.ielts.org/researchers/score_processing_and_reporting.aspx).
- Leask, B., Ciccarelli, A. & Benzie, H. (2003). Pathway to tertiary learning: A framework for evaluating English language programs for undergraduate entry. *EA Journal*, 21(1), 17-30.
- Mok, M., Parr, N, Lee, T. & Wylie, E. (1998). A Comparative Study of IELTS and ACCESS Test Results. In Wood, S. (ed.) *International English Language Testing System Research Reports*. Sydney: ELICOS Association.
- O'Neill, T.R., Buckendahl, C. W., Plake, B.S. & Taylor, L. (2007). Recommending a Nursing-Specific Passing Standard for the IELTS. *Language Assessment Quarterly*, 4(4), 295-317.

Turner, J. (2004). Language as academic purpose. *Journal of English for Academic Purposes, 3(2)*, 95-109.

Ziguras, C. (2007, February 14). Poor grasp of the medium. *The Australian Higher Education Supplement*, p. 25.