

Tracking international students' English proficiency over the first semester of undergraduate study

Authors: Ms Pamela Humphreys, Dr Michael Haugh, Dr Ben Fenton-Smith, Dr Ana Lobo, Dr Rowan Michael and Dr Ian Walkinshaw
Griffith University, Australia

Grant awarded: Round 16, 2010

Keywords: *International students, English language proficiency, IELTS, English language testing*

Abstract

This paper reports on a study exploring variation and change in language proficiency amongst international undergraduate students who had been identified as requiring English language support. Specifically, it investigates changes in IELTS scores and in students' perceptions of language proficiency in their first semester of study.

The study employed a concurrent mixed methods design in which quantitative and qualitative data were collected simultaneously and analysed separately before comparing results. Quantitative data was collected using an IELTS Academic test at the beginning and end of one semester, while qualitative data comprised two rounds of focus group interviews conducted in the same semester. Fifty-one participants undertook both IELTS tests. The initial round of focus groups was attended by 10 participants and the final round by 15.

This study found that the main improvement in proficiency as measured by IELTS was in Speaking. All four sub-scores of the Speaking test showed statistically significant gains: Fluency and Coherence and Pronunciation showed gains of almost half an IELTS band score and these were found to be highly statistically significant. There was little shift in Writing scores except in the sub-score of Lexical Resource and only marginal mean score gains in Listening and Reading. The study distinguished between low-scorers, mid-scorers and high-scorers. The low-scorers obtained significantly higher scores after one semester of study, perhaps reflecting the more rapid progress often made at lower levels of language proficiency, while the mean improvement amongst mid-scorers and high-scorers was not found to be statistically significant.

In investigating the relationship between IELTS scores and GPA, Listening and Reading were found to be strongly correlated with GPA in the first semester of study, while Speaking and Writing were not. Further investigation of correlation between their IELTS scores and GPAs found that this strong correlation between GPA and Listening and Reading was maintained in their second semester of study but not in their third semester. This finding points to a relationship between language proficiency test scores and academic achievement for students in their initial year of study, but primarily with the receptive macro-skills, which may have implications for setting entry requirements if borne out by larger studies.

Focus group data suggested that students did not have unrealistic expectations of academic study, even if their perceptions of their English proficiency did not always match their actual IELTS levels. Students were able to articulate a range of strategies they had developed to raise proficiency while at university, as well as a range of obstacles that hindered language development. A key finding in comparing the focus group data with the IELTS scores was that proficiency is a complex and contested notion.

Publishing details

Published by IDP: IELTS Australia © 2012.

This new online series succeeds *IELTS Research Reports Volumes 1–13*, published 1998–2012 in print and on CD. This publication is copyright. No commercial re-use. The research and opinions expressed are of individual researchers and do not represent the views of IELTS. The publishers do not accept responsibility for any of the claims made in the research.

Web: www.ielts.org

Author biodata

Ms Pamela Humphreys

Pamela Humphreys has held many roles in ELT over two decades and has worked in Europe, Asia and Australia. She is currently the Academic Manager of Language Testing and University Programs at Griffith English Language Institute where she manages the IELTS Test Centre and co-ordinates several credit-bearing language enhancement courses. Pamela is the author of the workbooks for the *IELTS Express* series (Cengage). She has worked for many years for Cambridge ESOL as a materials writer and as Assistant Principal Examiner for IELTS.

Dr Michael Haugh

Dr Michael Haugh is a Senior Lecturer in Linguistics and International English in the School of Languages and Linguistics at Griffith University. He has previously taught English and Japanese at schools and universities in Australia, New Zealand, Japan and Taiwan. Michael's research interests include pragmatics and intercultural communication. He has published numerous articles on topics in pragmatics, a number of edited volumes on face and politeness, and is the author of a number of forthcoming books including *The Pragmatics of the English Language* (Palgrave Macmillan, co-authored with Jonathan Culpeper), *Revisiting Communicative Competence* (Cambridge Scholars Press), and *Understanding Politeness* (Cambridge University Press, co-authored with Dániel Kádár).

Dr Ben Fenton-Smith

Dr Ben Fenton-Smith is a Lecturer in English at the School of Languages and Linguistics at Griffith University. Prior to this, he was Director of Research and Professional Development at the English Language Institute of Kanda University of International Studies in Japan (2006–10). He has published on applied linguistics in such journals as *Discourse and Society*, *ELT Journal*, *Journal of Academic Language and Learning*, and *The Language Teacher*.

Dr Ana Lobo

Dr Ana Lobo is an Associate Lecturer in English at the School of Languages and Linguistics at Griffith University. She has been involved in English and Spanish language teaching for over 10 years and is a Portuguese/English translator. Her research interests include tertiary student retention, second language teaching and learning and the practice of translation and interpreting in Australia. Her publications include *Will We Meet Again? Examining the Reasons Why Students are Leaving First Year University Courses*, *The International Journal of Learning* (2012).

Dr Rowan Michael

Dr Rowan Michael is a Lecturer in English in the School of Languages and Linguistics at Griffith University. He has previously taught English and Chinese at schools and universities in Australia, China and Kazakhstan. Rowan's research interests include language teacher training at a distance, language teaching and learning (Chinese and ESL/EFL), and interaction in online environments. His publications include *Developing Northwest China through Distance Education* (2010).

Dr Ian Walkinshaw

Dr Ian Walkinshaw is a Lecturer in English at the School of Languages and Linguistics at Griffith University, where he co-ordinates the English Language Enhancement Courses (ELEC) program. He has been involved in English language teaching and applied linguistics since 1994 as a teacher, curriculum designer and researcher, and has worked in Japan, Vietnam, Britain, New Zealand and Australia. His publications include *Learning Politeness: Disagreement in a Second Language* (Peter Lang, 2009).

IELTS Research Program

The IELTS partners, British Council, Cambridge English Language Assessment and IDP: IELTS Australia, have a longstanding commitment to remain at the forefront of developments in English language testing.

The steady evolution of IELTS is in parallel with advances in applied linguistics, language pedagogy, language assessment and technology. This ensures the ongoing validity, reliability, positive impact and practicality of the test. Adherence to these four qualities is supported by two streams of research: internal and external.

Internal research activities are managed by University of Cambridge English Language Assessment Examinations' Research and Validation unit. The Research and Validation unit brings together specialists in testing and assessment, statistical analysis and item-banking, applied linguistics, corpus linguistics, and language learning/pedagogy, and provides rigorous quality assurance for the IELTS Test at every stage of development.

External research is conducted by independent researchers via the joint research program, funded by IDP: IELTS Australia and British Council, and supported by Cambridge English Language Assessment.

Call for research proposals

The annual call for research proposals is widely publicised in March, with applications due by 30 June each year. A Joint Research Committee, comprising representatives of the IELTS partners, agrees on research priorities and oversees the allocations of research grants for external research.

Reports are peer reviewed

IELTS research reports submitted by external researchers are peer reviewed prior to publication.

All IELTS research available online

This extensive body of research is available for download from www.ielts.org/researchers.

Introduction from IELTS

This study by Humphreys and her colleagues was conducted with support from the IELTS partners (British Council, IDP: IELTS Australia, and Cambridge English Language Assessment) as part of the IELTS joint-funded research program. Research studies funded by the British Council and IDP: IELTS Australia under this program complement those conducted or commissioned by Cambridge English Language Assessment, and together inform the ongoing validation and improvement of IELTS.

A significant body of research has been produced since the joint-funded research program started in 1995, over 90 empirical studies having received grant funding. After undergoing a process of peer review and revision, many of the studies have been published in academic journals, in several IELTS-focused volumes in the *Studies in Language Testing* series (<http://research.cambridgeol.org/research-collaboration/silt>), and in *IELTS Research Reports*. To date, 13 volumes of *IELTS Research Reports* have been produced.

The IELTS partners recognise that there have been changes in the way people access research. In view of this, since 2011, *IELTS Research Reports* have been available to download free of charge from the IELTS website, www.ielts.org. However, collecting a volume's worth of research takes time, delaying access to already completed studies that might benefit other researchers. This has been recognised by other academic publication outlets, which publish papers online from the moment they are accepted, before producing them in print at a later time. As a natural next step therefore, individual *IELTS Research Reports* will now be made available on the IELTS website as soon as they are ready.

This first report under the new arrangements considers changes in the IELTS scores of undergraduate international students during their first semester of study at an Australian university, the students' own perceptions of their English language proficiency development, and the relationship of their test scores to academic outcomes. It forms part of a larger, longitudinal study being conducted by the university which will track such changes over several years of undergraduate study.

The contexts where language proficiency is a consideration and where language proficiency exams are required have always been of interest to IELTS. In this study, there is a dual context: a country, and a university operating within that country. In Australia there is intense scrutiny and debate regarding the language proficiency and other needs of international students, how these might be addressed, and to whom the responsibility should fall. Numerous reports, symposiums, reviews, and the establishment of a new national regulatory and quality agency – the Tertiary Education Quality and Standards Agency (TEQSA) – provide evidence of this. Humphreys et al provide an excellent overview of this context, which will be of interest not only to readers in Australia, but also to those in other countries that attract international students and where similar issues may exist.

Within the general context of Australia, the authors discuss the response of a particular university, which drew from the *Good Practice Principles for International Students in Australian Universities* (Arkoudis et al 2008) and adopted a whole-of-university approach to improving students' English language skills. The Griffith English Language Enhancement Strategy (GELES) is a comprehensive strategy that will be worth following in the longer term.

In the study, IELTS was used to track students' proficiency growth over time. It was found that, among the four skills, improvement was greatest in speaking and greatest for lower initial proficiency students, both findings perhaps not unexpected. IELTS results were also correlated with students' grade point averages, and a moderate correlation was found between Listening and Reading scores and academic outcomes.

This study thus contributes to the literature on students' growth in language proficiency (Craven 2012; Elder & O'Loughlin 2003; Green 2005; O'Loughlin & Arkoudis 2009; Storch & Hill 2008) and on the predictive validity of IELTS (Cotton & Conrow 1998; Dooley & Oliver 2002; Ingram & Bayliss 2007; Kerstjen & Nery 2000; Ushioda & Harsch 2011). While the outcomes of individual studies such as this one are understandably limited by their relatively short duration and sample sizes, the overall picture created by the studies taken together shows IELTS to be a fit for purpose exam in making admissions decisions.

Collaboration between the university, the research team, and IELTS was a key feature of this study. It was valuable as a means of achieving knowledge and understanding about the quality and usefulness of the IELTS test, the nature of language development and assessment, and measures to enhance the experiences and outcomes of international students. The study found a reality that is "a complex tapestry of multiple intersecting conceptualisations of proficiency and multiple underlying variables", which future studies will no doubt continue to consider.

MS JENNY OSBORNE **DR GAD LIM**
IELTS Research **Senior Research**
Coordinator **and Validation Manager**
IDP: IELTS Australia **Cambridge English**
 Language Assessment

December 2012

References to the IELTS introduction

Arkoudis, S, Atkinson, C, Baird, J, Barthel, A, Ciccirelli, A, Ingram, D, Murray, D & Willix, P, 2008, *Good practice principles for English language proficiency of international students in Australian universities, Final report*, Department of Employment, Education and Workplace Relations, Australian Government

Cotton, F & Conrow, F, 1998, 'An investigation into the predictive validity of IELTS amongst a group of international students studying at the University of Tasmania' *IELTS Research Reports Volume 1*, IELTS Australia Pty Limited, Canberra, pp72–115

Craven, E, 2012, 'The quest for IELTS 7.0: Investigating English language proficiency of international students in Australian universities', *IELTS Research Reports Volume 13*, IELTS Australia, Canberra and British Council, London, pp1–61

Dooley, P & Oliver, R, 2002, 'An investigation into the predictive validity of the IELTS Test', *Prospect*, 17(1), pp36–54

Elder, C & O'Loughlin, K, 2003, 'Investigating the relationship between intensive English language instruction and band score gain on IELTS', *IELTS Research Reports Volume 4*, ed R Tulloch, IELTS Australia Pty Limited, Canberra, pp207–254

Green, A, 2005, 'EAP study recommendations and score gains on the IELTS Academic writing test', *Assessing Writing*, 10, pp44–60

Ingram, D & Bayliss, A, 2007, 'IELTS as a predictor of academic language performance, Part 1', *IELTS Research Reports Volume 7*, IELTS Australia Pty and British Council, Canberra, pp137–199

Kerstjen, M & Nery, C, 2000, 'Predictive validity in the IELTS test', *IELTS Research Reports Volume 3*, IELTS Australia Pty Limited, Canberra, pp85–108

O'Loughlin, K & Arkoudis, S, 2009, 'Investigating IELTS exit score gains in higher education', *IELTS Research Reports Volume 10*, ed J Osborne, IELTS Australia, Canberra and British Council, London, pp1–90

Storch, N & Hill, K, 2008, 'What happens to international students' English after one semester at university?' *Australian Review of Applied Linguistics*, 31(1), pp04.1–04.17

Ushioda, E & Harsch, C, 2011, *Addressing the needs of international students with academic writing difficulties: Pilot project 2010/11, Strand 2: Examining the predictive validity of IELTS scores*, retrieved from <http://www2.warwick.ac.uk/fac/soc/al/research/groups/ellt/a/projects/strand_2_project_report_public.pdf>

TABLE OF CONTENTS

INTRODUCTION FROM IELTS	3
1 BACKGROUND AND CONTEXT	6
1.1 Institutional context.....	7
2 LITERATURE REVIEW: PREDICTIVE VALIDITY STUDIES AND SCORE GAIN STUDIES	8
2.1 Focus	8
2.2 Methodology.....	8
2.3 Instruments.....	8
2.4 Sample size.....	8
2.5 Stage and duration of studies	8
2.6 Proficiency of graduating students.....	8
2.7 Investigating score gain.....	9
2.8 Absence of score gain	9
2.9 Academic outcomes and proficiency test score	9
2.10 Macro-skill as predictor of academic success	10
2.11 Other variables impacting academic performance	10
2.12 Literature summary.....	11
3 RESEARCH AIMS	11
4 METHODOLOGY	11
4.1 General approach.....	11
4.2 Quantitative methodology	12
4.2.1 The use of IELTS.....	12
4.2.2 Procedure.....	12
4.2.3 Participants.....	13
4.2.4 Analysis.....	13
4.3 Qualitative methodology.....	14
4.3.1 Rationale for focus groups.....	14
4.3.2 Study participants	14
4.3.3 Procedure.....	14
4.3.4 Questions asked and their rationale.....	14
4.3.5 Data analysis.....	15
5 FINDINGS	15
5.1 Variation and change in IELTS scores over one semester of undergraduate study.....	15
5.1.1 Test 1 and Test 2 descriptive statistics	15
5.1.2 Variation in IELTS scores across Tests 1 and 2.....	16
5.2 The nature of change in IELTS scores across Tests 1 and 2	20
5.2.1 Principal Components Analysis (PCA)	20
5.2.2 Correlation and regression analysis.....	21
5.3 The relationship between IELTS scores and academic achievement.....	22
5.4 Students' views on their English language learning experiences	24
5.4.1 Macro-skills	24
5.4.2 Micro-skills.....	29

6	DISCUSSION	31
6.1	Proficiency change over initial semester of study	31
6.2	Variation in English language proficiency	32
6.3	English language proficiency and academic achievement	33
6.4	Students' views of their English language experiences	34
6.5	Students' perceptions of proficiency compared with proficiency as shown by IELTS	35
6.6	Limitations	35
7	CONCLUSION	36
	ACKNOWLEDGMENTS	36
	REFERENCES	37
	APPENDIX: Focus Group Interview Protocol	40

List of tables and figures

Table 1:	Pathway into university	12
Table 2:	Nationality breakdown	13
Table 3:	Academic group	13
Table 4:	Bachelor degree	13
Table 5:	Means and standard deviations for IELTS scores at Test 1 and Test 2	15
Table 6:	Results of paired t-test for difference in mean scores for Speaking at Test 1 and 2	15
Table 7:	Cross-tabulations of Listening Test 1 and Test 2	16
Table 8:	Cross-tabulations of Reading Test 1 and Test 2	16
Table 9:	Cross-tabulations of Writing Test 1 and Test 2	17
Table 10:	Cross-tabulations of Speaking Test 1 and Test 2	17
Table 11:	Cross-tabulations of Overall score Test 1 and Test 2	17
Table 12:	Paired t-test for Fluency and Coherence	19
Table 13:	Paired t-test for Pronunciation	19
Table 14:	Paired t-test for Lexical Resource	19
Table 15:	Paired t-test for Grammatical Range and Accuracy	19
Table 16:	Paired t-test for Lexical Resource (Writing Task 2)	20
Table 17:	Unrotated factor matrix	20
Table 18:	Rotated factor matrix (Varimax rotation with Kaiser normalisation)	20
Table 19:	Total variance explained by two-factor model	21
Table 20:	Correlations between improvements in Speaking, Writing, Reading and Listening	21
Table 21:	Correlations between Test 1 scores and improvement across Tests 1 and 2	21
Table 22:	Correlations between Test 1 scores and first semester GPA including ELEC	22
Table 23:	Correlations between Test 2 scores and first semester GPA including ELEC	22
Table 24:	Correlations between Test 1 scores and first semester GPA excluding ELEC	22
Table 25:	Correlations between Test 2 scores and first semester GPA excluding ELEC	22
Table 26:	Correlations between Test 1 scores and second semester GPA	23
Table 27:	Correlations between Test 2 scores and second semester GPA	23
Table 28:	Correlations between Test 1 scores and third semester GPA	23
Table 29:	Correlations between Test 2 scores and third semester GPA	23
Figure 1:	Good Practice Principles for International Students in Australian Universities	6
Figure 2:	Griffith English Language Enhancement Strategy	7
Figure 3:	Scattergram of Fluency and Coherence subscores for Tests 1 and 2	18
Figure 4:	Scattergram of Pronunciation for Tests 1 and 2	18
Figure 5:	CEFR and IELTS comparison	33

1 BACKGROUND AND CONTEXT

As the third largest source of export income in Australia (COAG, 2010), the value and contribution of the international education sector to Australia's economy is considerable. In 2010–11 over half a million international students studied in Australia and contributed \$16.3 billion to the economy (Australia Unlimited, 2012) with 44% of approved visas being granted for study in the university sector (Australia Bureau of Statistics, 2011; Australian Education International, 2012). Although there are now indications of a slowing market, the success of international student recruitment has already led to a focus on standards and stronger outcomes for international students both during their degree and, increasingly, at the point of graduation. The level of English language proficiency of international students in particular has been the subject of intense debate over the past few years amongst academics, higher education policy drivers and in General Skilled Migration policy. In the media, international students have been painted as a source of contempt for their perceived lack of adequate English language skills (Devos, 2003) yet simultaneously (and paradoxically) regarded as valuable in relieving the financial pressures facing Australian universities. The tension between quantity and quality is both evident and increasing.

Many of the current changes in higher education policy in Australia in relation to international students can be traced back to the findings presented by Birrell, Hawthorne and Richardson (2006) in their report to the Department of Immigration and Citizenship, which raised serious concerns about the language standards of international students not only when they gain entry to Australian tertiary institutions but also when they graduate from them. Following this report, a National Symposium was organised by Australian Education International (AEI) and the International Education Association of Australia (IEAA) (Hawthorne, 2007). Since then, there have been moves to implement positive changes

in both policy and practice in Australian universities in regards to addressing issues of English language development amongst international students. In part, this movement has arisen in response to the publication of the Good Practice Principles for English Language Proficiency of International Students in Australian Universities (see Figure 1) in a report to the Department of Employment, Education and Workplace Relations (DEEWR) (Arkoudis, Atkinson, Baird, Barthel, Ciccarelli, Ingram, Murray & Willix, 2008). The report outlined the need to tackle such issues at an institutional level, forcing universities to examine their support practices and entry requirements due to “the potential to compromise English standards in terms of academic entry, progression and exit” (Hawthorne, 2007, p23). The principles put the responsibility for ensuring adequate language skills squarely on the institution from the point of enrolment to graduation as well as on the student.

At around the same time, the Review of Australian Higher Education was published (Bradley, Noonan, Nugent, & Scales, 2008). It also noted that ongoing language support should be provided to international students and integrated into the curriculum.

Rather than a blueprint or a ‘silver bullet’ to solve all issues related to language proficiency, the Good Practice Principles have been described as ‘a starter gun’ and a launching point for discussions aimed at substantial change (Barrett-Lennard, Dunworth, & Harris, 2011, pA-99). Since the publication of these principles, universities in Australia have attempted to amend perceived deficiencies in language abilities in a number of ways, many arguably assuming that the principles were in fact a blueprint and a set of standards to be met, as opposed to principles of best practice. In 2010, the Australian Universities Quality Agency (AUQA) steering committee recommended that six of the principles be converted into standards for all students (not only international students) and for all providers (not solely universities) under the title English Language Standards for Higher Education (Barthel, 2011).

1. Universities are responsible for ensuring that their students are sufficiently competent in the English language to participate effectively in their university studies.
2. Resourcing for English language development is adequate to meet students' needs throughout their studies.
3. Students have responsibilities for further developing their English language proficiency during their study at university and are advised of these responsibilities prior to enrolment.
4. Universities ensure that the English language entry pathways they approve for the admission of students enable these students to participate effectively in their studies.
5. English language proficiency and communication skills are important graduate attributes for all students.
6. Development of English language proficiency is integrated with curriculum design, assessment practices and course delivery through a variety of methods.
7. Students' English language development needs are diagnosed early in their studies and addressed, with ongoing opportunities for self-assessment.
8. International students are supported from the outset to adapt to their academic, sociocultural and linguistic environments.
9. International students are encouraged and supported to enhance their English language development through effective social interaction on and off campus.
10. Universities use evidence from a variety of sources to monitor and improve their English language development activities.

Figure 1: Good Practice Principles for International Students in Australian Universities

The decision on the conversion to standards was delayed due to the federal election and due to the commencement of a new national regulatory and quality agency, TEQSA (Tertiary Education Quality and Standards Agency), and so the sector awaits further developments in this area. In the meantime, the quest to ensure compliance should the standards be implemented continues and interest in empirical investigation in this area is increasing.

1.1 Institutional context

Griffith University is a large university in South-East Queensland, Australia. It appears in the Top 400 of the Academic Ranking of World Universities (ARWU) and the university ranks among Australia's top 10 research universities (according to the Excellence in Research for Australia 2010 evaluation www.griffith.edu.au/research/research-services/research-policy-performance/excellence-research-australia). Of its 44,000 students, approximately one quarter are international (Griffith Fast Facts, 2012), a percentage which is not unusual within the Australian context. In 2010, the university introduced a whole-of-university approach known as the Griffith English Language Enhancement Strategy (GELES) to address the Good Practice Principles and to further develop English language skills throughout the course of students' studies (see Figure 2). Its introduction represents a significant investment of resources by the university. As a strategy, the GELES aims to enhance English language support for international students or domestic students from a non-English-speaking background. The five-strand strategy comprises the following optional and compulsory elements.

1. Griffith UniPrep: a voluntary three-week intensive academic language and literacy course delivered before semester to international students.
2. English Language Enhancement Course (ELEC): a compulsory course for international students who enter with an Overall IELTS score (or equivalent) less than 7.0 or via a non-test pathway.
3. English HELP (Higher Education Language Program): free additional English language

support via one-to-one consultations and group workshops.

4. StudentLinx: opportunities for international students to interact with local students and the local community, and to establish social and intellectual ties across languages and cultures.
5. IELTS4grads: a subsidy for international students to take an IELTS Academic test on completion of a degree at the university.

A key component of the strategy is the introduction of a compulsory credit-bearing discipline-specific English Language Enhancement Course (ELEC). The course is designed to be completed by students in their first semester of study on entry into either the first or second year of their program. In this way, the entire GELES as a strategy aims to:

- provide English support to international students in their first semester of study
- ensure international students understand their responsibilities in continuing to develop their English language competence throughout their degree
- provide immersion experiences that encourage integration between domestic, 'native' English language speakers and international students
- demonstrate that Griffith's international students graduate with strong English language competence. (Fenton-Smith (2012) provides an example of how the Good Practice Principles have been applied within ELEC.)

The present study is therefore positioned within a large Australian university comprising a high number of international students at a time when there is considerable scrutiny of language proficiency not only at entry but at other key stages of university degrees. The literature has long recommended that institutions conduct their own studies concerning the link between English proficiency levels and academic success and that they make their own decisions about acceptable English language proficiency levels (Dooley, 1999; Graham, 1987). The Good Practice Principles and the Bradley Review have served as the most recent national catalysts for monitoring English language proficiency of international students while the GELES provides the institutional imperative.

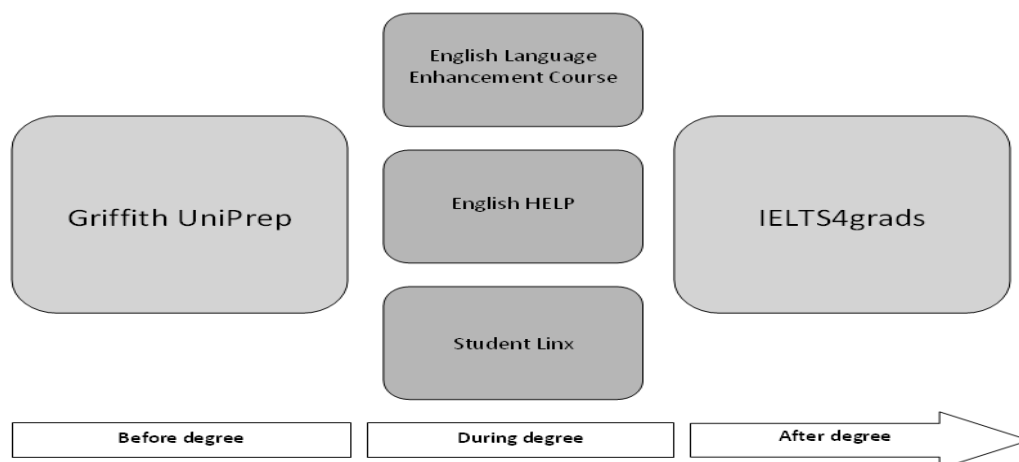


Figure 2: Griffith English Language Enhancement Strategy

2 LITERATURE REVIEW: PREDICTIVE VALIDITY STUDIES AND SCORE GAIN STUDIES

2.1 Focus

This report investigates English language proficiency change over the first semester of undergraduate study at an Australian university using the IELTS Academic test and the relationship of this proficiency test score to academic success as measured by Grade Point Average (GPA). The literature review therefore focuses on both score gain studies and predictive validity studies. Score gain studies investigate the degree of shift between two points in time using the same testing instrument at both points, whereas predictive validity studies administer a test to attempt to predict something about future behaviour such as academic success. This review demonstrates the scope of such studies by comparing the methods, instruments, sample size and stage and duration of the study. It then progresses to highlight the range of findings from these studies including: proficiency score gain, academic performance and specific macro-skills as predictors of academic success. Issues that have been commonly noted, such as the difficulty of measuring proficiency change of international students in a tertiary context, are also examined.

2.2 Methodology

Principal studies have utilised a variety of methods depending on the focus of the research and the data collected. As proficiency is often measured via test scores, and academic success is often measured in terms of GPA, it is not surprising that most studies are quantitative. Indeed many have relied solely on quantitative data (Allwright & Banerjee, 1997; Archibald, 2001; Avdi, 2011; Cotton & Conrow, 1998; Dooley, 1999; Dooley & Oliver, 2002; Feast, 2002; Green, 2005; Humphreys & Mousavi, 2010; Light, Xu, & Mossop, 1987; Read & Hayes, 2003). Mixed methods studies are becoming increasingly common, as they allow investigation beyond score gains or correlation, through which researchers might better explain any relationship or change (Craven, 2012; Elder & O'Loughlin, 2003; Ingram & Bayliss, 2007; Kerstjen & Nery, 2000; O'Loughlin & Arkoudis, 2009; Rose, Rose, Farrington, & Page, 2008; Storch & Hill, 2008; Woodrow, 2006). Some researchers have interviewed staff as well as students to highlight other factors that contribute to score changes and/or academic success (Ingram & Bayliss 2007; Kerstjen & Nery, 2000; O'Loughlin & Arkoudis, 2009; Woodrow, 2006).

2.3 Instruments

Within the literature there is great variability in the application of test instruments, both in the range of tests chosen and also the way individual instruments are used and results analysed. Most studies tracking score gains or investigating language proficiency as a predictor for academic success have used IELTS Academic as the testing instrument. However, the full test is not always administered and in some cases live tests are not used (Archibald, 2001; Green, 2005; Read & Hayes, 2003).

One large study (n=376) investigating proficiency tests as a predictor of academic success with international students in the USA used TOEFL (Light et al, 1987) while Storch and Hill's study (2008) used DELA (Diagnostic English Language Assessment) developed at the University of Melbourne. This can make the comparability of findings challenging.

2.4 Sample size

Studies to date into the predictive validity of tests for academic success or investigating score gains have varied considerably in sample size. Sample sizes range from 17 pre-sessional students (Read & Hayes, 2003) to 2594 university students (Cho & Bridgeman, 2012). Typically, studies use between 40 and 100 participants and most are reliant on non-probability convenience sampling.

2.5 Stage and duration of studies

There is also considerable variation in the stage that the study is undertaken (for example: pre-sessional - ie intensive language courses prior to tertiary study known as ELICOS in Australia; over first semester at university; over the entire university degree) and in the duration of studies (from one month to over three years). Some studies have focused on score gains in pre-sessional IELTS preparation or EAP courses and therefore predominantly on lower level learners (Archibald, 2001; Elder & O'Loughlin, 2003; Green, 2005; Read & Hayes, 2003). English language behaviour in the university context has also been a focus of numerous studies, many concentrating on change over one semester (Avdi, 2011; Ingram & Bayliss, 2007; Kerstjen & Nery, 2000; Light et al, 1987; Storch & Hill, 2008; Woodrow, 2006). Others have focused on one academic year (Cotton & Conrow, 1998; Dooley, 1999; Dooley & Oliver, 2002; Feast, 2002; Ushioda & Harsch, 2011). While some of these studied undergraduates (Craven, 2012; Dooley, 1999; Dooley & Oliver, 2001; Kerstjen & Nery, 2000), others concentrated on postgraduates (Allwright & Banerjee, 1997; Avdi, 2011; Light et al, 1987; Storch & Hill, 2008; Ushioda & Harsch, 2011; Woodrow, 2006) and some comprised both undergraduate and postgraduate cohorts (Cotton & Conrow, 1998; Feast, 2002; Humphreys & Mousavi, 2010; O'Loughlin & Arkoudis, 2009).

2.6 Proficiency of graduating students

Increasingly, there has been a focus on testing proficiency at the point of graduation, particularly in Hong Kong (Berry & Lewkowicz, 2000; Qian, 2007) and Australia (Craven, 2012; Humphreys & Mousavi, 2010; O'Loughlin & Arkoudis, 2009) though only the latter two studies trace proficiency changes over an entire university degree. There is currently a paucity of literature investigating language proficiency change across an entire degree program especially in English-speaking higher education contexts. This may be due to the challenges of longitudinal research or because graduating proficiency and what happens to language ability during degrees is a relatively recent focus. Other research literature has also begun to discuss the issue of graduating proficiency.

Barrett-Lennard, Dunworth and Harris (2011) argue that insufficient consideration has been given to the levels of English language proficiency of graduates and that “few measures are in place to ensure that graduating students have attained a level of proficiency that employers will accept” (p103), while Benzie (2010) has called for wider perspectives on the debate of language proficiency in higher education, citing access to adequate levels of language experience during degrees to ensure improved language communication skills among graduates.

2.7 Investigating score gain

Many studies investigate score change over time using test re-test methods. Green's (2005) retrospective study of over 15,000 test scorers who had taken the test more than once showed “considerable individual variation in rate of gain” (p58). Elder and O'Loughlin (2003), Storch and Hill (2008) and Craven (2012) also demonstrate strong variability with some students making no progress at all between pre- and post-testing even over entire degrees. It has been found that proficiency gains are not linear and that “improvements seen in mean scores do not apply equally at all band levels” (Green, 2005, p11). Studies consistently show that the lowest scorers on an initial test improve most by post-test and that the highest scorers at pre-test increase the least or even regress by post-test. Green (2005), for instance, found that:

Candidates with Writing scores at band 5 or below at Time 1 tended to improve their results at Time 2. Those obtaining a band 7 or 8 at Time 1 tended to receive a lower score at Time 2, while those starting on band 6 tended to remain at the same level. (p57)

Arkoudis and O'Loughlin (2009) concur and suggest that this may be due to regression to the mean or because language acquisition occurs more easily at lower levels of proficiency. Band 6 is described as a threshold or plateau level beyond which it is hard to progress (Craven, 2012; Elder & O'Loughlin, 2003; Green, 2004). Green (2005) claims, for instance, that:

...if a student obtains an IELTS Writing band score of 6 on entry to a two-month pre-session (200h) English course, then takes an IELTS test again at course exit, they are more likely to obtain a score of 6 again than to advance to band 7. (p58)

Elder and O'Loughlin (2003) also found that at band 6, candidates have less than a 50% chance of increasing while those below 5.5 saw measurable improvement. Green (2005) suggested that the L1 background of candidates may have an effect.

Some of the above studies occurred before half band scores were awarded for Speaking or Writing. However, even with all four macro-skills being reported with increased degrees of granularity, Craven (2012) argues that stakeholders need to be aware of how difficult it is to progress to band 7 and above.

Pre-session studies that have investigated score gains have not always been via live or complete IELTS tests. Read and Hayes (2003) report only the average improvement on the Reading, Writing and Listening components and did not test Speaking. They found an increase of 0.35 of a band (from 5.35 to 5.71) following one month of instruction but the gains were not found to be statistically significant. Archibald (2001) focused only on writing and found discourse argumentation and organisation (the two most genre-specific criteria) increased most. Elder and O'Loughlin (2003), using a live IELTS test, found the average amount of improvement to be 0.5 of an overall IELTS band. However, the median increase was zero on Writing and Speaking whereas it was 0.5 for Listening and Reading. University-level studies seem to concur with the latter finding. The small number of studies tracing score gains over the course of a degree show the greatest gains in Reading and Listening and the least in Writing, though not all students improved (Arkoudis & O'Loughlin, 2009; Craven, 2012).

2.8 Absence of score gain

The absence of score gain does not necessarily indicate that improvement has not occurred. Storch and Hill (2008) posit that the increase may not be large enough to be captured. Green (2005) also suggests that tests such as IELTS are “not designed to be sensitive to relatively limited short-term gains in ability or to the content of particular courses of instruction” (p58). Elder and O'Loughlin (2003) propose that Standard Error of Measurement may better account for score gain or lack thereof. All three points highlight the complexity in examining score gain.

Storch and Hill (2008), using DELTA, found different outcomes when measured against “discourse measures” (fluency, accuracy and complexity) compared to proficiency measures (fluency, content, form). No statistically significant difference was found between pre- and post-test on the discourse measures whereas on proficiency measures it was found to be statistically significant. They attribute these different outcomes to the “collapsing” of features within criteria on proficiency scales in which more than one area of language is judged within a single criterion yet only one score is awarded.

2.9 Academic outcomes and proficiency test score

The literature presents contradictory findings as to whether English language ability measured by proficiency tests is a predictor of academic success (Cotton & Conrow, 1998; Graham, 1987). Some have found little or no correlation between test score and Grade Point Average (GPA). Craven (2012), for instance, identified no clear predictor of which students will (or will not) improve their proficiency during their degree while Cotton and Conrow (1998) stated that no positive correlations were found between IELTS scores and academic outcomes.

Some studies have shown that those allowed entry to university despite scoring below the cut-off obtain low academic scores (Ushioda & Harsch, 2011) but others have found that such students did not fare worse over one semester than those who had exceeded the minimum requirement (Dooley, 1999; Fiocco, 1992, as cited in Dooley 1999; Light et al, 1987).

Many studies show some degree of a correlation between test scores and academic outcomes as measured by GPA. Ushioda and Harsch (2011) found a highly significant correlation between coursework grades of postgraduates in various disciplines and their Overall IELTS scores used for entry (n=95) and also that IELTS Overall scores and IELTS Writing scores best predicted academic coursework grades, explaining over 33% of the variance in academic coursework grades. Yet many other studies have evidenced weak predictive validity. A study of 376 students in the USA using TOEFL scores showed weak predictive validity for GPAs and concluded that commencing test scores were not an effective predictor, though there was higher correlation for humanities, arts or social science majors than for those studying science, maths or business (Light et al, 1987).

Cho and Bridgeman in their large-scale study of 2594 students in the US found the correlation between TOEFL iBT and GPA was not strong but concluded that even a small correlation might indicate a meaningful relationship. Kerstjen and Nery (2000) and Feast (2002) found a significant positive but weak relationship between the English language proficiency of international students and their academic performance. Woodrow (2006), on the other hand, identified weak but significant correlations between IELTS and GPA in postgraduate Education students, especially in Writing and Listening while Elder (1993) found that the strongest predictor of language proficiency and academic outcomes occurred where students were scoring band 4.5.

As indicated above, studies to date show that there are inconsistencies in finding strong correlation between language proficiency scores and academic performance. Ingram and Bayliss (2007) argue that "it is not surprising that attempts to correlate test scores with subsequent academic results have been inconsistent in their outcomes" (p5) because IELTS predicts *language behaviour* in academic contexts not academic performance. Not only is measuring language proficiency change difficult, but as Woodrow (2006) points out, academic achievement is a complex issue.

2.10 Macro-skill as predictor of academic success

Of the four macro-skills (Reading, Writing, Listening, and Speaking), the two receptive skills of Reading and Listening have generally been shown to have correlation to academic success. Kerstjen and Nery (2000), for example, found Reading to be the only significant predictor of academic success. In the same study, academic staff felt Reading (and Writing to a lesser extent) should be given special

consideration in the selection process of international students. Reading has been cited by others as the macro-skill that best predicts academic success, though often the correlation is only moderate (Avdi, 2011; Cotton & Conrow, 1998; Dooley & Oliver, 2002; Rose et al, 2008). Other studies found Listening to be a useful predictor (Elder & O'Loughlin, 2003; Woodrow, 2006) while Ushioda and Harsch (2011) found a highly significant correlation between IELTS Writing as well as Reading and coursework grades.

2.11 Other variables impacting academic performance

A large number of studies posit that variables beyond language are likely to contribute to success at university. Kerstjens and Nery (2000), for example, concluded that less than 10% of academic performance may be attributed to English proficiency as measured by IELTS. According to Ingram and Bayliss (2007), it is "impossible to account for all the variables" (p5) and language is an *additional* variable. Arkoudis and O'Loughlin (2009) termed these additional variables "enabling conditions" and cite agency, language socialisation, language support and contact with other English speakers outside of university classes. Motivation/agency is considered a key factor (Avdi, 2011; Cotton & Conrow, 1998; Craven, 2012; Elder & O'Loughlin, 2003; Ingram & Bayliss, 2007; Kerstjen & Nery, 2000; Light et al, 1987; O'Loughlin & Arkoudis, 2009; Rochecouste, Oliver, Mulligan & Davies, 2010; Woodrow, 2006).

Sociocultural factors, cultural adjustment and the need for intercultural skills are also regularly cited (Briguglio, 2011; Cotton & Conrow, 1998; Fiocco, 1992, as cited in Dooley & Oliver, 2002; Ingram & Bayliss, 2007; Kerstjen & Nery, 2000; Rochecouste et al, 2010). The use of English outside of class is important (Cotton & Conrow, 1998; Craven, 2012; Elder & O'Loughlin, 2003; O'Loughlin & Arkoudis, 2009) and even the language background of others in class may have an impact as, according to Storch and Hill (2008), students need "an input-rich environment to improve" (p4.13). In some cases age was cited as a factor (Avdi, 2011; Craven, 2012).

Some have therefore cautioned against using quantitative scores alone for admission to university (Allwright & Banerjee, 1997; Dooley, 1999; Green, 2005; O'Loughlin, 2011), arguing that multiple sources of evidence of students' abilities should be sought. The English Language Learning and Teaching Council found that academic success is linked to a plethora of variables, of which learning strategy use and affective variables represent just a few (Rochecouste et al, 2010, p2). In summary, then, the claim made by Criper and Davies (1998) below appears to be generally shared in the research literature:

Language plays a role but not a major dominant role in academic success once the minimum threshold of adequate proficiency has been reached. Thereafter it is individual non-linguistic characteristics, both cognitive and affective, that determine success. (p113)

While a minimum threshold appears key (Elder, Erlam, & Von Randow, 2002), what this threshold *is* remains a contested and contentious issue.

It is also important to note that recruitment is frequently an issue in longitudinal studies attempting to examine international student proficiency change. Craven (2012) contacted 2000 potential participants but only found 48 that were eligible, ultimately testing only 40 participants. O'Loughlin and Arkoudis (2009) were required to extend their recruitment period and, after a second round of recruitment, had a sample of 63. Attrition in longitudinal studies exacerbates these initial recruitment issues, tending to result in relatively small sample sizes.

2.12 Literature summary

In summary, there is considerable variation in previous studies in terms of methodology, instruments, sample size, stage and duration and findings. However, there are some consistent findings: score gains are not guaranteed even over an entire degree; lower scorers generally see the greatest improvement while reaching high bands is challenging; there is some evidence of weak positive correlations between language proficiency and academic success; reading and listening both appear to be key in terms of score gains over time and as the more likely predictors of academic success. However, there are many variables which appear to impact proficiency change and thus such findings need to be treated with due caution.

3 RESEARCH AIMS

This paper reports on Phase 1 of a larger study investigating language proficiency change among undergraduate international students. Phase 1 encompasses quantitative and qualitative data collected over the first semester of university study while Phase 2 will track longer-term changes over the entire degree. It is expected that the greatest gains will be seen by the end of Phase 2. This study is unique in that it focuses solely on international undergraduate students identified as requiring early linguistic intervention in their first semester of study. It therefore addresses key points raised at the 2007 National Symposium organised by AEI and IEAA. The project attends to Hawthorne's 2007 call for the "[d]evelopment of more effective mechanisms to audit students' English language entry and academic progression standards" (2007, p6) through measuring the impact of an intervention strategy. Relatively little research has been conducted into the issues related to language standards of international students at graduation or regarding changes in language proficiency during degree courses (Arkoudis & Starfield, 2007; O'Loughlin & Arkoudis, 2005). This study will start to address this shortfall in the research literature.

Phase 1 of the broader study focuses on collecting and analysing quantitative and qualitative data on international students who have completed one semester at university and who have experienced both optional and compulsory strands of GELES, to ascertain any changes in actual IELTS scores or in students' perceptions of language proficiency and self-efficacy. The quantitative data included both pre- and post-semester IELTS Academic test scores that were analysed by macro-skill and sub-score (where available). The qualitative data consists of nine focus group sessions (four held at the beginning of semester and five held at the end), which were analysed for changes in students' views about their English proficiency.

The Research Aims addressed in Phase 1 of the study are:

1. To measure change in English language proficiency over one semester of international students at Griffith University using the IELTS test.
2. To explore variation in language proficiency of initial semester students at Griffith University using the IELTS test.
3. To investigate the correlation between language proficiency as shown through IELTS test scores and overall academic outcomes as measured by GPA.
4. To explicate commencing students' views on their English language learning experiences over one semester.
5. To investigate similarities and differences between students' perceptions of learning English for university study and their language proficiency as shown through IELTS test scores.

4 METHODOLOGY

4.1 General approach

The study employed a concurrent mixed methods design, ie, comprising quantitative and qualitative data collected simultaneously and in which the datasets are analysed separately before the results are compared (Creswell, 2008). Mixed methodology research has been said to provide an improved understanding of the research problem and better inferences through its breadth and depth (Tashakkori & Teddlie, 2003). In our approach, the different datasets and modes of analysis are treated as offering distinct interpretive windows on English language proficiency issues amongst international students. Neither approach is prioritised over the other. Thus, in making comparisons we are looking for synergies between the two approaches rather than seeking to use one to "explain" the other.

Quantitative data was collected using the IELTS Academic test in addition to participants' academic results in the form of cumulative Grade Point Averages (GPA), along with background data on the participants. Qualitative data comprised two rounds

of focus group interviews. Data collection therefore comprised pre-semester scores (referred to as Test 1) and focus group data at the beginning of semester followed by post-semester scores (referred to as Test 2) and focus group data from the end of semester. The rationale for the two-pronged assessment tool approach is to mitigate the fact that the IELTS test is “not sensitive enough to indicate language preparedness for specialised uses” (Hirsh, 2007, p206), while the focus group data alone is unable to provide reliable evidence of the participants’ level of proficiency in English.

The study took place between June and November 2010. Ethical approval was gained from the Griffith University Human Research Ethics Committee before recruitment. Participation was limited to students who were required to take the English Language Enhancement Course (ELEC) in Semester 2, 2010. This course is compulsory in the first semester of study at the university for those entering via a pathway program in which no formal test is required or for those who enter with a proficiency test score of less than IELTS 7.0 (or equivalent according to university admissions policies). This study therefore employs purposeful sampling by targeting those deemed by the university to have the weakest language skills at the commencement of their undergraduate studies.

4.2 Quantitative methodology

4.2.1 The use of IELTS

IELTS Academic was used as the test of English language proficiency. IELTS is a high stakes international language proficiency test. In Australia it is used in particular to provide evidence of English language proficiency for entry to educational institutions, for professional registration (eg teachers, nurses, engineers) and for migration. According to the IELTS website, 1.7 million candidates accessed the test in 2011 alone and its use is increasing (www.ielts.org).

IELTS was chosen as a measure of English language proficiency in this research project for a variety of reasons. Firstly, IELTS is an external commercially-available proficiency test which provides standardised reporting that is understood by a range of stakeholders. This also removes any conflict of interest that might be inherent in an institution-specific test. Secondly, the IELTS test, as a highly regarded and widely-accepted test internationally, makes the setting of benchmarking standards comparatively straightforward. Finally, the IELTS test also has the advantage of currency and usefulness for the test-taker, which is particularly important for the recruitment of volunteers for the end of degree testing which is part of the broader project. Finally, IELTS is considered the preferred test (Coley, 1999) and, in some cases, is the only test accepted for post graduation purposes.

4.2.2 Procedure

Prior to semester, students were recruited from three entry pathways, ie, test or non-test means by which students evidence meeting the language condition to enter the university. Participants were recruited from the two key non-test pathway providers into the university (named ‘Pathway 1’ and ‘Pathway 2’) and by inviting students who had met the language entry condition via IELTS to participate (named ‘Proficiency Test’).

Initially the intention was to recruit a minimum of 30 students from each of the three pathways to provide a cohort of around 90. However, recruitment proved difficult as the majority of students targeted had already met the conditions for entry to the university and did not require a proficiency test score. To increase the number to a viable cohort, it was agreed that the IELTS subgroup would not be required to take Test 1 but rather that verified scores submitted for entry to the university would be utilised. One test score in this batch was 12 months old but most were less than five months old. Seventy-three students were recruited from the three pathways. However, between Test 1 and Test 2 considerable attrition occurred with 22 opting not to return for Test 2, thus reducing the cohort to 51 participants as shown in Table 1 below. The pathway with the greatest attrition (Pathway 1) suggests that they undertook Test 1 as a contingency for entry to the university before final grades were known in their pathway program (Test 1: n=31; Test 2: n=19).

Pathway	Total
Proficiency test	17
Pathway 1	15
Pathway 2	19
Total	51

Table 1: Pathway into university

Students were provided with IELTS application packs and ethics-approved documentation and invited to apply at the Griffith IELTS test centre following standard procedures. Research candidates were integrated into the public candidature for two test dates in July (Test 1) and two test dates in November (Test 2) 2010. IELTS examiners were not informed that research candidates were being tested at any point. By way of incentive for taking part in Test 1 and Test 2, students were offered free pizza at each test, a \$30 supermarket voucher for test completion, and a guaranteed free IELTS test at graduation (Test 3) for Phase 2 of the longitudinal study. Each IELTS test was free of charge to the participants.

4.2.3 Participants

The initial group of 73 participants comprised 57 females and 16 males. Due to attrition, 41 females and 10 males undertook both tests and the report focuses on these 51 participants. Of the participants, 55% were Chinese with the remainder predominantly from Asia as can be seen in Table 2.

Country	Total
China	28
Hong Kong	4
Japan	1
Korea	5
Mongolia	1
Philippines	1
Sri Lanka	3
Taiwan	4
Ukraine	2
Vietnam	2
Total	51

Table 2: Nationality breakdown

Tables 3 and 4 summarise information regarding the academic groups and specific degrees in which the cohort was enrolled. Students were mainly from the Business/Commerce group (90%) with only five from other academic groups. A range of degrees were represented though predominantly accounting/finance and hotel management. They were all undergraduate students who were at the beginning stages of undergraduate degrees. Thus, although the cohort of participants could not be claimed to be strictly representative of all international students since the participants themselves volunteered, these proportions are broadly in line with the numbers of students enrolled in ELEC from different elements of the university.

Group	Total
Business/Commerce	46
Health	4
SEET (Science, Environment, Engineering and Technology)	1
Total	51

Table 3: Academic group

Degree	Total
Accounting	13
Accounting & Finance	1
Banking & Finance	1
Business	1
Commerce	1
Event management	4
Finance	1
Graphic Design	1
Hotel Management	10
Human Resource Management	1
International Business	3
International Tourism & Hotel Management	7
Management	1
Marketing	1
Nursing	2
Psychological Science	1
Psychology	1
Sustainable Enterprise	1
Total	51

Table 4: Bachelor degree

4.2.4 Analysis

The IELTS test provides a score for each macro-skill (Listening, Reading, Writing and Speaking) as well as an Overall score calculated from the average of the four macro-skills. (Note that where macro-skills are capitalised in this report, they refer to the papers of the IELTS test.) IELTS test scores for Test 1 and Test 2, Grade Point Averages (GPA) of the same semester and relevant biodata were entered into SPSS for statistical analysis. All scores were exported from the IELTS test centre database except for the Proficiency Test Pathway Test 1 scores, which were extracted from the university database. Biodata and GPA data were also extracted from the university database.

Descriptive statistics were first explored for Test 1 and Test 2. Cross-tabulations and paired t-tests were carried out to investigate possible change between Test 1 and Test 2. This was followed by multivariate analysis, correlation and regression analysis, and factor analysis that were used to investigate the nature of the changes and variation in scores between Tests 1 and 2. To analyse test scores compared to academic outcomes, correlations were tested between Test 1 and GPA for the same semester and Test 2 and GPA for the same semester. In both cases correlations between GPA scores were tested (both including the grade for ELEC and excluding it). Tests for correlation were also carried out to compare the IELTS test scores acquired with the GPA of the two subsequent semesters.

4.3 Qualitative methodology

4.3.1 Rationale for focus groups

Focus groups were used to understand participants' self-perceptions of their productive and receptive English proficiency during the research period, and to provide detail about perceived shifts in their linguistic development over one semester. Although other methods (eg one-to-one interviews) were considered for this purpose, focus groups were selected because a greater volume of data could be gathered at one time, and because participants could "comment on each other's point of view, ...challenging each other's motives and actions" (Kidd & Parshall, 2000, p294) in a way that might alienate interviewees if employed in a dyadic interview format.

The advantages of focus groups were weighed against their shortcomings. Being a group discussion forum, focus groups may stifle individual dissenting voices (Kitzinger, 1996). Additionally, too much moderator control can inhibit discussion, while too little control can result in the topic not being discussed in sufficient detail (Agar & MacDonald, 1995). This issue was managed in the present study by constructing a framework for questioning (see Section 4.3.4) and ensuring that all discussion was grounded in this framework.

4.3.2 Study participants

The participants in the focus group sessions were international students taking ELEC. They represented all four strands of ELEC – Business and Commerce; Health; Science, Environment, Engineering and Technology; and Arts and Social Sciences.

Ten respondents attended the initial round of focus groups: six females and four males, from China (4), Taiwan (2), Russia (1), Brazil (1), Hong Kong (1) and Turkey (1). Fifteen participants attended the final round of focus groups – from China (5), South Korea (3), Taiwan (2), Hong Kong (2), Russia (1), Israel (1) and Brazil (1). There were seven females and eight males.

4.3.3 Procedure

Focus group sessions were held at the beginning and end of a 13-week semester between July and October 2010. The initial round of focus groups comprised four sessions, while the final round consisted of five sessions. Sessions lasted approximately one hour and were digitally recorded.

Each focus group session had between one and four participants, as well as an interviewer. The interviewers were members of the research team or postgraduate students. No interviewer was directly involved in teaching the participants' courses. Participants were given a free movie voucher as compensation for attending and were provided with refreshments during the focus group sessions.

4.3.4 Questions asked and their rationale

Before all focus group sessions, the core aims were explained to the participants. The participants were informed that all data would be de-identified, and that the interview would have no bearing on their grades for any course. They were also given the opportunity to ask questions before each session commenced.

In the initial round of focus groups, participants were asked their nationality, the length of time they had been in Australia, their major subject and which English macro-skills they found easiest and most challenging. In order to elicit information about language-related issues they faced at an English-medium university, participants were asked whether they believed their English proficiency was good enough for studying in Australia. They were also asked about the importance of continuing to increase their English proficiency while they were studying at university, and whether they viewed this as their own responsibility or that of the institution. These questions solicited their initial perceptions about language self-study and about the value of the university's English language enhancement resources, including English HELP and StudentLinX (described in Section 1.1), and academic literacy workshops.

Lastly, the participants were asked to list some of the factors both within and outside the university which had either enhanced or inhibited their English proficiency (eg accommodation with Australian speakers of English or with other speakers of the participants' L1).

The questions asked in the final round of focus groups explored the participants' use of the various English language resources available at Griffith University for increasing their English language proficiency and helping them adapt to university life. Another line of questioning investigated whether participants believed their English ability had increased during the semester, complementing the quantitative findings gained through IELTS testing. Participants' views on the importance of continuing to increase their English proficiency – and whether they or the university were responsible for this – were solicited a second time for comparison with the initial data.

Finally, the participants' opportunities for speaking or listening to English outside the university since the beginning of the semester were elicited to provide further illustrative detail about the contexts in which the participants were using English. Their expectations about their future linguistic development were also elicited. The focus group interview protocols for both stages can be found in the Appendix.

4.3.5 Data analysis

The discussions were transcribed and the transcriptions were then coded for common themes by four members of the research team using NVivo 8 qualitative data analysis software. A coding scheme was first devised by the team based on coding and discussion of one of the transcribed texts. Subsequently, each of the other texts was coded twice using the agreed framework. The double-coding was 'blind', ie, neither coder had knowledge of the other's work. The two coders later compared their results and, where opinions differed, consensus was reached through discussion. Consequently, the overall degree of agreement between the four coders was 98.15%.

5 FINDINGS

The results of this study are organised into four main sections. The first two sections (5.1 and 5.2) summarise changes in IELTS test scores across one semester and discuss variation in language proficiency at entry (Research Aims 1 and 2). The third section (5.3) explores correlations with academic achievement as measured by GPA (Research Aim 3). The fourth section (5.4) summarises the findings from focus groups where international students discussed their views on learning and using English across their first semester of study (Research Aim 4). The qualitative and quantitative findings are compared in the Discussion section (6.5) that follows (Research Aim 5).

5.1 Variation and change in IELTS scores over one semester of undergraduate study

The current minimum score at Griffith University required to meet the language condition for undergraduate study is an Overall minimum of IELTS 6.0 with no macro-skill below 5.5. The IELTS scores of the 51 undergraduate student participants at the beginning (Test 1) and end (Test 2) of their first semester of study at Griffith University are reported in Table 5, including their scores in Listening, Reading, Writing, Speaking and Overall. All scores are expressed from 1.0 to 9.0 in increments of 0.5. We start by reporting changes in Overall means before exploring variation in these scores, as well as variation in changes in these scores between Tests 1 and 2. We then focus on changes in subscores for Speaking where the most change was observed between Tests 1 and 2, as well as subscores for Writing where significant shifts were not picked up by the combined score.

Paired samples test

	Paired differences					t	df	Sig. (2-tailed)
	Mean	Std Deviation	Std Error	95% Confidence Interval of the Difference				
				Lower	Upper			
Pair 1 Speaking 2 – Speaking 1	.363	.794	.111	.139	.586	3.262	50	.002

Table 6: Results of paired t-test for difference in mean scores for Speaking at Test 1 and 2

5.1.1 Test 1 and Test 2 descriptive statistics

Table 5 shows the means and standard deviations for each macro-skill and Overall scores for Test 1 and Test 2. The means for all five scores were higher in Test 2 (ie, end of semester) than in Test 1 (ie, beginning of semester) suggesting some level of overall improvement. The highest mean in Test 1 was for Listening (6.196) and it remained the highest at Test 2 (6.216), although it saw the smallest change amongst the four macro-skills. Reading was the second lowest at both Test 1 and Test 2 while Writing scored the lowest mean on average on both tests. Speaking saw the greatest upward shift and equalled Listening as the highest mean in Test 2. High standard deviations were observed for Listening and Reading at both Test 1 and Test 2.

It can also be seen that on average all four macro-skills and the Overall score increased, though these were mostly marginal increases. There were very marginal increases for Listening of 0.02, Reading of 0.069 and Writing of 0.029. The greatest increase in mean between Tests 1 and 2 was in the scores for Speaking (0.36) and there was a marginal increase in mean Overall scores of 0.107.

	Test 1 mean	Test 1 SD	Test 2 mean	Test 2 SD
Listening	6.196	.959	6.216	.945
Reading	5.784	1.016	5.853	.808
Writing	5.706	.460	5.735	.483
Speaking	5.853	.658	6.216	.808
Overall	5.971	.569	6.078	.568

Table 5: Means and standard deviations for IELTS scores at Test 1 and Test 2

Paired t-tests were performed for each macro-skill and the Overall score to ascertain whether the within-subject degree of shift from Test 1 to Test 2 was statistically significant, with 2-sided tests used for this initial assessment looking at any level of significant change. Only the increase in Speaking was found to be statistically significant at the 0.05 level for the two-sided t-test ($t=3.262, df=50, p=0.002$) (see Table 6). Increases for Listening, Reading and Writing were neither large nor statistically significant.

We subsequently conducted one-sided paired t-tests to test for statistically significant levels of improvement. On a one-sided t-test, the improvement in mean score for Speaking was found to be highly significant ($t=3.262, df=50, p=0.001$).

5.1.2 Variation in IELTS scores across Tests 1 and 2

5.1.2.1 Cross-tabulations of IELTS test scores

Test 1 and Test 2 data were cross-tabulated to analyse variation in shifts for Listening, Reading, Writing, Speaking and Overall scores. There were no missing data ($n=51$), although some of the particularly low scores reported can be considered outliers (the reasons for these low-scoring outliers are considered further in the Discussion section). The cross-tabulations show that there is considerable within-subject variability, with the degree of shift being described below.

Listening scores

Table 7 shows that Listening scores ranged from 4.0 to 8.5 on Test 1 and 4.5 to 9.0 on Test 2 with the median at 6.0 for both tests. 18 participants scored the same in both Tests 1 and 2, with an equivalent number of participants having scores that went up as those that went down. However, the lowest scorers in each test were different participants, showing some unexpected within-subject movement.

Reading scores

Table 8 shows that Reading scores ranged from 1.0 to 7.5 on Test 1 and 4.5 to 8.0 on Test 2 with the median at 6.0 for Test 1 and 5.5 for Test 2. Nine participants scored the same in both Tests 1 and 2, with an equivalent number of participants obtaining scores that went up as those that went down. The lowest scorer in Test 1 scored considerably higher in Test 2 though a score of 1.0 indicates that this candidate barely attempted this paper, a point we will return to in the Discussion. Initial scores of 5.5 and below appeared quite likely to improve on Test 2. Interestingly, none of those who scored 6.5 on Test 1 scored 6.5 on Test 2, with six of the 12 obtaining higher scores and six obtaining lower scores in the second test.

	Listening 2										Total	
	4.5	5.0	5.5	6.0	6.5	7.0	7.5	8.0	8.5	9.0		
Listening 1 4.0			1									1
4.5		1										1
5.0		4	2	1								7
5.5	1	1	3	2	1							8
6.0		1	1	4	3	2						11
6.5				5		2						7
7.0			1	1	3	4			1			10
7.5				1			2					3
8.0							1			1		2
8.5								1				1
Total	1	7	8	14	7	8	3	1	1	1	51	

Table 7: Cross-tabulations of Listening Test 1 and Test 2

	Reading 2								Total	
	4.5	5.0	5.5	6.0	6.5	7.0	7.5	8.0		
Reading 1 1.0				1						1
4.5	2	2								4
5.0		4	2	1						7
5.5	1		4	5						10
6.0		1	7	2	1	1				12
6.5		1	2	3		4	1	1		12
7.0				1	2					3
7.5						1	1			2
Total	3	8	15	13	3	6	2	1	51	

Table 8: Cross-tabulations of Reading Test 1 and Test 2

Writing scores

Table 9 shows that in the Writing test, scores ranged from 5.0 to 7.0 on Test 1 and 4.5 to 7.0 on Test 2 with the median at 5.5 for Test 1 and median 5.5 for Test 2. Writing was the macro-skill that showed the narrowest range of scores. Again, there is an unusual backward movement for one candidate who scored 6.5 at Test 1 but 4.5 at Test 2, which suggests lack of motivation for Test 2 rather than true attrition of two bands. Overall, while 16 participants scored the same in Writing for both Tests 1 and 2, the scores of 20 participants went up while only 15 went down.

Speaking scores

Table 10 shows that Speaking scores ranged from 4.5 to 7.5 on Test 1 and 5.0 to 8.0 on Test 2 with the median score being 6.0 for both tests. Overall, while 12 participants scored the same in Speaking for both Tests 1 and 2, the scores of 28 participants went up while only 11 went down.

Overall scores

Table 11 shows that Overall scores ranged from 4.5 to 7.0 in Test 1 and 5.0 to 7.5 in Test 2. The median was 6.0 for both Test 1 and Test 2. Overall, while 21 participants scored the same Overall for both Tests 1 and 2, the scores of 19 participants went up while only 11 went down.

		Writing 2					Total
		4.5	5.0	5.5	6.0	6.5	
Writing 1	5.0	1	1	6	1		9
	5.5		1	7	8	1	17
	6.0		1	8	8	3	21
	6.5	1		1	1		3
	7.0					1	1
Total		2	3	22	18	5	51

Table 9: Cross-tabulations of Writing Test 1 and Test 2

		Speaking 2						Total
		5.0	5.5	6.0	6.5	7.0	7.5	
Speaking 1	4.5	1	1	1				3
	5.0			1	1	1		3
	5.5	3	5	6	2	2		18
	6.0	3		6	2	3	2	16
	6.5		1	0	1	1	2	5
	7.0		0	1	2			5
	7.5				1			1
Total		7	7	15	9	7	4	51

Table 10: Cross-tabulations of Speaking Test 1 and Test 2

		Overall 2					Total
		5.0	5.5	6.0	6.5	7.0	
Overall 1	4.5			1			1
	5.0		4	1			5
	5.5	1	4	4			9
	6.0	1	6	10	3	1	21
	6.5			3	4	4	11
	7.0					3	4
Total		2	14	19	7	8	51

Table 11: Cross-tabulations of Overall score Test 1 and Test 2

5.1.2.2 Improvement of lower scorers (5.5 and below)

The cross tabulations show that there is considerable within-subject variation both upwards and downwards. Additionally, it shows that initial lower scorers often improved most by Test 2 while higher scorers were more likely to receive a lower score in Test 2 than they obtained in Test 1. While this may be due in part to regression to the mean, it is also likely to be due to the fact that lower proficiency is easier to improve than “plateau” level (ie, IELTS 6.0) and higher level proficiency (ie, 6.5 and above). This is particularly evident in Listening and Speaking. In Listening, out of 17 low scorers in Test 1 (ie, 5.5 and below), eight improved on their initial score, seven obtained the same score while only two decreased in Test 2. For Speaking, 16 out of 24 low scorers in Test 1 improved by Test 2, five remained at the same level while only three decreased. A similar but less obvious pattern occurred with Reading where, out of 21 low scorers (ignoring the outlier), five improved their score, 11 remained the same and one regressed. The pattern continued with Writing; out of 26 low scorers, 16 increased their scores; eight remained the same while two obtained lower scores.

This shift upwards by initial low-scorers (5.5 and below) on the four macro-skills was found to be statistically significant using one-sided paired t-tests, testing for improvement between Test 1 and 2. The mean improvements in Listening ($t=1.852, df=16, p=0.041$) and Reading ($t=1.851, df=21, p=0.039$) were found to be significant on a one-sided t-test, although given the small sample size this should be treated with due caution. The mean improvements in Writing ($t=4.170, df=25, p=0.0005$) and Speaking ($t=4.071, df=23, p=0.0005$) were found to be very highly significant on a one-sided t-test, a result which is more robust given the slightly larger samples and the very high level of significance.

5.1.2.3 Improvement in Speaking subscores

The research team had access to the subscores for the participants who had taken both tests at the Griffith IELTS test centre (n=35). Further analysis was thus performed on Speaking by drilling down to the analytical subscores to explore these changes in more detail. IELTS Speaking is scored against four key indicators: Fluency and Coherence, Lexical Resource, Grammatical Range and Accuracy and Pronunciation. The mean scores for Fluency and Coherence and Pronunciation increased the most with improvement shown by the line of best fit in the scattergrams presented for each in Figures 3 and 4 respectively.

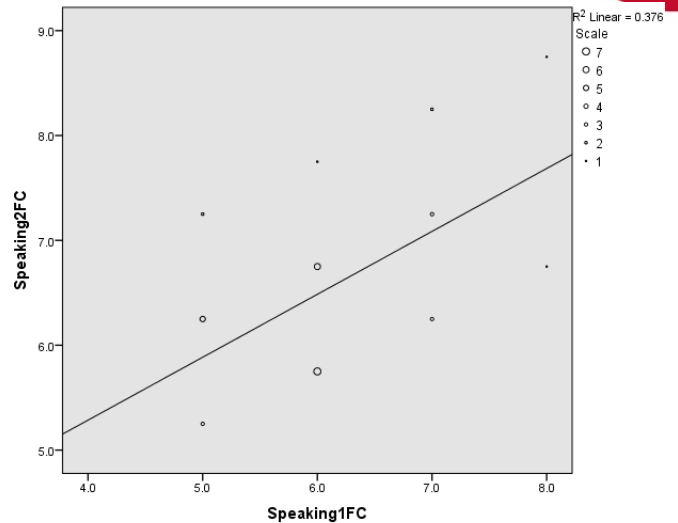


Figure 3: Scattergram of Fluency and Coherence subscores for Tests 1 and 2

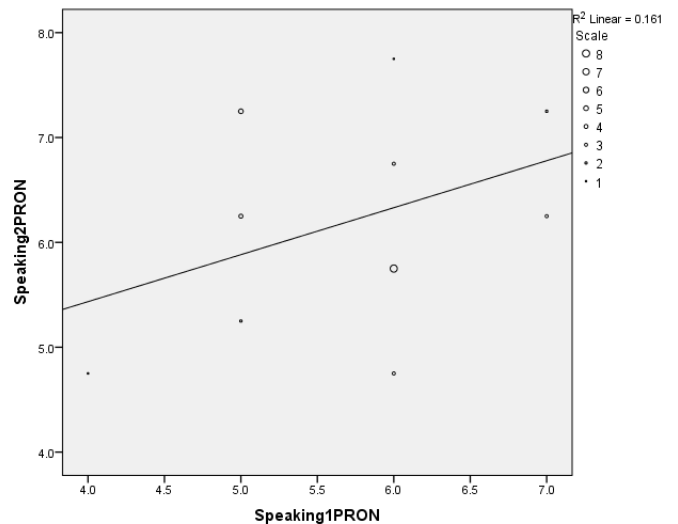


Figure 4: Scattergram of Pronunciation for Tests 1 and 2

These changes in mean subscores for Speaking were found to be statistically significant. Fluency and Coherence increased by 0.4857 of a band on average and was found to be highly significant on a one-sided test ($t=3.513$, $df=34$, $p=0.000$) as shown in Table 12.

Pronunciation increased by 0.45 of a band on average and this change was also found to be highly statistically significant ($t=2.758$, $df=34$, $p=0.005$) as shown in Table 13.

Lexical Resource increased by 0.285 of a band on average and was significant ($p=0.034$) on a one-sided t-test as shown in Table 14.

Grammatical Range and Accuracy increased by 0.314 of a band on average and was significant ($p=0.031$) on a one-sided test as shown in Table 15.

Overall, there were upward shifts in all four Speaking subscores. One-tailed pair-sample t-tests where these changes were treated as improvements found these upward shifts were very highly significant for Fluency and Coherence and Pronunciation, and significant for Lexical Resource and Grammatical Range and Accuracy. However, the shift was of greater statistical significance for Fluency and Coherence and Pronunciation than the subscores for Lexical Resource and Grammatical Range and Accuracy.

	Paired Differences					t	df	Sig. (1-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
Speaking2FC - Speaking1FC	.486	.818	.138	.205	.767	3.513	34	.000

Table 12: Paired t-test for Fluency and Coherence

	Paired Differences					t	df	Sig. (1-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
Speaking2PRON - Speaking1PRON	.457	.981	.166	.120	.794	2.758	34	.005

Table 13: Paired t-test for Pronunciation

	Paired Differences					t	df	Sig. (1-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
Speaking2LR - Speaking1LR	.286	.894	.151	-.0212	.595	1.892	34	.034

Table 14: Paired t-test for Lexical Resource

	Paired Differences					t	df	Sig. (1-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
Speaking2GRA - Speaking1GRA	.3143	.963	.163	-.017	.645	1.930	34	.031

Table 15: Paired t-test for Grammatical Range and Accuracy

	Paired Differences					t	df	Sig. (1-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
Writing2LR2 – Writing2LR	.143	.494	.0834	-.0267	.3124	1.712	34	.048

Table 16: Paired t-test for Lexical Resource (Writing Task 2)

5.1.2.4 Improvement in Writing subscores

Further analysis was performed on Writing to drill down to the analytical subscores. IELTS Writing is scored against four key indicators across two tasks: Task Achievement/Response, Coherence and Cohesion, Lexical Resource and Grammatical Range and Accuracy. In examining changes in mean subscores across Tests 1 and 2, only the change in mean for Lexical Resource for Task 2 (a short essay task) was found to be statistically significant ($t=1.712, df=34, p=0.048$) according to a one-sided paired t-test as shown in Table 16.

5.2 The nature of change in IELTS scores across Tests 1 and 2

One-way ANOVAs found no statistically significant differences in the four macro-skills scores and Overall IELTS score according to field of study, language background or gender, consistent with Craven’s findings (2012). The sample was fairly homogeneous for age and length of time between Tests 1 and 2 so these were not tested. In this section we thus move to examine patterns of improvement in the four macro-skills relative to each other.

5.2.1 Principal Components Analysis (PCA)

We first investigated whether the improvement in the four macro-skills (Listening, Reading, Writing and Speaking) followed the same pattern using an exploratory factor analysis approach, specifically Principal Components Analysis (PCA) (O’Loughlin & Arkoudis, 2009). The unrotated factor matrix for improvement in Listening, Reading, Writing and Speaking is reported in Table 17.

	Factor 1	Factor 2
Listening improvement	.840	
Reading improvement	.818	
Writing improvement		.686
Speaking improvement		-.714

Table 17: Unrotated factor matrix

Two factors emerged, with Reading and Listening loading together on the first component (ie, strong positive correlation), and Speaking and Writing loading on the second component (ie, weak negative correlation). A rotated analysis was also conducted to spread variance more equitably across the main factors. However, the rotated analysis gave a very similar result here, most likely because there were only four variables (O’Loughlin & Arkoudis, 2009). The rotated factor matrix is reported in Table 18.

	Factor 1	Factor 2
Listening improvement	.847	
Reading improvement	.859	
Writing improvement		-.728
Speaking improvement		-.752

Table 18: Rotated factor matrix (Varimax rotation with Kaiser normalisation)

Once again it is clear that improvement in Reading and Listening are grouped together, while improvement in Writing and Speaking are grouped outside of this factor. This means, on the one hand, that there is no correlation between the average improvement in Reading and Listening and the average improvement in Speaking and Writing. On the other hand, it means that Speaking and Writing are themselves likely to be distinct factors in their own right. The percentage of variation accounted for by both unrotated and rotated two-factor loading of variance is reported in Table 19 on the following page.

The high loading of variance on these two factors (64%) is indicative of the usefulness of a two-factor analysis.

	Initial Eigen values			Extraction sums of squared loadings			Rotation sums of squared loadings		
	Total	% of Variance	Cumulative%	Total	% of Variance	Cumulative%	Total	% of Variance	Cumulative%
1	1.492	37.299	37.299	1.492	37.299	37.299	1.458	36.450	36.450
2	1.072	26.795	64.094	1.072	26.795	64.094	1.106	27.644	64.094
3	.900	22.510	86.604						
4	.536	13.396	100.000						

Table 19: Total variance explained by two-factor model

To check the level of correlation between the two skills that load on to the first factor (ie, Listening and Reading) versus the two skills that load on to the second factor (ie, Writing and Speaking), we tested correlations between all degrees of improvement. The results of this are reported in Table 20.

		Speaking improvement	Writing improvement	Reading improvement	Listening improvement
S Imp	Pearson Correlation	1	-.101	.020	.090
	Sig. (2-tailed)		.479	.891	.532
W Imp	Pearson Correlation	-.101	1	-.051	-.067
	Sig. (2-tailed)	.479		.724	.640
R Imp	Pearson Correlation	.020	-.051	1	.459**
	Sig. (2-tailed)	.891	.724		.001
L Imp	Pearson Correlation	.090	-.067	.459**	1
	Sig. (2-tailed)	.532	.640	.001	

Table 20: Correlations between improvements in Speaking, Writing, Reading and Listening

The only significant correlation observed was that between improvements in Listening and Reading ($r^2=0.459$, $n=51$, $p=0.001$). It appears then that the pattern of scores for Listening and Reading is different from the other two scores. This means that Listening and Reading constitute one dimension of 'communicative language ability', while Speaking and Writing constitute two other distinct dimensions.

5.2.2 Correlation and regression analysis

As shown in Table 21 below, the initial scores for each macro-skill correlate strongly with the improvement in their respective mean scores across Tests 1 and 2 as expected. Additionally, there was a strong correlation between the initial score for Reading and mean improvement in Listening along with the expected mean improvement for Reading.

		Speaking improvement	Reading improvement	Listening improvement	Writing improvement
S1	Pearson Correlation	.393**	.048	.073	.042
	Sig. (2-tailed)	.004	.737	.612	.772
R1	Pearson Correlation	-.105	.661**	.309*	-.011
	Sig. (2-tailed)	.463	.000	.027	.938
L1	Pearson Correlation	.062	.004	.371**	-.053
	Sig. (2-tailed)	.664	.977	.007	.710
W1	Pearson Correlation	-.257	.045	-.165	.577**
	Sig. (2-tailed)	.069	.754	.248	.000

Table 21: Correlations between Test 1 scores and improvement across Tests 1 and 2

These results do not indicate, however, whether the improvement was more likely to be by the lower or higher scoring participants. Further regression analysis, specifically Ordinary Least Squares (OLS) regression was undertaken to test this hypothesis. Improvements in the four macro-skills were treated as the dependent variable and the initial Test 1 scores for each macro-skill were treated as the explanatory variables. However, the initial test scores were not found to account for a significant degree of the variation in improvement for any of the four macro-skills. While the increase in scores between Tests 1 and 2 were found to be significant for low-scorers (ie, 5.5 and below) (see Section 5.1.2.2), we did not find support for the hypothesis that those participants with lower initial scores improved more than those with higher initial scores.

5.3 The relationship between IELTS scores and academic achievement

In this section we explore the relationship between the scores for each of the four macro-skills and Overall scores with the GPAs of the participants in the semester in which they took Tests 1 and 2, as well as in the two subsequent semesters. The correlations between the IELTS scores for the four macro-skills, along with academic achievement as measured by

GPA (Research Aim 3), are reported in the following tables. In Table 22, correlations between Test 1 scores and the GPAs of participants which include their grade in ELEC in their first semester of study are reported. In Table 23, correlations between Test 2 scores and the GPAs of participants which include their grade in ELEC in their first semester of study are reported.

The GPAs of participants that include their ELEC grades correlate strongly with their scores in Listening for both Tests 1 and 2 (Test 1: $r^2=.416$; $p=0.004$; Test 2: $r^2=.396$; $p=0.004$) and strongly for Reading in Test 2 (Test 1: $r^2=.312$; $p=0.007$; Test 2: $r^2=.401$; $p=0.004$), but not for Writing or Speaking.

As ELEC was of a qualitatively different nature to the other courses the participants were taking, in that it focused particularly on developing English language, we also explored correlations between the GPAs of participants that did not include their ELEC grade in their first semester of study. In Table 24, correlations between Test 1 scores and the GPAs of participants which do *not* include their grade in ELEC in their first semester of study are reported, while in Table 25, correlations between Test 2 scores and those GPAs are reported.

		Listening 1	Reading 1	Writing 1	Speaking 1
Cumulative GPA Semester 2 2010	Pearson Correlation	.416**	.312**	.048	.201
	Sig. (2-tailed)	.004	.007	.685	.089

Table 22: Correlations between Test 1 scores and first semester GPA including ELEC

		Listening 2	Reading 2	Writing 2	Speaking 2
Cumulative GPA Semester 2 2010	Pearson Correlation	.396**	.401**	.128	.208
	Sig. (2-tailed)	.004	.004	.371	.142

Table 23: Correlations between Test 2 scores and first semester GPA including ELEC

		Listening 1	Reading 1	Writing 1	Speaking 1
Cumulative GPA Semester 2 2010	Pearson Correlation	.385**	.221	.061	.144
	Sig. (2-tailed)	.001	.060	.609	.224

Table 24: Correlations between Test 1 scores and first semester GPA excluding ELEC

		Listening 2	Reading 2	Writing 2	Speaking 2
Cumulative GPA Semester 2 2010	Pearson Correlation	.343**	.339**	.048	.194
	Sig. (2-tailed)	.014	.015	.739	.172

Table 25: Correlations between Test 2 scores and first semester GPA excluding ELEC

The results were almost identical to those correlations with GPAs including ELEC. The GPAs of participants that exclude their ELEC grades correlate very strongly with their scores in Listening for both Tests 1 and 2 (Test 1: $r^2=.385$; $p=0.001$; Test 2: $r^2=.343$; $p=0.014$) and Reading for Test 2 ($r^2=.339$; $p=0.015$), while there was a weaker near-significant correlation Reading in Test 1 ($r^2=.221$; $p=0.06$). There was no correlation between GPAs excluding ELEC and the participants' scores in Writing or Speaking for Tests 1 or 2. It appears then that those participants who had higher scores in Listening (both Tests 1 and 2) and Reading (particularly Test 2) tended to have higher GPAs in their first semester of study. The participants' scores in Writing and Speaking, on the other hand, did not appear to be a useful predictor of academic achievement as measured by GPA in their first semester of study.

The degree of correlation between the GPAs of participants in their subsequent semesters of study and their Test 1 and Test 2 IELTS scores for the four macro-skills was also explored. In Table 26, correlations between Test 1 scores and the GPAs of participants in their second semester of study are reported, while in Table 27 correlations between Test 2 scores and their GPAs are shown.

Once again, the GPAs of participants in their subsequent semester of study correlate very strongly with their scores in Listening (Test 1: $r^2=.545$; $p=0.000$; Test 2: $r^2=.464$; $p=0.001$) and for Reading (Test 1: $r^2=.433$; $p=0.000$; Test 2: $r^2=.503$; $p=0.000$) in both Tests 1 and 2. While the Writing scores once again did not correlate with the GPAs of participants in the second semester of study, their scores of Speaking did this time correlate with their GPAs (Test 1: $r^2=.234$; $p=0.05$; Test 2: $r^2=.303$; $p=0.034$) although not as strongly for Listening as Reading. In other words, participants who had higher scores on entry in Listening, Reading and Speaking tended to have higher GPAs in their second semester of study.

Their score in Writing, on the other hand, did not appear to be a useful predictor of academic achievement in their second semester either.

In their third semester of study, however, statistically significant correlations between IELTS scores in their first semester of study were no longer found for any of the macro-skills except a weak correlation for Speaking in Test 1, which may warrant further investigation with a larger dataset. In Table 28 correlations between Test 1 scores and the GPAs of participants in their second semester of study are reported, while in Table 29 correlations between Test 2 scores and their GPAs are shown.

		Listening 1	Reading 1	Writing 1	Speaking 1
Cumulative GPA Semester 1 2011	Pearson Correlation	.545**	*.433**	.151	.234
	Sig. (2-tailed)	.000	.000	.209	.050

Table 26: Correlations between Test 1 scores and second semester GPA

		Listening 2	Reading 2	Writing 2	Speaking 2
Cumulative GPA Semester 1 2011	Pearson Correlation	.464**	*.503**	.180	.303
	Sig. (2-tailed)	.001	.000	.216	.034

Table 27: Correlations between Test 2 scores and second semester GPA

		Listening 1	Reading 1	Writing 1	Speaking 1
Cumulative GPA Semester 2 2011	Pearson Correlation	.145	.109	.151	.234*
	Sig. (2-tailed)	.231	.369	.209	.050

Table 28: Correlations between Test 1 scores and third semester GPA

		Listening 2	Reading 2	Writing 2	Speaking 2
Cumulative GPA Semester 2 2011	Pearson Correlation	.184	.224	-.112	-.094
	Sig. (2-tailed)	.205	.122	.445	.521

Table 29: Correlations between Test 2 scores and third semester GPA

Initial IELTS scores in Listening and Reading (and subsequently Speaking) were therefore, at least for this group, only a robust predictor of academic achievement for the first year of undergraduate study. It was notable that scores for Writing did not correlate with academic achievement, a point we will be revisiting in the Discussion.

5.4 Students' views on their English language learning experiences

This section presents qualitative findings from two rounds of focus groups: one held at the beginning of a semester (four focus groups, comprising 10 students) and another at the end (five focus groups, comprising 15 students). Although the students spoke about a range of topics, we focus here on students' views about their English skills, with a particular interest in whether their views changed over the course of the semester. By "English skills", we refer primarily to the four macro-skills of reading, writing, speaking and listening. However, in analysing the data we noticed that students considered grammar and vocabulary as particular areas of concern. We have therefore included a discrete section on micro-skills (grammar and vocabulary) to represent these views, as well as referring to lexico-grammar within macro-skills as needed.

5.4.1 Macro-skills

5.4.1.1 Listening

Beginning of semester

Students conceptualised two types of listening as presenting challenges in the upcoming semester: (i) listening to academic English in structured learning environments; and (ii) listening to Australian English in everyday contexts and in conversations with fellow students.

Regarding the former, the issue concerned the difficulty of comprehending dense, discipline-specific content within the constraints and affordances of lectures and tutorials. One student put it this way:

[1] From my first studying in Griffith, I found it's like to understand what the teacher said in tutorial, is quite difficult, than to listen to in the lecture. Because in the lecture we have the text book and we can prepare for it before the lecture. But for the tutorial, all the questions that we can't imagine before the tutorial. So, when I has a tutorial, I can't understand what the tutor said, and it including many professional vocabulary, and so it's quite difficult to understand.

This student makes the point that dynamic learning environments such as tutorials put a greater stress on listening abilities, since the back-and-forth of interaction is inherently unpredictable ("all the questions that we can't imagine"). This is compounded by the difficulty of discipline-specific content ("many professional vocabulary").

Lectures, on the other hand, do not require an immediate response, and comprehension can be enhanced by targeted preparation ("we have the text book and we can prepare for it").

Another student mentioned that variable teaching quality can also affect students' comprehension of classes:

[2] ... some of our teachers are really good teachers [...], I mean like, some of the students are able to understand very well. But some not. And they are good, because they are academy [...] so I don't think the school will [...] fire them or will [...] ask them to change 'cause they have been, so many years like doing the same thing, and they may be big deal or some agent of that, and you can't expect them to change...

In contrast to the first comment, in which the student believed there were actions that could be taken to improve comprehension of lectures, implicit in this comment is a sense that the student lacks agency. The speaker suggests that some lecturers are of such high standing ("they are academy" ... "they are big deal") that they will never change the way they teach, irrespective of the needs of the audience, and nor will the university take steps to force that staff member to change.

Away from the academic domain, many students mentioned that listening to everyday Australian English has its own challenges. For example, one student emphasised accent and speaking speed as inhibiting comprehension:

[3] Because they have a strong accent. So it's kind of difficult for me to understand why they are talking about and also they speak really quickly. And even though I can tell them can you repeat or something like that, but they normally do not do that.

This student observes that locals are reticent to play the role of language instructor. Another participant described at length the frustrating experience of trying to get Australian interlocutors (including well-meaning friends) to clarify meanings:

[4] ... we may have more questions of what he or she just say. Like what's the meaning of that word? Just say again, and then more words and more words, like that. So sometimes it can take a while, and some people get frustrated, or they just forget about it. You know. Like that doesn't help us to learn. Quite often, especially some, just Australian friends. They, or when they're talking and you heard something sounds interesting, and you ask them "What was that, what did you just say?" "Oh, don't worry about it." You know.

It is noteworthy that this student feels that the local students' interactional style has inhibited his learning ("that doesn't help us to learn").

Gaps in cultural knowledge were also seen as a factor inhibiting listening ability. The following student refers to a number of non-linguistic factors that make

listening and comprehending her fellow Australian students a difficult task:

[5] I take the bus. And every time take a bus a lot of our age students they talking and I just listen, and sometimes I can't understand. Because we have a different culture, we have a different education. So we have a different thinking. So, if I want to join their topic, it a little bit, quite difficult.

However, the group did not see themselves as lacking the capacity to overcome the constraints mentioned above. Several students were of the belief that simply by staying and studying in Australia, they would gradually understand more. For example:

[6] I'm in Australia so I had to get used to speaking English everywhere I go. My job to go so supermarket and everything, so listening, speaking. With living in Australia, it gets easier, because it's something that you need to do every day.

[7] ... listening is most easiest for me. Because you have to use it every single day in Brisbane in Australia.

Some students referred to concrete strategies that could improve their general listening proficiency. For example, this student discussed how the local media could be used to improve listening comprehension and cultural understanding in general:

[8] It's cool because at the beginning when I arrive here, you watch TV, you listen to the radio, you don't understand anything. But it's cool. That once you listen or watch TV for a while you begin to understand. And even get the jokes. Jokes are different from country to country. So you can begin to get what they're saying. Because Australians speak fast. So I think listening to radio and watching TV is the best thing that you can do to learn. I have a friend [...] he came here with poor English. And he learned a lot watching Hi-Five, because [it's] for children and they speak slower, easy words, so he learned a lot watching children's TV.

End of semester

There was a perception across the group that one semester of studying and living in Australia had been beneficial for their listening ability. Most students felt that they comprehended more in both academic and everyday settings as a result of continual exposure to authentic input and by growing used to the interactional demands of commonly occurring situations.

Academic English was still viewed as a very challenging form of discourse, but most students felt it was more accessible after one semester's engagement. One student went so far as to claim that in "my first lecture, I didn't understand teacher, for me they speak very fast, but now I understand everything". More typically, students were cautiously optimistic that progress was being made, coming off a low base, due to continued exposure to academic input. For example:

[9] And my listening is awful, was awful in that time, [...] because in Griffiths University we have a lot of like lecturers and I just normally listen [...] So it's, I think probably I improve my listening skill, yeah.

[10] Yeah, I think it has improved, especially for the reading and listening because every day I need to read a lot of book to prepare for the lecture and on the lecture and tutorial, if you want to study well you need to listen to the teachers carefully which improve you listening.

Some students made the interesting observation that in lectures their identity as international students dissolved and that this had a positive impact on their listening proficiency. In lectures, they simply became audience members like everybody else, no longer spoken to in the simplified English of TESOL instructors, but rather in the authentic discourse of real university courses:

[11] ...with the lecture because it's improving my knowledge as well as my listening. Because I told you I want to be familiar with like the original pronunciation but I can hear that in my lecture because they don't care about I'm like international or local student.

[12] I don't know why but normally the language lecturer, [...] he pronounces really clear so I can understand almost everything. But normally and the contents is also obvious [...] but actually in other lecture, they like talking about other content and sometimes like really hard to understand the vocabulary, the pronunciation but they keep going because it's not language school, like language lesson. So they keep going and I have to know like keep up with their like talking and you have to keep listening.

Another positive discovery over the semester was lecture capture technology: "many course have a Lectopia and so I always stay at home and listen to the Lectopia". It was mentioned by several members of the focus group cohort, one of whom said it was utilised by "my friends ... also a lot of people". Another participant explained its attractiveness for international students:

[13] Yeah, I have one friend, her friend just stay at home and listen to the record because they think if they cannot understand they can just repeat again and again. It's very convenient for them to understand it. [...] I think it's very embarrass to raise the hand to ask the question so they prefer to stay at home and listen it. And write an email to the lecturers.

Lecture capture offers international students the opportunity to listen repeatedly to lectures and replay difficult segments. The above student acknowledges the interactional norms of Australian university lectures by referring to hand-raising and questioning, but states that such interaction can be embarrassing for her friend(s), and that there are avenues for listening and interacting which are not face-to-face and which may be preferable for some students.

On the boundary of academic and everyday English are conversations with local students in tutorials. One student mentioned that she could understand the conversational English of local students in tutorials better than before:

[14] Some tutorials, I, we have one local student, like first time even I didn't understand. He asked me like what is your name but I didn't understand. [...] And then now I understand normally, the first time I was really embarrassed I don't understand what you're saying, really embarrassed.

Overall, the cohort believed that their ability to understand everyday Australian English had improved. One strategy for improvement mentioned by several participants was having a job that entailed interaction with the local community. One student described the benefit of his job in a local restaurant: "the customer is speaking English is so fast so you have to listen very carefully". A few students also mentioned having made some local friends and that this had benefited their listening comprehension.

5.4.1.2 Reading

Beginning of semester

When asked in the initial round of focus groups which of the four macro-skills of speaking, listening, reading or writing posed the biggest challenge for them as international university students, a number of participants identified reading. Although they had no problems reading newspapers, novels or other standard texts, they often struggled to comprehend academic textbooks, reference books and journal articles: "If I read a book, like a history book or like a story book or fiction or something like this, it's easy. But academic reading is quite difficult for me".

The cohort's responses suggest that there was an initial deficit in their discipline-specific lexical knowledge, in that they were suddenly exposed to a great deal of unfamiliar vocabulary when they began to read academic texts. One participant said that:

[15] It's very hard to read or to understand some of the words, the terminology they're using [in academic words]. 'Cause we are not used to those words, we never, maybe we never heard of them.

Another student commented that "I face [...] a lot of words that I don't know in the academic journal article because it's too hard." (Vocabulary is discussed further in Section 5.4.2.2.)

In addition, participants found that the syntactic complexity and the abstract, theoretical content of many academic texts made the main thrust of a text difficult to discern. This caused consternation early on:

[16] [I] always panic when I get the paper because I don't even know what's that about.

[17] A lot of information if you don't find out the key words or what's the key points, you might just read and memory will be blank. I don't know what that's talking about, what's the key point?

End of semester

The participants in the final round of focus groups reported that their overall academic reading proficiency had increased over the course of the semester. Several participants even listed reading as the macro-skill they now had the least difficulty with: "I would say my reading skill is improving well, the best actually". This contrasts with the initial round of focus groups, in which many participants regarded reading as their most problematic skill.

Some participants reported an increase in reading speed over the semester: one recalled that during her first week at university it had taken her 30 minutes to read one page of text, whereas she could now read a page in three minutes.

The perceived increase in reading proficiency and speed was partly attributed to the large volume of reading which the participants did for their degree courses: "I think it has improved, especially for the reading [...] because every day I need to read a lot of book to prepare for the lecture".

The amount of academic reading was perceived to produce a concomitant increase in their store of discipline-specific vocabulary, which once processed could be employed for future academic reading: "You have to read some text books, right, and like when you're reading them you have to learn a lot of new words and I think the things can only get better".

5.4.1.3 Writing

Beginning of semester

Overall, the students' views on their academic writing ability at the commencement of university study could be described as cautious and realistic. That is, they generally felt that they had a significant challenge ahead in this area. Only one student expressed outright confidence in his/her writing ability, proclaiming, "For me writing is easy because I like writing. Become writer, I can't stop". More typical was the view that they entered the English-medium university environment with basic writing proficiency but would henceforth grapple with higher expectations. As one student put it, "I'm good enough to write an academic paper, not at a high level of course, but on a beginner level". Many students had already formulated perceptions about specific areas of concern in their writing. For example, some participants mentioned vocabulary:

[18] ...we have a lot of academic writing and I don't think my vocabulary is enough to write a real academic writing.

Others referred to the fact that academic writing in English is a fundamentally different mode of expression to other types of discourse, eg: "You are learning how to write English academically. It's something that is different from personal writing".

Academic writing was also problematised within the frame of cultural differences. One student noted that "for me, the big challenge is the writing. Because I cannot write an academic writing or something like a letter because they have a different style".

The student's use of the pronoun "they" is indicative of the view that written output can mark one as being a member of the "non-English speaking background" class: ie, "they" write differently to "us". Another student focused on how one's writing reflects differences in ways of thinking:

[19] *Because for second language for me, we always write in Chinese which means thinking in Chinese, but here we should write in English, think in English. Different logic.*

Similarly, differences in educational culture were pinpointed by one student as affecting writing:

[20] *I don't know about Griffith, but I think most Asia countries we are not focused on academic writing. It's gonna be more focused on the content itself. Like right answers like that kind of thing. So we need to learn how to write, like essay properly, and like how to use English properly.*

Another concern was the time, and therefore workload, associated with producing good written output. Even at the beginning of the semester some students were feeling overloaded and worried about having sufficient time (and support) to complete their written assignments satisfactorily:

[21] *Yeah, university is quite busy. I'm actually quite busy with university work. I'm writing essay, this and that study.*

[22] *I have two assignment and I have to write, so that is a big challenge for me.*

[23] *I think it's the limited time. We don't have time to do and also to provide, for example the English HELP is only one hour. But sometimes we have three or four essay to do...*

End of semester

At the end of the semester, several participants felt that they were better writers because they had received some in-course, explicit instruction on how to write academically for their degree subjects. As one put it, "it's quite good because they tell me how to write an assignment and tell you how about the structure". Some students elaborated on specific skills that they had found useful for completing assignments:

[24] *... essay is really integral part of the courses, writing syntactical, knowing how to structure your argument, how to find what you need to write, stuff like that. So yeah, [...] I didn't have much ideas on how to go about writing. It was good that some kind of courses that actually gives you those...*

[25] *I think most part they teach you how to analyse the topic and structure your essays, so I think it's help me to increase a lot...*

However, some students were critical of the explicit writing instruction because they felt it repeated the kinds of instruction they had received in academic English courses prior to commencing their degree:

[26] *I've already done like a foundation study before, so I found the essay structure thing is sort of redundant for me and I think it would be great if like the lecturer would combine like the structure and reference stuff into one lecture, and I think that would be enough.*

[27] *I think some writing skills, because I take that already, yeah, before, so [...] it's like all the stuff, some of them things repeat and repeat and repeat again, so not for me really should fresh.*

5.4.1.4 Speaking

Beginning of semester

In the initial round of focus groups, speaking was perceived as less challenging than the other macro-skills, partly because comprehension of information presented at a university (through lectures, tutorials and texts) emphasises listening and reading skills rather than speaking skills, while assessment items most often focus on writing skills. The following comment reflects this perception:

[28] *For me I think [...] speaking is easier than reading and writing because during the course I study academically than, writing and reading tend more academic, so lots of vocabulary to [memorise].*

The implication is that the participants initially perceived reading, writing and listening to be more demanding than speaking in an academic context.

From the outset, participants were well aware of the need to increase their oral proficiency by communicating in English whenever possible. Some achieved this by living, working or socialising with other users of English as a second language whose first language was different from their own. They were thus compelled to communicate with one another using English as a lingua franca, which they found valuable for increasing their oral proficiency:

[29] *Have friends of different nationalities, even English not their first language. You gonna be forced to speak English. [...] I don't think there is a secret formulae for that, just speak English.*

Participants also realised that maintaining L1-only social networks would hinder their linguistic development:

[30] *My girlfriend, she's Korean, and her Korean friends come to Australia without any English. And they just hang around like that, and they don't speak English at all. So they're not going to learn, never.*

As well as communicating with other ESL users in English, the cohort also increased their oral proficiency through interaction with Australian speakers of English. They often found this agreeably easy:

[31] *It's really easy to engage in a conversation with someone. Like you are at the bus stop. Australians don't mind talking to you.*

[32] *I have a lot of opportunity to have a conversation with Aussie guys.*

[33] *If you are trying to find something in a shop or something, and you just ask people or you know, just talk to them and or quite often they will talk to you as well.*

The only reported disadvantage was that Australian speakers of English tended not to correct the participants' linguistic or phonological irregularities, despite their requests to have these pointed out.

[34] *I have Australian friends, and [I] ask "can you correct my grammar mistakes when I speak" and everybody tell "oh your English very good", and I tell you "very polite". Not good for my English. But [...] friends don't correct mistake, they want just speaking.*

Some participants circumvented this issue through being accommodated in homestays with a family of Australian speakers of English. Homestay families were normally aware of international students' desire to increase their English proficiency and often pointed out perceived errors:

[35] *I have a Australian home-stay family, so [...] when I say something wrong, they always correct me. And they always correct my, how to say, pronunciation as well. So I learn every day, I learn something.*

However, although they generally had little difficulty engaging with Australian speakers of English in everyday social contexts, a small number of participants encountered difficulties interacting with domestic Australian students at their university:

[36] *I mean like you been asking questions or when you're asking for help or anything like that and you have been rejected or you have been ignored. And you feel upset and dumb.*

The participant who made the above comment was unsettled by the perceived reluctance of domestic Australian students to interact with her: "You just feel that people don't want to talk to you, don't want to answer your questions, they feel you're annoying, you know. [...] You might kind of try to keep yourself away from them".

End of semester

Several participants had commented initially that they found speaking less problematic than the other macro-skills. In the final round of focus groups, however, only one participant listed speaking as their

least problematic skill, while two participants said that speaking was the skill they had most difficulty with, though it must be borne in mind that the composition of focus groups altered somewhat between the beginning and end of semester. This may be linked to a perception that lectures and even tutorials afforded the participants few opportunities to speak and that development of their speaking skills was impeded as a result. One participant said that "the environment of the university is more focused on the lecture and that even what I see in the tutorial is not people actually discuss or express their opinions or anything. [...] There is not much speaking involved [so] it's really difficult to improve your speaking ability I think". Another participant believed that the emphasis at university on writing and reading had a detrimental effect on their speaking ability:

[37] *Because we just attend the lecture I think for me my level is just writing and reading is increased but my speaking is getting worse because I don't have a lot of time with friends to speak English, so every time I just attend the lecture and go home. I think I don't have time to practice my English.*

One common form of assessment which employs spoken discourse is oral presentation. Participants were required to give an oral presentation as part of the assessment for their ELEC, as well as for some other courses in their degree programs. A number of participants found giving oral presentations beneficial for their speaking proficiency, partly due to the format of the oral presentation exercise which demanded structured, informed, persuasive and fluent delivery from the presenters:

[38] *Because you have to look at so many people and then you have to explain your opinion and then the context for that, so if you do more and more time practice you will be more confident with what you are said, explain about.*

Another contributing factor was the feedback they received from tutors and students after giving their presentation: "[I get] massive feedback from my tutor, from other students. I see what's wrong in my speaking. I can improve in future."

A common theme in the final round of focus groups, repeated from the beginning of the semester, was that L2 proficiency increased by speaking English with other ESL users:

[39] *The workers who work with me they speak [English] quite good because they also work like in a multicultural environment.*

Allied to this was the issue (also mentioned in the initial round of focus groups) that some respondents tended to interact with people from their own country in their L1, despite the potential detriment to their L2 development. This issue sometimes manifested in tutorial activities such as group discussions: students with a shared linguistic background would jointly

carry out the group activity in their L1. "My speaking [hasn't improved] because most of [my classmates are] from the same nationality [...] so we don't always speak English in the class". The issue was also apparent when participants were accommodated with students from the same country:

[40] I'm afraid I live with Chinese, so every time we speak Chinese because they always told me, say because in the at school they already speak English, they won't relax because they speak English, a kind of stressful for them, so they go back home they just speak Chinese to me. Yeah, even though I want to speak English they always say shut up.

So although some members of the cohort pursued opportunities to speak in the L2 to other ESL speakers, others took the path of least resistance and communicated in the L1.

As with the initial round of focus groups, the participants reported having numerous opportunities to interact with Australian speakers of English in a variety of social contexts such as organised activities (eg church groups or study groups) as well as one-time encounters (eg in a supermarket or waiting for a bus). Some participants interacted in English when they went to pubs or clubs because "Australians really want to talk when they're drinking alcohol". Several participants had part-time work, often at busy restaurants, where the pressure to perform in an L2 environment pushed them to increase their communicative competence:

[41] I believe my speaking is really improved when I work because I'm working in a café and then I'm taking orders so then I have to you know speak English every time in my workplace. So this is really helpful for me.

However, as some respondents had stated at the beginning of the semester, it was often more difficult to interact with Australian speakers of English in the university context than in other purely social contexts. The issue was particularly apparent in lectures or tutorials, where domestic Australian students appeared to become tired of being asked for clarification by international students:

[42] I always say I'm sorry, can you repeat all and she say it's quite annoyed, so sometimes it's difficult for the international student want to make friend with them. Because they say you always say pardon and it's yeah annoying.

Even in non-academic milieus such as in on-campus halls of residence some participants reported having difficulty initiating conversations with domestic Australian students:

[43] Makes it kind of hard to get to the native students because they also stick to themselves quite a lot so [...] kind of hard to [...] learn how to talk to them.

For their part, some of the participants did not initiate interaction with domestic Australian students either, as this excerpt reveals:

[44] Facilitator: I mean do you speak to domestic students, do you have conversations with English speaking students?

Student 1: Not really much.

Student 2: Not much.

The potential face-loss of a failed communicative encounter may have been a causative factor in this reluctance.

5.4.2 Micro-skills

5.4.2.1 Grammar

Beginning of semester

Where grammar was mentioned, it was usually associated with the notion of mistakes and errors, and the need for these to be weeded out: ie, the belief that good language learning involves direct feedback on grammatical deficiencies. Students expressed dissatisfaction that such instruction was rarely given in university courses. For example, the following student objects to receiving correction on specific grammar mistakes when there is no generalised attention to grammar within the curriculum:

[45] But, I think, like, with the academic skills, I think it's also really important to improve students' vocabulary and that's [...] grammar as well and in most of the course there are not any grammar information. Just teachers say you know, when we do something wrong they correct, but I think the course also needs to include that kind of things.

Another student feels that grammar should be a central concern of instructors, and if consistently left unchecked, will impede her progress as a competent English user:

[46] Yeah, because these are the main things we should be perfect. After that we can improve our English correctly. You know, like much better, because if I continue to do this same mistakes and nobody corrects them and nobody gives me the information that I need to know about it, so I will continue like this. And I will do the same mistakes again and again.

End of semester

Backing up the sentiments above, there were some comments expressing disappointment that more emphasis was not placed on grammar in their ELEC in the past semester, as evidenced in the following exchange:

[47] *Facilitator: So the question is: Can you think of other language skills or academic skills that you wish that you could have learned in this course by doing this course?*

Student: Ah, grammar.

F: Ah-ha, grammar.

S: We had, I think we had one tutorial, and it was like two weeks ago, that we really touched grammar: like we wrote and then we put it on the board and checked the grammar. And it was only one, and the changes that, like things that we won't notice, like two past tense in one, like stuff like that and one tutorial is just not enough.

F: Okay.

S: And just do, okay, this is wrong and bye-bye.

F: Okay, so more grammar.

S: Yeah.

It is difficult to know whether students felt the same way about their regular discipline-specific courses, as they did not mention grammar in relation to them. It may be that students only expect attention to grammar in ESL/EAP-related courses. Nevertheless, several students did mention that they had turned to learning support services as a way of having grammar issues attended to in various assignments. In cases where grammar was addressed directly by those services, they expressed satisfaction. For example:

[48] *So the tutoring in English HELP, it helps me, they cracked my mistake of the grammar and [...] gave me some ideas of how to write the assignments.*

[49] *I did English HELP and [...] it's really helped me to improve my English, especially the grammar.*

In cases where grammar was not attended to, they expressed displeasure:

[50] *So for me, it was really good but I take the one times of English HELP but I wasn't really satisfied with it because firstly it was grammar checking but they don't, they didn't see my grammar. They just trying to you know, change it, all the essays so they kind of ignore my essay, I wasn't feeling like good and also he or she, I mean like he was like trying to [...] you know kind of restructure my essay. So and then he didn't finish like checking all [...], because she only focus on like one paragraph. So it [...] wasn't a really good opportunity for me, so I'm not booking anymore.*

[51] *Yeah, I also wanted to check my grammar not the content but she tried to change my content and she tried to change my opinion and even, I tried to write my essay but sometimes she was angry because it's not in the point here and this I shocked. Because I was trying to write a good essay but yeah.*

5.4.2.2 Vocabulary

Beginning of semester

As mentioned in Section 5.4.1.2, participants at the beginning of the semester struggled with the high volume of new technical or discipline-specific vocabulary with which they had to become familiar. Their comments indicate that this initial lexical deficiency impacted on their ability to produce and comprehend academic discourse.

On the productive side, the participants needed to quickly expand their discipline-specific lexicon in order to write or speak about technical subjects as part of their assessment. As we saw in Section 5.4.1.3, some members of the sample were concerned about their ability to manage this, with one participant commenting that “We have a lot of academic writing and I don't think my vocabulary is enough to write a real academic writing”.

In terms of receptive skills, reading academic texts was perceived as relatively unproblematic because participants encountering unfamiliar vocabulary often had time to refer to a dictionary. Listening was viewed as more challenging, since students rarely had time to look up unfamiliar terms they heard in lectures or tutorials. This meant that they faced difficulties comprehending the content of lectures:

[52] *[Lectures are] hard for me to understand [because of] you know, vocabulary. And [...] when the lecture says something I don't know, I am not able to check it and understanding quickly.*

Aware of the urgent need to increase their discipline-specific lexical knowledge, several participants did a great deal of course-related reading, which incrementally increased their store of lexical knowledge.

[53] *'Cause when I read book sometimes, I saw new words, I should look dictionary to understand. Like if I will study these words, I don't need to look dictionary up.*

Other participants increased their vocabulary through reading novels, internet websites or “anything you find interesting”. This may have been effective for increasing general vocabulary, but its value for developing academic lexical knowledge is less clear.

There were a variety of opinions about whether a bilingual (eg Chinese–English) or monolingual (eg English–English) dictionary was more appropriate for increasing L2 vocabulary. A student favouring the use of monolingual dictionaries stated that:

[54] *English–English dictionary it's what helps me because I'm not just translating the words, I'm seeing the meaning of that word in English. And if the explanation of that word, there is another word I don't know, I'm going to be forced to go to that other word. And then I learn more.*

However, another student said that she was unlikely to use a monolingual dictionary, arguing that when she looked up an unfamiliar word she was often confronted with an explanation containing even more unknown words:

[55] If you look up and it comes up with heaps and heaps of other words you don't understand, you can spend like ten minutes or half an hour reading one paragraph.

End of semester

There was relatively little comment about vocabulary in the final round of focus groups. Some members of the sample perceived that their store of general vocabulary had increased over the semester due to reading books and interacting with people in English: "I know if I read a book I know I will learn new words, how the words are combined together, and then it will improve [...] definitely". This participant displays awareness of the need to understand meta-lexical aspects of vocabulary, ie, how to decode words, identify their components (eg prefixes) and apply this knowledge to inferring the meaning of other unfamiliar words.

Some respondents also perceived that in general their ability to produce and comprehend spoken and written academic discourse had increased over the semester. Nevertheless, comprehension and production of such a large volume of new technical and discipline-specific vocabulary continued to be an issue:

[56] Technical vocabulary you need study, for example I [have to] study academic words, business, more economical vocabulary, and management like marketing, special focus.

None of the participants believed that their discipline-specific lexical knowledge had reached a stage where further study was no longer warranted.

6 DISCUSSION

This section interprets the key results from this study by first responding to the five research aims and then outlining some of its limitations.

6.1 Proficiency change over initial semester of study

This section focuses on the first research aim, namely to measure change in English language proficiency over one semester of international students at Griffith University using the IELTS test. This study found that, on average, the mean scores were higher in Test 2 than Test 1 in all four macro-skills, though these were mostly marginal increases. There was therefore little measurable improvement in proficiency on average, as measured by the IELTS Academic test, during the initial semester of undergraduate study except in Speaking. This outcome was not surprising given the short timeframe in which acquisition could occur. In previous studies (Arkoudis & O'Loughlin, 2009; Craven, 2012), proficiency was tracked over the entire university degree. Despite one to three years between Test 1 and Test 2 in these studies, they found Overall mean band score increases of just 0.413 and 0.3 respectively during the degree.

In our study, it is likely that acquisition *did* occur but that in some cases, the gain was not measurable on the IELTS scale. IELTS reports scores in terms that are meaningful to stakeholders, but underlying this seemingly simple reporting mechanism is a complex system of analytical scoring which is weighted and averaged, based on extensive research and trialling, so as to report one numerical score per macro-skill. This belies the difficulty of moving from one band to another. In Listening and Reading, for example, IELTS score processing and reporting indicates that it is possible for a candidate to score at the bottom of the range of one band score in Test 1 and at the top of the range of the same band score in Test 2, which would indicate proficiency gain, but without translating to improvement in IELTS terms (http://www.ielts.org/researchers/score_processing_and_reporting.aspx). As band scores, rather than raw scores, were entered into the database for Listening and Reading at the time of writing, it was not possible to investigate if this was in fact the case. This is not a criticism of the test but a reality of the necessity of reporting in meaningful terms, which necessitates threshold cut-offs.

In light of the above, to find a statistically significant improvement in Speaking is an interesting finding. Having investigated what contributed most to improvement in Speaking, we found that, while all four subscores showed statistically significant gains, Fluency and Coherence and Pronunciation showed gains of almost half a band score and these were found to be highly statistically significant. Fluency and Coherence mean scores increased by 0.49, Pronunciation by 0.46, Grammatical Range and Accuracy by 0.31 and Lexical Resource by 0.29 of a band score. The only other published study to date

which explored subscore increases was Craven (2012), who found Grammatical Range and Accuracy to have the greatest mean improvement by subscore at 0.35, followed by Pronunciation (0.23), Lexical Resource (0.1) and Fluency and Coherence (0.05), though these were not found to be statistically significant. The increase in Grammatical Range and Accuracy was therefore similar between the two studies despite considerable differences in the time period over which the two studies took place, while our study showed greater gains in the other three subscores.

In terms of Writing gains, we found little change on average between Test 1 and Test 2 due in part to within-subject variability. Previous studies found that Writing saw the least improvement between Test 1 and Test 2. Arkoudis and O'Loughlin (2009), for example, found Writing only increased by 0.2 of a band score, though at that time Writing and Speaking were still reported in whole bands only. Craven (2012) also found minimal increases in Writing with a mean increase of just 0.11 at the end of degree. At subscore level, we found only Lexical Resource in Writing Task 2 showed a statistically significant improvement though it was small at 0.14. Craven found a slightly greater increase in Lexical Resource (0.2), though it was not reported as being statistically significant. Similar to Craven, we found isolated improvement and small gains for some candidates in Writing though an absence of score gain was not unexpected for the reasons cited earlier.

Focus group data shows that students who felt that it was important to engage with external activities expected an improvement in their speaking skills. They also appeared to understand that speaking and interacting with people predominantly in their L1 could be detrimental to their English language development. Spending four months in an English language environment where English is required in the university setting does seem to provide an opportunity for an increase in speaking proficiency to occur. However, we cannot confirm what was specifically driving the increase in this group and, as previously noted, the research literature consistently shows that many variables impact proficiency gain.

The above commentary raises questions about what we are really observing in terms of proficiency change. Our study used IELTS to begin to explore what occurs in the initial semester of study where the closest relationship between IELTS score and academic outcomes was observed (see Section 6.3). The use of a standardised test such as IELTS provides comparability across degrees and institutions, and IELTS is currently the most common yardstick for measuring English language proficiency by employers and professional bodies at and beyond graduation in Australia. However, the IELTS Academic test measures general academic proficiency and may not reflect what students have actually been exposed to or learned in their first semester of university study. For example, discipline-specific vocabulary and genre-specific writing required within the discipline are not tested in IELTS as that is not the purpose of the test.

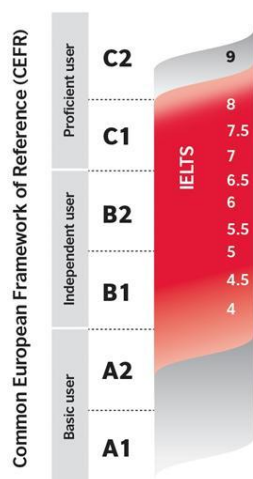
Clearly, there is a complex relationship between general academic proficiency and the discipline-specific demands of university degrees, but this matter is beyond the scope of this report.

6.2 Variation in English language proficiency

In this section we discuss our findings in relation to our second research aim, namely *to explore variation in language proficiency of initial semester students at Griffith University using the IELTS test*. As expected, we found variation in IELTS scores, ranging from what we term “low-scorers” (IELTS band 5.5 and below), “mid-scorers” (6.0 or 6.5) through to “high-scorers” (7.0 and above). The greatest concern in the higher education sector has been in relation to low-scorers and mid-scorers, and whether they are adequately prepared for tertiary study.

We found that the low-scorers had significantly higher scores after one semester of study, but that the mean improvement amongst mid-scorers and high-scorers was not statistically significant. However, unlike O'Loughlin and Arkoudis (2009), we did not find evidence through regression tests that low-scorers were more likely to improve than those with higher scores (cf Craven, 2012). In other words, we found an absolute difference but not a relative one. This can be attributed to the significant amount of variation in scores across Tests 1 and 2 amongst the mid- and high-scorers. That is, mid- and high-scorers were just as likely to obtain the same score or drop a band or two as improve in Test 2. Changes in mean score across Test 1 and 2 amongst the mid- and high-scorers thus arguably reflect regression to the mean that is attributable to measurement error (Green, 2005; O'Loughlin & Arkoudis, 2009). We would argue, however, that the statistically significant mean improvement amongst the low-scorers is not as well explained with reference to regression to the mean. Instead, it is more likely to be a reflection of the more rapid progress expected at lower levels of language proficiency (cf Green, 2005).

Recent work in the English Profile project offers empirical validation of this “intermediate plateau” that ELT experts have long acknowledged. In the project, the Common European Framework of Reference (CEFR) is used for referring to proficiency levels, where users are divided into “basic user” (A1/A2), “independent user” (B1/B2) and “proficient user” (C1/C2). Figure 5 on the following page shows the official CEFR to IELTS comparison.



Retrieved from www.ielts.org/researchers/common_european_framework.aspx

Figure 5: CEFR and IELTS comparison

The English Profile project is investigating the levels at which grammatical and lexical features of language have a tendency to be under control (ie, become “critical”) using the 40 million-word Cambridge Learner Corpus (Hawkins, 2010; McCarthy, 2011). A key finding is that there is a steeper trajectory at the CEFR B2 to C1 levels (ie, IELTS 6.5/7.0), exacerbating the difficulty for users to move to the level of proficient user (McCarthy, 2011). This has also been previously noted from the research literature (Arkoudis & O’Loughlin, 2009; Elder & O’Loughlin, 2003; Green, 2005). Briguglio (2011) goes further and argues that progressing to an IELTS band 7 does not happen naturally but requires “extra and sustained measures” (p321). On the other hand, the trajectory is less steep at the CEFR B1 to low B2 levels (cf IELTS 5.0/5.5). In other words, there is solid empirical evidence that users find it easier to move from CEFR B1 to B2 than from CEFR B2 to C1, a finding that is reflected in our study.

We noted earlier that very low scorers (ie, less than 5.0) could arguably have been treated as outliers in that these scores were most likely due to a lack of motivation to complete the IELTS test. However, we were reluctant to remove these from the statistical analysis as we found on closer examination that individual participants were not necessarily consistent low-scorers across the four macro-skills of Listening, Reading, Speaking and Writing. Instead, we found that there was evidence of lack of engagement to either Test 1 or Test 2 in a number of cases for particular sections of the IELTS test. The most striking case of this was where one participant scored IELTS band 1.0 in Test 1 for Reading, but band 6.0 in Test 2. Another case was where one participant dropped from IELTS band 6.0 in Test 1 for Writing to 4.5 in Test 2. A few other participants dropped or increased by a band or more across Tests 1 and 2 in particular macro-skills, which is also evidence of lack of buy-in at either Test 1 or 2,

particularly given that they were mid-scorers or even high-scorers for other macro-skills. Of the 11 participants who were very low-scorers for one or more of the macro-skills, we found that seven had a GPA well above the pass of 4.0 in their first and second semesters of study. Indeed, the participant who scored IELTS 4.5 in Writing in Test 2 (after scoring 6.0 in Test 1) had a GPA of 6.0 (ie, distinction level) in her first semester of study. This suggests that either there was a lack of engagement with the test, or possibly that in some cases students are able to “compensate” for a weak macro-skill through higher proficiency in the other macro-skills. The remaining four participants who were very low scorers in one or more of the macro-skills were failing in their first two semesters of study (ie, they had GPAs less than 4.0), and so were the only true very low scorers in this sample.

Our final finding in relation to variation in IELTS scores across and amongst those scores and prior to, and at the end of, their first semester was the pattern of scores across the macro-skills. Our analysis indicates that Listening and Reading formed one coherent factor that explained this variance, while Speaking and Writing formed two other weakly related factors. In other words, not only does Speaking contrast with the other three macro-skills as found by O’Loughlin and Arkoudis (2009), but Writing should also be treated as distinct from Listening and Reading, in contrast to what O’Loughlin and Arkoudis (2009) found. This is more consistent with the received view in second language acquisition that receptive skills should be treated as distinct from productive skills. It is also echoed in our finding in regards to the relationship between English language proficiency and academic achievement, a point to which we now turn.

6.3 English language proficiency and academic achievement

In this section we discuss our findings in relation to our third research aim, namely to investigate the correlation between language proficiency as shown through IELTS test scores and overall academic outcomes as measured by GPA.

Our key finding was that while Listening and Reading were strongly correlated with the GPAs of students in their first semester of study, Speaking and Writing were not. In other words, we found evidence of a relationship between English language proficiency in the receptive macro-skills and academic achievement for students in their first semester of study. The emphasis on the importance of Listening and Reading amongst participants in the focus groups was thus vindicated by this strong correlation between those macro-skills and GPAs. Our findings thus echo those of Kerstjen and Nery (2000) and Cotton and Conrow (1998) who found weak to medium positive correlations between scores in Reading and academic performance and, to some degree, with Ushioda and Harsch (2011) who found a highly significant correlation between coursework grades and IELTS Reading as well as Writing.

This finding contrasts markedly, however, with Craven's (2012) analysis where she found no clear relationship between the IELTS scores of participants and their GPA. Nonetheless, it is important to note that the GPAs of the participants in our sample were only from their first year of study, and indeed we found that the strong correlation between IELTS scores and GPA evident in their first two semesters of study broke down in their third semester of study. In other words, as students enter their second year of study, other factors appear to be more influential on their GPAs than their initial English language proficiency in Listening and Reading. This means that there is likely to be a tighter relationship between IELTS scores and academic achievement in the initial semesters of study, which is consistent with the use of IELTS to gate-keep entry into tertiary institutions.

It was perhaps not surprising that the participants' scores in Writing did not correlate strongly with their GPAs in their first year of study. This is partly because the type of writing tasks assessed at university are likely to be discipline-specific, in contrast to those more generic academic tasks in the Writing components of the IELTS test as found by Moore (2004). Since the relationship between general academic proficiency and discipline-specific proficiency is a complex one, as we have already noted, a strong correlation is not necessarily expected between IELTS scores in Writing and GPA. It is also perhaps partly due to the fact that in large first-year classes, assessment is very likely to include more tasks that require relatively less extended writing, thereby naturally placing greater weight on the students' ability to comprehend assessment tasks than produce extended discourse. The relationship between proficiency in Writing and academic achievement clearly requires more research which draws on other kinds of data, including for instance a detailed breakdown of the actual requirements on language proficiency of assessment tasks that make up those GPAs, as well as other more discipline-oriented measures of language proficiency.

6.4 Students' views of their English language experiences

This section discusses our findings in relation to our fourth research question, namely *to explicate commencing students' views on their English language learning experiences over one semester*. Four themes consistently emerged in the qualitative data.

First, students seemed to be aware of the complex relationship between various "types" of proficiency and believed that these affected learning. They referred to a general academic proficiency as measured by IELTS, an academic proficiency needed for disciplinary study and a more general proficiency for "everyday life". They also discussed the interconnectedness of these dimensions of proficiency, stating for example that one's ability to listen to and comprehend academic lectures or interact in tutorials was linked to the kinds of listening and speaking one did at home or in a part-time job.

Secondly, students did not appear to have unrealistic expectations of academic study, even if their perceptions of their English proficiency did not always match their levels as predicted by IELTS (explained in more detail in the following section). For example, students referred to academic reading as voluminous and requiring the difficult decoding of lengthy, complex and/or abstract text, and they perceived academic writing as time-consuming and culturally or rhetorically foreign. In general, they did not perceive themselves to be at a level of L2 mastery which would allow them to comfortably negotiate the challenges of academic life ahead.

Third, students were able to articulate a range of obstacles that hindered language development. Examples of this include: the role of colloquialisms, culture and L1 social groups in constraining the advancement of speaking skills; the effect of discipline-specific vocabulary on understanding written and spoken texts; the impact of native-speaker reticence to engage and provide feedback on the development of an error-reduced discourse; and the constraints of learning environments (eg tutorials) on language performance. Previous research has explored how these factors can be variables that impact on student success (Cotton & Conrow, 1998; Elder & O'Loughlin, 2003; Haugh, 2008; Ingram & Bayliss, 2007; Kerstjen & Nery, 2000; Lobo, 2012; O'Loughlin & Arkoudis, 2009).

Finally, students were able to articulate a range of strategies that they had developed to raise their proficiency while studying at university. These strategies included: becoming accustomed to the amount of reading through experience; receiving explicit instruction on academic writing; mixing with local students/people; identifying and acquiring discipline-specific vocabulary; listening to local media; and using Lectopia technologies. Participants were aware of the importance of communicating in English as much as possible, particularly for improving their speaking skills. Participants who lived, worked or socialised with others who spoke English acknowledged that being "forced" to speak the language helped them to improve their communication skills and proficiency in English.

A general perception was that those students who were motivated to undertake activities outside the university using English were more likely to improve their proficiency, even if they found it difficult to understand the Australian accent and colloquial language used by the local community.

6.5 Students' perceptions of proficiency compared with proficiency as shown by IELTS

This section discusses our findings in relation to our fifth research aim, namely to investigate similarities and differences between students' perceptions of learning English for university study and their language proficiency as shown through IELTS test scores.

The key finding was that there appeared to be some degree of divergence between the self-reported perceptions of students about changes in their level of proficiency in the four macro-skills and the mean IELTS scores of the larger cohort from which they were drawn. The students reported for instance that their listening and reading had improved over the semester, yet the mean IELTS scores for Listening and Reading were only marginally better in Test 2 at the end of their first semester of study. Some of them also claimed that their writing had improved, yet there was also only a marginal increase in the mean IELTS score for Writing. Another perception was that while they found speaking the least difficult at the beginning of the semester, by the end of the semester they had changed their view that speaking was the least problematic macro-skill. Once again, this diverged from the test score results, which found a significant increase in mean score for Speaking between Tests 1 and 2.

There are a number of possible explanations for these apparent divergences. The first possibility is that these reported improvements in listening, reading and writing were not sufficiently large to impact on the IELTS band scores of Listening, Reading and Writing. The second possibility is that the students were not able to accurately self-report their level of proficiency in the four macro-skills. While these two factors no doubt played some part in these divergences, we would suggest from close analysis of the responses of students in the focus groups that the students were in fact talking about various dimensions of proficiency, only some of which are encompassed by the IELTS Academic test. The students seemed to be aware of distinctions between regular language proficiency, the kind of "general academic" language proficiency they had previously acquired, and the more discipline-specific language proficiency required for study at university. The IELTS Academic test is primarily focused on the second broad dimension or type of proficiency, although certain sections of the test relate to the first. In other words, proficiency is not a straightforward, unidimensional construct. It encompasses a complex array of different dimensions that become more or less salient depending on the context in which the construct of proficiency is being situated.

Thus, while "general academic" proficiency may be most salient in the case of pre-sessional students, in the case of students commencing their studies at university, discipline-specific language proficiency also comes to the fore and, arguably, regular everyday language proficiency, as they perhaps have more opportunities to interact with local students.

After graduation, on the other hand, yet another dimension of proficiency, namely the professional communication skills that Murray (2010) makes reference to become more critical. A key finding here is that while students may not be able to reliably assess their own level of proficiency, which is understandable, they *are* aware of these kinds of distinctions. The upshot of this is that proficiency is ultimately a complex and contested notion, a point which is not always well appreciated by all stakeholders.

6.6 Limitations

It has been previously stated that the overall sample size for this study was small and the recruitment of participants challenging. Both Craven (2012) and O'Loughlin and Arkoudis (2009) also found participant recruitment to be problematic and, as a result, only managed to test small numbers in their studies. One explanation for the difficulty of recruitment in our study is an understandable lack of motivation to sit a test for research purposes when the score is not directly useful for the participant. Storch and Hill (2008) state that:

One problem with studies which compare pre- and post-test scores is that they are based on the assumption that all participants will be equally motivated to complete the test to the best of their ability on both occasions. Test-takers tend to perform better on a test when the results have high stakes. (p413)

Engagement at the final stage of a degree is likely to be greater as students may see the test as a useful tool for future employment or migration purposes in the Australian context and we expect greater engagement for Phase 2 of the study.

Although we found some evidence of lack of engagement for certain sections of the test, fluctuations in motivation were not necessarily systematic and appeared opaque. Focus groups did indicate students were concerned about their language proficiency, yet in some instances marked changes in scores across the tests were evidence of a lack of concern about the results of the IELTS test. Additionally, those who had entered the university by IELTS had been required to evidence the minimum requirement of Overall 6.0 (no subscore below 5.5) and one of the major pathways in the study also requires evidence of a formal test score of 5.5 (no subscore below 5.5) of maximum one-year validity for entry to the program. In reality, the scores in the study were likely to have been depressed overall but this was less of a concern as we were investigating relative change.

Participant attrition was also of some concern to this study, as both the IELTS testing and the focus groups saw students drop out of the research for a variety of reasons. This is often a factor affecting longitudinal research as participants shift their focus or encounter problems which make it impossible for them to continue their participation. As a result of the sample sizes, some of the data should be viewed with caution.

Familiarity with the IELTS test was a variable that was not controlled in the study. Students were not required to prepare for either Test 1 or Test 2 though they could have opted to do so individually. While the IELTS pathway students (n=16) are likely to have prepared for Test 1, the other two pathways (n=35) may well not have prepared at all as they were enrolled in programs that provided entry without a formal test. It is possible that some of these participants had never taken IELTS before. At Test 2, it is highly likely that students did not prepare for the test as the score had no institutional implications at the end of the first semester of study. Additionally, they may have forgotten some aspects of the test, such as the importance of time management in the Writing test, having spent a semester concentrating on the requirements of university study. While it is not necessary to complete a preparatory course to score well in IELTS, familiarity with the tasks is considered to be advantageous for the test-taker. The participants were purposely not offered workshops for this research as the researchers believed that this may have unduly influenced test outcomes. In so doing, it was hoped that the scores would more accurately reflect participants' true proficiency.

The focus group interview data are limited by the relatively small sample size. As with the quantitative data, participant attrition was also a factor in collecting focus group data; several participants who attended the initial round of focus groups did not attend the final round, and new volunteers had to be sought. Hence, the descriptive findings should be read as suggestive of trends rather than as definitive results.

7 CONCLUSION

It is often assumed that international students entering their first year of study are relatively uniform in their level of English language proficiency. Our study indicates that there is a great deal of variation, not only amongst students but also between the scores in the four macro-skills of the same student. Consistent with other studies, we have found that while some students improve their English language proficiency (as measured by IELTS) over the course of their first semester of study, others do not, and some even appear to regress. We would suggest that this variability in English language proficiency is a reality that universities must come to grips with.

We have also suggested that we need to focus research on English language proficiency at particular times in the "life cycle" of a university student. In our study we have focused on general academic proficiency in their initial semester of study. However, English language proficiency clearly means something different for various stakeholders during students' subsequent two to three years of undergraduate study, where there is much greater emphasis on discipline-specific language proficiency, particularly by academics. On graduation, however,

there is more likely to be emphasis on several dimensions of language proficiency including general proficiency, general academic proficiency and discipline-specific proficiency, particularly by employers and members of the community. International students appear to have some awareness of these different views on proficiency. Research in this area thus needs to reflect the complex and contested nature of proficiency.

We would further suggest that the strong correlation between scores in Listening and Reading and GPAs of students in their first year of study (in contrast to the lack of correlation between their GPAs and scores in Speaking and Writing) possibly points to the need to place greater emphasis on minimum entry scores for Listening and Reading. While these findings would need to be replicated in a larger sample if they are to properly influence university policies on English language requirements, it is interesting to note that we have found evidence in our study that scores in Listening and Reading should not be interpreted in the same way as scores in Speaking and Writing by stakeholders, including university administrators.

We noted at the outset of this report that this study of changes in English language proficiency over the initial semester of undergraduate study is part of a larger, longitudinal study of changes in English language proficiency over the course of undergraduate study. While we would expect to see greater evidence of improvement in English language proficiency over the course of a whole degree program, which can vary from two to three years depending on the students' prior study, the jury remains out on the degree and nature of this improvement. The lesson from this study, and the research literature more broadly, is that any such results need to be interpreted as reflecting a complex tapestry of multiple intersecting conceptualisations of proficiency and multiple underlying variables.

ACKNOWLEDGMENTS

The research team is grateful to the IELTS partners for funding towards this study. We wish to thank the English Language Working Party for their co-funding of the project, staff at the Griffith IELTS Test Centre for their assistance as well as the students who participated in the research. We also gratefully acknowledge the statistical advice provided by Dr Helen Klieve.

REFERENCES

- Agar, M & MacDonald, J, 1995, 'Focus groups and ethnography', *Human Organisation*, 54(1), pp78–86
- Allwright, J & Banerjee, J, 1997, *Investigating the accuracy of admissions criteria: A case study in a British university*, Lancaster: Centre for Research in Language Education, Lancaster University
- Archibald, A, 2001, 'Managing L2 writing proficiencies: Areas of change in students' writing over time' in *International Journal of English Studies*, 1(2), pp153–174
- Arkoudis, S, Atkinson, C, Baird, J, Barthel, A, Ciccarelli, A, Ingram, D, Murray, D & Willix, P, 2008, *Good practice principles for English language proficiency of international students in Australian universities, Final report*, Department of Employment, Education and Workplace Relations, Australian Government
- Arkoudis, S & Starfield, S, 2007, 'In-course language development and support', paper presented at the *National Symposium: English Language Competence of International Students*, Sydney, August 2007
- Australia Unlimited, 2012, retrieved from <<http://www.australiaunlimited.com/society/educations-unlimited-potential>>
- Australian Bureau of Statistics, 2011, retrieved from <<http://www.abs.gov.au/ausstats>>
- Australian Education International (AIE), 2012, International student data 2012, retrieved from <<https://www.aei.gov.au/research/International-Student-Data/Pages/InternationalStudentData2012.aspx>>
- Avdi, E, 2011, 'IELTS as a predictor of academic achievement in a Master's Program' in *EA Journal*, 26(2), pp42–49
- Barrett-Lennard, S, Dunworth, K & Harris, A, 2011, 'The Good Practice Principles: Silver bullet or starter gun?', *Journal of Academic Language & Learning* 5(2), ppA-99–A-106
- Barthel, A, 2011, 'Good practice in academic language development: From principles to standards', paper presented at the *10th Biennial Conference of the Association for Academic Language and Learning*, University of Adelaide, November 2011.
- Benzie, H, 2010, 'Graduating as a 'native speaker': International students and English language proficiency in higher education', *Higher Education Research & Development*, 29(4), pp447–459
- Berry, B, & Lewkowicz, J, 2000, 'Exit-tests: Is there an alternative?' in *Hong Kong Journal of Applied Linguistics*, 5(1), pp19–49.
- Birrell, B, Hawthorne, L & Richardson, S, 2006, *Evaluation of the General Skilled Migration categories report*, Australian Government: The Department of Immigration and Citizenship (DIAC) Report, retrieved from <<http://www.immi.gov.au/media/publications/research/gsm-report/index.htm>>
- Bradley, D, Noonan, P, Nugent, H & Scales, B, 2008, *Review of Australian Higher Education*, retrieved from <<http://www.deewr.gov.au/HigherEducation/Review/Pages/ReviewofAustralianHigherEducationReport.aspx>>
- Briguglio, C, 2011, 'Quality and the English language question: Is there really an issue in Australian universities?' in *Quality in Higher Education*, 17(3), pp317–329
- Cho, Y & Bridgeman, B, 2012, 'Relationship of TOEFL iBT scores to academic performance: Some evidence from American universities' in *Language Testing*, 29(3), pp421–422
- Coley, M, 1999, 'The English language entry requirements of Australian universities for students of non-English speaking background' in *Higher Education Research & Development*, 18(1), pp7–17
- Cotton, F & Conrow, F, 1998, 'An investigation into the predictive validity of IELTS amongst a group of international students studying at the University of Tasmania' *IELTS Research Reports Volume 1*, IELTS Australia Pty Limited, Canberra, pp72–115
- Council of Australian Governments (COAG), 2010, *International students strategy for Australia*, retrieved from <http://www.coag.gov.au/reports/docs/aus_international_students_strategy.pdf>
- Craven, E, 2012, 'The quest for IELTS 7.0: Investigating English language proficiency of international students in Australian universities', *IELTS Research Reports Volume 13*, IELTS Australia, Canberra and British Council, London, pp1–61
- Creswell, JW, 2008, *Educational research: Planning, conducting and evaluating quantitative and qualitative design*, Pearson Education, New Jersey

- Criper, C & Davies, A, 1988, 'ELTS validation project report', *ELTS Research Report 1*(1), The British Council and University of Cambridge Local Examinations Syndicate, Cambridge
- Devos, A, 2003, 'Academic standards, internationalisation and the discursive construction of 'the international student'' in *Higher Education Research & Development*, 22(2), pp155–166
- Dooley, P, 1999, 'An investigation into the predictive validity of the IELTS Test as an indicator of future academic success' in *Teaching in the disciplines/ learning in context*, eds K Martin, N Stanley & N Davison, pp114–118, Proceedings of the 8th Annual Teaching Learning Forum, The University of Western Australia, February 1999, retrieved from <<http://lsn.curtin.edu.au/tlf/tlf1999/dooley.html>>
- Dooley, P & Oliver, R, 2002, 'An investigation into the predictive validity of the IELTS Test', *Prospect*, 17(1), pp36–54
- Elder, C, 1993, 'Language proficiency as a predictor of performance in teacher education', *Melbourne Papers in Language Testing*, 2, pp68–85
- Elder, C & O'Loughlin, K, 2003, 'Investigating the relationship between intensive English language instruction and band score gain on IELTS', *IELTS Research Reports Volume 4*, ed R Tulloh, IELTS Australia Pty Limited, Canberra, pp207–254
- Elder, C, Erlam, R & Von Randow, J, 2002, 'Enhancing chances of academic success amongst first year undergraduates from diverse language backgrounds', *The 6th Pacific Rim - First Year in Higher Education Conference: Changing Agendas "Te Ao Hurihuri"*, The University of Canterbury in conjunction with Queensland University of Technology, Christchurch, New Zealand, July 2002
- Feast, V, 2002, 'The impact of IELTS scores on performance at university', *International Education Journal*, 3(4), pp70–85
- Fenton-Smith, B, 2012, 'Facilitating self-directed learning amongst international students of health sciences: The dual discourse of self-efficacy', *Journal of Academic Language and Learning* 6(1), pp64–76
- Graham, J, 1987, 'English language proficiency and the prediction of academic success', *TESOL quarterly*, 21(3), pp505–521
- Green, T, 2004, 'Making the grade: Score gains on the IELTS writing test', *Research Notes*, 16, University of Cambridge ESOL examinations, pp9–13
- Green, A, 2005, 'EAP study recommendations and score gains on the IELTS Academic writing test', *Assessing Writing*, 10, pp44–60 (DOI:10.1016/j.asw.2005.02.002)
- Green, A & Weir, C, 2002, *Monitoring score gain on the IELTS Academic writing module in EAP programmes of varying duration, Phase 1 report*, University of Cambridge Local Examinations Syndicate, Cambridge
- Green, A & Weir, C, 2003, *Monitoring score gain on the IELTS Academic writing module in EAP programmes of varying duration, Phase 2 report*, University of Cambridge Local Examinations Syndicate, Cambridge
- Griffith Fast Facts, 2012, retrieved from <<http://www.griffith.edu.au/about-griffith/fast-facts>>
- Griffith University Capability Statement, 2011, retrieved from <http://www.griffith.edu.au/__data/assets/pdf_file/0003/355836/Briefing-on-Griffith-Uni-rankings,-strengths-and-credentials-2011-26-10-Oct.pdf>
- Haugh, M, 2008, 'The discursive negotiation of international student identities', *Discourse: Studies in the Cultural Politics of Education*, 29(2), pp207–222
- Hawkins, J, 2010, 'Criterial features in the learning of English', *English Australia conference proceedings*, Gold Coast, Australia, September 2010
- Hawthorne, L, 2007, 'Outcomes: Language, employment and further study', paper presented at the *National Symposium: English Language Competence of International Students*, Sydney, August 2007
- Hirsh, D, 2007, 'English language, academic support and academic outcomes: A discussion paper', *University of Sydney Papers in TESOL*, 2(2), pp193–211
- Humphreys, P & Mousavi, A, 2010, 'Exit-testing – a whole of university approach', *Language Education in Asia*, 1(1), pp8–22, retrieved from <http://dx.doi.org/10.5746/LEIA/10/V1/A03/Humphreys_Mousavi>
- Ingram, D & Bayliss, A, 2007, 'IELTS as a predictor of academic language performance, Part 1', *IELTS Research Reports Volume 7*, IELTS Australia Pty and British Council, Canberra, pp137–199
- Kerstjen, M & Nery, C, 2000, 'Predictive validity in the IELTS test', *IELTS Research Reports Volume 3*, IELTS Australia Pty Limited, Canberra, pp85–108

- Kidd, PS & Parshall, MB, 2000, 'Getting the focus and the group: Enhancing analytical rigor in focus group research', *Qualitative Health Research*, 10(3), pp293–308
- Kitzinger, J, 1996, 'Introducing focus groups', in *Qualitative Research in Healthcare*, eds N Mays & C Pope, BMJ Publishing, London, England, pp36–45
- Light, RL, Xu, M & Mossop, J, 1987, 'English proficiency and academic performance of international students', *TESOL Quarterly*, 21(2), pp251–261
- Lobo, A & Gurney, L, 2012, *An investigation of the links between international students' expectations and reality in the English Language Enhancement Course*, manuscript submitted for publication (copy on file with authors)
- McCarthy, M, 2011, 'Learner grammars based on corpus evidence', paper presented at *English Australia Conference*, Adelaide, September 2011
- Moore, T, 2004, 'Dimensions of difference: A comparison of university writing and IELTS writing', *Journal of English for Academic Purposes*, 4(1), pp43–66, (DOI: 10.1016/j.jeap.2004.02.001)
- Murray, D & O'Loughlin, K, 2007, 'Pathways - Preparation and Selection', paper presented at the *National Symposium: English Language Competence of International Students*, Sydney, August 2007
- Murray, N, 2010, 'Considerations in the post-enrolment assessment of English language proficiency: Reflections from the Australian context', *Language Assessment Quarterly*, 7(4), pp343–358 (DOI: 10.1080/15434303.2010.484516)
- O'Loughlin, K & Arkoudis, S, 2009, 'Investigating IELTS exit score gains in higher education', *IELTS Research Reports Volume 10*, ed J Osborne, IELTS Australia, Canberra and British Council, London, pp1–90
- O'Loughlin, K, 2011, 'The interpretation and use of proficiency test scores in university selection: How valid and ethical are they?', *Language Assessment Quarterly*, 8(2), pp146–160 (DOI:10.1080/15434303.2011.564698)
- Qian, D, 2007, 'Assessing university students: Searching for an English language exit test', *RELC Journal*, 38(1), pp18–37, (DOI:10.1177/0033688206076156)
- Read, J & Hayes, B, 2003, 'The impact of the IELTS test on preparation for academic study in New Zealand', *IELTS Research Reports Volume 4*, ed R Tulloh, IELTS Australia Pty Limited, Canberra, pp153–206
- Rochecoste, J, Oliver, R, Mulligan, M & Davies, M, 2010, *Addressing the ongoing English language growth of international students*, Australian Government Office for Learning and Teaching
- Rose, D, Rose, M, Farrington, S, & Page, S, 2008, 'Scaffolding academic literacy with indigenous health sciences students: An evaluative study', *Journal of English for Academic Purposes*, 7, pp165–179
- Storch, N & Hill, K, 2008, 'What happens to international students' English after one semester at university?' *Australian Review of Applied Linguistics*, 31(1), pp04.1–04.17
- Tashakkori, A & Teddlie, C, 2003, *Handbook of mixed methods in social and behavioural research*, Sage, California
- Ushioda, E & Harsch, C, 2011, *Addressing the needs of international students with academic writing difficulties: Pilot project 2010/11, Strand 2: Examining the predictive validity of IELTS scores*, retrieved from <http://www2.warwick.ac.uk/fac/soc/al/research/groups/ellta/projects/strand_2_project_report_public.pdf>
- Woodrow, L, 2006, 'Academic success of international postgraduate education students and the role of English proficiency', *University of Sydney Papers in TESOL*, 1, pp51–70

APPENDIX

Focus Group Interview Protocol

Stage 1: Early in Semester

Introduction

My name is X and I am working with the School of Languages and Linguistics at Griffith University on this research project. I would like to ask you some questions about how you are developing your English language skills here at Griffith. I'd also like to ask you some questions about the Language and Communication courses that you are all taking.

This interview is being recorded, but no-one else will know who is speaking. Your names won't be used in this research. This interview has nothing to do with your grade in your Language and Communication course. Do you have any questions before we start?

1. Opening questions

What country are you from?

How long have you been in Australia?

What's your major subject at Griffith University?

Which English skills (speaking, listening, reading or writing) are easiest for you? Which are most challenging?

2. Questions about using English in Brisbane and at Griffith University

When you began this course did you think your English was at a high enough level for studying in Australia? Why/why not?

Do you feel that it is important to improve your English skills while you are at university? If so, is it the university's responsibility to help you do this, or is it your responsibility to do it yourself? Or both?

What factors at the university have helped you to improve your English (e.g. talking with Australian friends)? How have they helped you?

What factors at the university have prevented you from improving your English? How have they prevented you?

What opportunities do you have for speaking or listening to English outside university?

What factors outside the university have helped you to improve your English (e.g. talking with Australian friends)? How have they helped you?

What factors outside the university have prevented you from improving your English? How have they prevented you?

3. Questions about the Language and Communication course

What do you think about taking the Language and Communication course?

Do you think this course will help you to study in your own discipline? If yes, in what ways?

If no, why not?

Do you think this course will help you to improve your English language proficiency? If yes, in what ways?

If no, why not?

How useful do you think this course will be in helping you to improve your IELTS score?

Focus Group Interview Protocol

Stage 2: Late in Semester

Introduction

My name is X and I am working with the School of Languages and Linguistics at Griffith University on this research project. I would like to ask you some questions about how you are developing your English language skills here at Griffith. I'd also like to ask you some questions about the Language and Communication courses that you are just completing.

I'd just like to remind you that this interview is being recorded, but no-one else will know who is speaking. Your names won't be used. This interview is not connected to your grade in your Language and Communication course. Do you have any questions before we start?

1. Opening questions

Did you come to the early-semester focus group session? Did you come to the mid-semester session?

What country are you from? *(for first-time participants)*

How long have you been in Australia? *(for first-time participants)*

What is your major subject at Griffith University? *(for first-time participants)*

Which English skills (speaking, listening, reading or writing) are easiest for you? Which are most challenging? *(for first-time participants)*

2. General questions about the Language and Communication courses

At the beginning of this course, what kind of skills or knowledge were you expecting to learn?

Has the course met your earlier expectations? If not, how is it different?

How useful do you think the course has been for study in your own discipline?

How useful do you think the course has been for improving your English skill? Which skills have increased, if any? Which skills have not increased?

How useful do you think the course has been for improving your IELTS score?

Can you think of any other language or academic skills that you would have liked to learn about in this course?

Can you think of any particular aspects of the course that should be removed or changed?

3. Questions about assessment tasks

How useful did you think the portfolio tasks were for learning about study in your discipline and for improving your English skills?

How useful did you think the oral presentation was for learning about study in your discipline and for improving your English skills?

How useful did you think the university service reflection task was for learning about study skills or for improving your English skills?

(For 5904LAL only) How useful did you think the quizzes were for learning about study skills or for improving your English skills?

4. Questions about other English language resources

Have you accessed the English language resources at Griffith University such as English HELP, Learning Services workshops or Student LINX this semester? If so, how often? If not, why not?

[If the respondent has previously accessed these resources] How useful do you think these resources have been for you? Which resource has been most useful for you?

Have any other factors at Griffith University helped you to improve your English during this semester? If so, how have they helped you?

5. Questions about learning and using English

In general, do you feel that your English language ability has improved over this semester? Or has it gotten worse? Or is it the same as previously?

In the two previous sessions, we asked you if you felt that you needed to improve your English skills while you are at university. We also asked whether you thought it was the university's responsibility to help you do this, or whether you thought it was your own responsibility. What is your attitude towards this issue now? Has it changed, or is it the same? *[Encourage Ss to elaborate.]*

Have you had any new opportunities for speaking/listening to English outside university since the semester started? *[Encourage Ss to elaborate.]*

Do you plan to do anything in the future to keep up your English skills? If so, what do you plan to do?