Research in Hospitality Management 2012, 1(2): 53–63
Printed in South Africa — All rights reserved

RHM
ISSN 2224-3534
http://dx.doi.org/10.2989/RHM.2012.1.2.1.1117

The moving finger: Asian international hotel school students developing English language writing skills

Peter Graham

Stenden Rangsit University, Rattanakunakorn Building, 52/347 Muang Ake, Phaholyothin Road, Lak-Hok, Pathumthani, Bangkok, Thailand
E-mail: pfr.graham@gmail.com

The number of international English as a Second Language hospitality students completing their degree programs in Australasia has risen dramatically in the past decade. One factor that may be motivating students to undertake tertiary studies in Australasia is the expectation that this form of immersion in English language degree programmes will lead to improved English language skills. However, existing research, such as research on the impact of study abroad programs on learners' second language (L2) skills, has produced mixed findings. Furthermore, most of this research has tended to focus on progress in L2 speaking skills. To date there has been very little research on the impact of living and studying in the L2 environment on learners' L2 writing development. This paper reports on a study that used a test–retest design to investigate changes, if any, in the learners' academic writing after one semester of study in an L2-medium university. The writing scripts of 25 students who did not access the formal language support programs offered by the university were analysed using a range of qualitative and quantitative measures. The study found that after a semester of study at the university, the learners' writing improved mainly in terms of structure and development of ideas. There was also some improvement in the formality of learners' language, but there was no evidence of improvement in linguistic accuracy or complexity. Strategies used to incorporate source materials also remained largely unchanged, with learners continuing to copy verbatim from sources and acknowledging sources incorrectly. A number of factors are put forward to explain these findings. These factors include the short duration of the study (one semester) as well as perhaps the absence of feedback.

Keywords: English language skills, hospitality students, impact of studying abroad, improvement, L2

De bewegende vinger: Aziatische internationale hotelschool studenten ontwikkelen Engelse schrijfvaardigheid

Het aantal internationale ESL hotelschool studenten die hun opleiding in Australasia afronden, is dramatisch gestegen in het afgelopen decennium. Eén factor die mogelijk leerlingen motiveert om een tertiaire studie in Australië te volgen, is hun verwachting dat deze Engelstalige opleidingen zal leiden tot een verbeterde kennis van de Engelse taal. Echter, resulteerde eerder onderzoek, zoals een onderzoek naar de invloed van een studie in het buitenland op L2 vaardigheden, in verschillende bevindingen. Daarnaast heeft het grootste deel van dit onderzoek zich vooral gericht op de vooruitgang in de L2 spreekvaardigheid. Tot op heden is er zeer weinig onderzoek gedaan naar de invloed van wonen en studeren in een L2 omgeving op de ontwikkeling van de L2 schrijfvaardigheid van de studenten. Dit artikel bespreekt een onderzoek die een test – hertest onderzoeksmethode gebruikt heeft, om zo eventuele veranderingen te onderzoeken bij cursisten, in het academisch schrijven, na een semester studie in een L2 omgeving. De verslagen van 25 studenten die geen toegang hadden tot de formele taal ondersteuningsprogramma's van de universiteit, werden geanalyseerd met behulp van een scala aan kwalitatieve en kwantitatieve maatstaven. Dit onderzoek wees uit dat de schrijfvaardigheid van studenten, na een semester studie aan de universiteit, vooral verbeterd was op het gebied van structuur en ontwikkeling van ideeën. Er was ook enige verbetering in het gebruik van formaliteit in de taal van de leerlingen, maar er waren geen tekenen van verbetering in de taalkundige juistheid of complexiteit. Strategieën met betrekking tot het verwerken van bronnen op de juiste manier bleven ongewijzigd, want studenten bleven bronnen letterlijk kopiëren en onjuist vermelden in hun bronvermelding. Een aantal factoren zijn naar voren gebracht om deze bevindingen te verklaren. Deze factoren omvatten onder andere de korte duur van het onderzoek (een semester) en de afwezigheid van feedback lijkt ook een oorzaak te kunnen zijn.

Trefwoorden: de invloed van studeren in het buitenland, gastvrijheid studenten, kennis van het Engels, L2, verbetering

幕后推手：亚洲国际酒店管理学院的学生提高英语写作 技巧

过去十年里， 选择在澳大利亚完成国际（非母语英语课程）酒店管理专业的学生数量显著增加。其中一个可能促使学生想在澳大利亚接受高等教育的原因是，他们期望这种沉浸在用英语学习专业课程中的方式能够提升他们的英语水平。然而现有的， 诸

如出国留学对学习者外语技能影响的研究，往往是结论不一； 而且， 此类研究多倾向于关注学习者外语口语能力的进步。迄今为止， 甚少有关于生活， 学习于目的语环境是否于发展学习者第二语言写作能力的研究。本文采取了测试-复测的研究设计，来调查经过一个学期目的语环境的大学生活后， 学习者学术写作能力的变化（假设有变化的话）。本文的研究样本是由大学提供的25份未接受过正式语言支持服务的学生的手稿；采用的是一系列定量与定性相结合的分析方法。研究显示， 经过在大学一个学期的学习后， 学生写作能力的进步主要体现在结构和构思上；此外， 也在礼节上有了进步。但是， 却没有任何证据显示他们是否在语言的精准度或表达的复杂度上有进步。那些用于整合吸收原材料的策略基本没有大的变化， 学习者还是逐字复制输入源并进行错误的反馈。笔者就这一发现给出了一些解释：例如学习的时限太短（通常是一学期）， 或许还因为意见反馈的缺乏。

关键词：酒店专业的学生，进步，第二语言，英语水平，出国留学的影响

Introduction

In the past decade, Australasian universities offering programs in English have experienced an exponential growth in the number of international students. For example, the University of Melbourne (2006) reported an 88% growth in its international student enrolments since 2000. A report by the Australian Department of Foreign Affairs shows that international education is now Australia's third-largest export industry (Australian Education International, 2008). Similar trends are found at universities in Asia with programs taught in English. Many of the students on international campuses are drawn from Asia, from countries such as China, Malaysia, Thailand, Indonesia and Vietnam.

In response to the growing number of international students and their perceived needs, universities catering to international students, as well as universities with Asian subsidiaries and who teach in English, have developed a range of language and academic support programs (see Melles et al., 2005). For example, the university in which this study was conducted offers a free-of-charge diagnostic test to incoming international students. The test results are used to generate recommendations regarding the type of language support, if any, that the student is likely to need. A range of support options is available: English for Academic Purposes (EAP) credit subjects, non-credit-bearing workshops and short courses, as well as individual consultations with language tutors who offer an editing type of service. However, neither the test nor the recommended support options are mandatory, and indeed many students do not follow the recommendations because of timetabling constraints (on enrolling in an EAP subject concurrently) and a perceived lack of time to attend workshops and consultations (see Storch and Hill, 2008). Other studies (e.g. Hirsch, 2007) also report low uptake of support options for similar reasons.

One of the assumed advantages of studying at an Australasian university with an international programme is that this immersion experience integrated with formal study will lead to improved English language skills. This is a reasonable expectation. Living and studying in the second language (L2) environment provides learners with exposure to rich and authentic language input and with opportunities to produce extensive and meaningful language output. In theories of second language acquisition, exposure to such input (Gass, 2003; Krashen, 1985) and practice in producing language (Swain, 1985; Swain and Lapkin, 1995) are generally accepted as essential conditions for successful L2 acquisition. However, output practice is now regarded as more important than input (e.g. Ellis, 2003), particularly for the development of productive skills such as writing (DeKeyser, 1997). Researchers building on Swain's work (e.g. Cumming, 1990; Muranoi, 2007) have argued that the need to produce language pushes learners to process language syntactically; that is, pay attention to the means of expression necessary to convey their intended meaning. For example, Cumming (1990: 483) writes: 'Composition writing elicits an attention to form-meaning relations that may prompt learners to refine their linguistic expression – and hence their control over their linguistic knowledge.'

However, findings from empirical research on the impact of living and studying in the L2 environment on students' L2 skills are mixed (see reviews in DeKeyser, 2007b; Freed, 1995), depending largely on how progress is measured. Moreover, where improvement in L2 is reported, it is mainly in speaking skills and for programs longer than one semester. DeKeyser (2007b) notes that research on the effects of study abroad on L2 listening, reading and writing skills is extremely scarce.

Studies that have investigated the effects of in-country intensive EAP courses that prepare students for further study or for proficiency tests such as the International English Language Testing System (IELTS) have also produced somewhat mixed results. These studies have tended, in the main, to examine the effects of the courses on learners' overall L2 proficiency. For example, studies by Read and Hayes (2003) in New Zealand and by Green and Weir (2003) in the United Kingdom reported marginal improvements in English language proficiency following intensive IELTS preparation and EAP courses. In contrast, Elder and O'Loughlin (2003), in a study conducted in New Zealand and Australia, reported a half-band improvement on IELTS following intensive EAP courses. Elder and O'Loughlin also noted that improvement was greatest on the listening subtest and evident in average gains for the entire cohort but that individual performances varied considerably.

Shaw and Liu's (1998) study, unlike the studies discussed above, investigated developments in learners' L2 writing skills. The researchers compared learners' writing in terms of a large number of linguistic features before and after a full-time preparatory EAP course (2–3 months long) in the United Kingdom. They reported that, although the learners' writing showed no significant changes in linguistic accuracy or complexity, the writing did become more formal, employing language associated with written rather than spoken language, such as fewer personal pronouns and contractions.

Hinkel's (2003) study, on the other hand, found that the writing of English as a Second Language (ESL) learners showed features that create an overall impression of text simplicity even after a long period of living and studying in the USA. Hinkel's large-scale study compared the writing produced by advanced ESL learners, most of whom had spent four years studying in USA community colleges and high schools, with the writing of native-speakers of English beginning their university study. Hinkel found that the ESL learners' writing showed a higher frequency of simple syntactic and lexical constructions (e.g. sentences with 'be' as the main verb, and vague nouns such as 'people') that are often associated with informal discourse. Thus, Hinkel suggests that the in-country experience provides ESL learners with a great deal of exposure to informal conversational discourse and that this may explain the prevalence of informal expressions and constructions in the learners' writing.

The above two studies focused only on the linguistic features of writing, and as such have considered only one aspect of what is considered 'good academic writing'. Influential theories of writing, such as genre theory (e.g. Christie, 1998; Hyland, 2003) and a growing volume of research on L2 writing (see an extensive review in Silva and Brice, 2004; Candlin and Hyland, 1999) have stressed that writing is a multidimensional, sociocognitive activity, where the processes involved and the features of the text produced are very much shaped by sociocultural norms and interpersonal relationships within the context in which the writing takes place. From this perspective, academic writing generally involves cognitive activities such as reading and synthesising information from a variety of sources, and producing a text that shows evidence of features associated with 'good academic writing' in an anglophone context. These features, evident in various writing assessment schemes (see Weigle, 2002), include, for example, a clear introduction and conclusion, the use of a formal register, and correct citations of the work of other authors.

The limited research on the effects of studying in an L2 setting on learners' L2 writing and the low uptake of support options provided by the university, mentioned earlier, provided the impetus for this study. The focus of the study was on the learners' writing, given that most assessment tasks are written assignments and exams. The study sought to investigate what features of academic writing, if any, develop as a result of studying in a degree program in an L2-medium university after one semester (10 weeks) and in the absence of formal language support.

Methods

The study used a test–retest design. The test used was a diagnostic test developed by the university's testing staff. Although the test is available to all incoming international students, it is generally students who have reached only the minimum threshold required for university entrance (i.e. IELTS score of 6.5) who are strongly encouraged by their faculties to sit the test. The main aim of the diagnostic test is to identify students who may require further language support and to recommend the appropriate form of support.

Participants

The data of 25 students were used in this study. This data came from a larger data set (n = 39) that investigated L2 developments in reading and writing of a group of international students after a semester of study. The 25 participants selected for this study were those who, despite recommendations generated by their test results suggesting that they should seek formal ESL support, sought only minimal or no help with their English during their semester of study at the university. The questionnaire the students completed as part of the study showed that of the 25 participants, 21 did not access any formal support. The other four students reported that they had attended a limited number of workshops and/or individual consultations with a language advisor; however, the support they received amounted to less than 5 h in total for the semester.

All 25 participants (17 females, eight males) were from Asia, from countries such as China, Thailand, Indonesia and Vietnam. All were graduate students studying hotel management.

Data

The main source of data was the essays produced by the learners on the diagnostic test on two occasions: just prior to the beginning of the semester (Time 1) and 12 weeks later, towards the end of the semester (Time 2). The writing test requires students to write an argumentative essay of at least 300 words, and they are given 55 min to do so. They are advised to spend 5 min on reading six short excerpts, 5 min on planning their essay, and 45 min on writing. They are also instructed that they can refer to the ideas contained in the excerpts, but that they should not copy phrases from the excerpts. The same version of the test was used on both occasions (although the students were not aware of that in advance). The essay topic concerned dietary procedures.

Another source of data used in this study was a background questionnaire the participants completed at Time 2. The questionnaire elicited information about the participants' first language, language learning background, and the type of ESL support, if any, they had accessed.

Data analysis

The essays produced by the learners at Times 1 and 2 were assessed and analysed using a range of quantitative and qualitative approaches. In analysing the essays, salient features of the writing were noted and subsequently subjected to further examination. For example, it became apparent that some learners relied on the language of the excerpts provided in their essays and that citations were not always accurate. Thus, the learners' essays were also analysed for use of source materials. The following subsections describe how the essays were assessed and analysed. Table 1 presents a summary of the approaches and measures used.

A closer analysis of language use, using a range of quantitative linguistic measures, identified aspects of language use that showed some improvements and others that showed no improvement over time. Descriptive statistics and, where relevant, statistical tests of significance are reported. Given the small sample size (n = 25), the large number of variables investigated, and the exploratory nature of the study, paired-sample t-tests were used (with conservative p values) rather than the more powerful MANOVA followed by ANOVAs and *post hoc* tests. When the differences in means seemed quite small, no statistical tests of significance were conducted.

Analytic scores given by raters

All the essays were assessed on three analytic criteria: fluency, content and form. These criteria (and scale) were developed by the testing centre and are similar to the criteria used by IELTS. The criterion of fluency assesses how well the essay communicates its ideas; that is, the overall essay structure, coherence and cohesion. Content refers to how well ideas are developed, and form refers to grammatical accuracy, range of vocabulary and sentence structures. For each of these three criteria, the student received a score on a scale of 1 to 9 with 9 representing an advanced level of performance. The three scores were then averaged to yield a single global writing score (as is the case in IELTS). When assessing the essays for this study, all identifying information (participants' names and times) was removed. The essays were numbered and assessed by two experienced raters independently. Scores awarded were then compared and any discrepancies were resolved through discussion.

Ortega (2003) points out that holistic ratings are unlikely to distinguish very clearly between performances at different points of time. A similar argument can be made about analytic ratings. Both types of ratings describe the quality of performance and tend to collapse several features under a single rubric, thereby assuming that they all develop at the same time. For example, in the rating scheme used in this study the criterion of content includes reference to how well arguments are developed, relevance and sufficiency of ideas. Thus, a more fine-grained analysis of the elements mentioned in each of the criteria was undertaken to investigate development in writing over time.

Qualitative analysis of structure and content

A qualitative analysis of the structure included examining more closely the information contained in the introduction (whether it contained some general background information and a clear focus), whether and how cohesion was achieved between and within ideas, and the nature of the conclusion (whether it related to the arguments discussed in the essay). In the qualitative analysis of the essays' content, attention was paid to the number of arguments and to whether they were clearly identified and sufficiently developed.

Quantitative analysis of language use: measures of fluency, accuracy, and grammatical and lexical complexity

The learners' use of language received the most attention. A range of quantitative linguistic measures of writing fluency, accuracy, and syntactic and lexical complexity were employed. These measures required all essays to be coded for T-units, clauses, and errors, and they required word choice to be considered more closely.

A T-unit is defined by Hunt (1966: 735) as 'one main clause plus whatever subordinate clauses happen to be attached to or embedded within it'. Written scripts were also coded for clauses, the researchers distinguishing between independent and dependent clauses. An independent clause is one that can be used on its own (Richards et al., 1992). In this study a dependent clause was one that contained a finite verb and a subject (Wolfe-Quintero et al., 1998). Thus, dependent clauses included adverbial (subordinate), adjectival (relative) and nominal clauses.

For example, the following T-unit from the data contains two clauses (shown separated by a slash), an independent clause and a dependent (subordinate) clause (beginning with 'just because'): *People cannot become all vegetarians/ just because they have to respect animal rights*. Following Wolfe-Quintero et al. (1998), three measures of fluency were used in this study: the total number of words (W), the number of T-units, and the length of the T-units measured in words per T-unit (W/T). The standard computer word count tool was used for these measures.

In order to assess accuracy, all essays were coded for errors. Errors were also categorised using Chandler's (2003) and Bardovi-Harlig and Bofman's (1989) error taxonomies as a guide to classify and distinguish between errors in syntax (e.g. errors in word order, and incomplete sentences), errors in morphology (e.g. verb tense, subject verb agreement, and use of articles), and errors in lexis (word choice). All errors in spelling and punctuation were ignored to minimise the possible overestimation of errors due to unclear handwriting. A number of accuracy scores were then calculated: the proportion of error-free T-units (EFT/T), the proportion of error-free clauses (EFC/C), and the total number of errors per total number of words (E/W). The last measure was used in order to address the concern that ratio scores do not distinguish between units (e.g. T-units) containing multiple errors and units containing single errors.

Two measures of grammatical complexity were used in this study: the ratio of clauses per T-unit (C/T) and the ratio of dependent clauses per clause (DC/C). The latter measure is based on the assumption that a greater level of embedding and subordination represents greater syntactic sophistication (Shaw and Liu, 1998; Wolfe-Quintero et al., 1998).

Table 1: Measures used in assessing and analysing the essays. Essays were scored on a scale of 1 to 9 for each criterion

Assessment criterion	Qualitative analysis of structure and content		Quantitative measures		Use of sources		
	Structure	Content	Fluency	Accuracy	Grammatical complexity	Lexical complexity	
Fluency	Introduction	No. of arguments	No. of words (W)	No. of errors per words (E/W)	Clauses per T-units (C/T)	Percentage of words appearing in AWL	No. of quotes and paraphrases
Content	Text cohesion	Development of arguments	No. of T-units (T)	Error-free T-units per T-units (EFT/T)	Dependent clauses per clauses (DC/C)	Frequency count of informal expressions	Type of paraphrases
Form	Conclusion		Words per T-units (W/T)	Error-free clauses per clauses (EFC/C)		Borrowed words/ total words	Correct citation

Lexical complexity or sophistication is an important aspect of academic writing. However, the analysis of vocabulary used in writing by L2 learners is continuing to challenge researchers (e.g. see Duran et al., 2004; McCarthy and Jarvis, 2007). In the present study two approaches to assessing use of vocabulary were employed: calculating the number of academic words used as a proportion of total words, and counting the number of informal expressions. The academic word list (AWL) developed by Coxhead (2000) was consulted. The list contains 570 word families that have been found to account for approximately 10% of words in a range of academic texts. Thus, lexical sophistication was operationalised in this study as the percentage of words in the students' essays that were found on the AWL. The other measure used was a frequency count of informal expressions. Informal expressions seemed quite prevalent in some of the learners' essays. Following the work of Shaw and Liu (1998) and Hinkel (2003), informal expressions identified and counted included contractions (e.g. I'm, it's), colloquial expressions (e.g. a lot, lots of, kind of, stuff), use of first person pronouns (I, me, my), addressing the reader (e.g. you should consider…), vague words (e.g. people, things, some), and inappropriate use of questions and exclamations (e.g. Who knows what chicken disease is? That example is worse than jail!).

In order to check for inter- and intra-rater reliability in coding, and following the advice of Polio (1997), guidelines were formulated stating clearly what constituted a T-unit, a clause, and an error. To check for inter-rater reliability, a random sample of four essays were coded by a second researcher; to check for intra-rater reliability, a random sample of 10 essays were coded a week later by the researcher. Reliability scores were calculated using simple percentage agreement. Inter-rater reliability for T-unit and clause identification was 94% and 88%, respectively. Inter-rater reliability for error counts was 82%, a score considered adequate when coding for accuracy (Cumming et al., 2005). Intra-rater reliability for T-unit and clause analyses was 96% and 95%, respectively, whereas it was 86% for error analysis.

Quantification of citation and paraphrase practices

Another salient feature of the students' essays was how ideas and language from the given excerpts were incorporated into their own writing and how these excerpts were cited. Appropriate use of sources and correct citations are important academic skills. A large and growing body of research has shown that students do not always incorporate sources correctly (e.g. Howard, 1995; Pecorari, 2003; Shi, 2004), and perhaps of greater concern is the amount of textual borrowing, or plagiarism, that is often evident in students' assignments and dissertations (Pecorari, 2003; Shi, 2004).

Thus, analysis for use of sources included coding for the type of strategies the learners employed to incorporate ideas from the excerpts and counting the number of borrowed words. This analysis was based on Keck's (2006) typology of attempted paraphrases. Keck (2006: 263) defined attempted paraphrases as 'an instance where the writer selects a specific excerpt of a source text and makes at least one attempt to change the language of the select excerpt.' Using the criterion of the proportion of words borrowed, Keck identified four types of attempted paraphrases: Near Copies (in which 50% or more of the words are borrowed from the original), Minimal Revisions (containing 20–49% of borrowed words), Moderate Revisions (containing 1–19% of borrowed words), and Substantial Revisions (which have no borrowed words).

In this study, three types of attempted paraphrase were identified based on the proportion of words borrowed from the excerpts provided: Near Copies [NC], Moderate Revisions [MR], and Substantial Revisions [SR]. This categorisation of paraphrases seemed to provide a sharper distinction between appropriate and inappropriate paraphrases than the four types suggested by Keck (2006). Thus, Near Copies were attempted paraphrases that contained 50% or more of borrowed words, Moderate Revisions were attempted paraphrases that contained 10–49% of borrowed words, and Substantial Revisions contained less than 10% of borrowed words. Direct quotes (with quotation marks) were identified and counted separately.

An inter-rater reliability check on identifying and coding attempted paraphrase, based on five randomly chosen essays, yielded 100% inter-rater reliability on paraphrase identification. There were, however, two sources of disagreements between the raters. The first source was determining where the paraphrase began. The second source of disagreement was how to count borrowed words, given that many of the words in the attempted paraphrases (e.g. animals) that also appeared in the excerpts were commonly used words and were closely related to the given essay topic. The subsequent discussion between the researcher and fellow coder led to the establishment of clear coding and counting guidelines.

The guidelines specified that words or phrases at the beginning of attempted paraphrases that served as cohesive links between sentences (e.g. On the other hand…) or which were used to show attribution (e.g. According to…) should not be counted in the word length of the attempted paraphrases. Phrases showing attribution and source citations, such as 'Graham 2011', were coded for their presence/absence and for whether they followed citation conventions correctly. Furthermore, in determining how many words to count as borrowed words, only strings of more than two consecutive words that were exactly the same as the words in the original excerpts were counted. The following examples illustrate the different types of attempted paraphrases and the way the number of borrowed words were counted.

Original excerpt:

> *Animals are valuable, but humans are more valuable – that is where our rights come from. If you have a choice between human suffering and the life of an animal, then the answer is pretty obvious to most people. (*Hospitality Digest *September 1995).*

Example 1:

> *If we should choose between human suffering and the life of an animal, nobody will have a thinking before they choose humans, that is where our rights come from.*

The length of the attempted paraphrase in Example 1 is 29 words, of which 16 words (in bold) are borrowed from the original excerpt. Given the high proportion of borrowed words (16/29; 55%) and that the source of the paraphrase is not cited, it was coded as a Near Copy without the source of the citation [NC – citation].

Example 2:

*If we have to choose between human suffering and animal's life, of course the answer is always humans (*Hospitality Digest*, 1995).*

The length of this attempted paraphrase is 18 words (excluding source citation), of which seven words are borrowed (i.e. 39%). Thus it was coded as a Moderate Revision paraphrase accompanied by a correct citation [MR + correct citation].

Example 3:

The statement shown in Hospitality Digest *addressed that compare with animals humans are more important.*

This seven-word paraphrase (beginning with the word 'compare') contains some words that appear in the original excerpt (e.g. human, more, animal), but these words do not form a string of more than two consecutive words.

Thus, it was coded as a Substantial Revision (with no borrowed words), with an incorrect (omitted details) introductory reporting phrase and, consequently, an incorrect citation [SR + incorrect introductory phrase].

Results

The results are presented in relation to the type of analysis and measures discussed above.

Analytic essay scores

A comparison of scores on each analytic criterion and the overall score at Time 1 (T1) with the scores at Time 2 (T2) is presented in Table 2, above. A paired samples *t*-test showed that the difference between T1 and T2 scores was statistically significant for all scores. Gain scores on fluency (i.e. difference between mean at T1 and T2) were slightly higher than gains on content and form.

Qualitative analysis of structure and content

A qualitative analysis of the written texts in terms of structure revealed that most essays at Time 1 (n = 18) but fewer at Time 2 (n = 14) began with some sort of general introduction, including a brief general statement describing the controversy followed by some sort of statement, either outlining a personal stance or what the essay would argue. At Time 2, the learners sometimes omitted the general introduction and made the focus clearer. However, overall introductions at Times 1 and 2 seemed quite similar (see Excerpt 1 below), thus suggesting that the learners have internalised a type of formulaic pattern for introductions, perhaps taught to them previously in EAP preparatory courses.

The following introduction produced by Respondent 1 is fairly typical of the introductions found in the essays. At Time 1 (Excerpt 1), the introduction begins by stating the issue (Sentence 1), followed by the opposing opinions that identify the salient points of contention (Sentences 2 and 3: Many people…However, some people). Sentence 4 states the focus of this essay, which is merely a restatement of the given topic.

Excerpt 1:

Nowadays the issue about whether animals should be killed and used for animal purposes is rising in the society. Many people argued that animals have equal right to human, therefore they cannot be used for human purposes. However, some people also argued that humans are more valuable than animals, so the equal rights between animals and humans are not exist. This essay will discuss about whether animals should have equal rights to humans by exploring some areas. [Respondent 1, T1]

A similar four-sentence introduction emerges at Time 2. Respondent 1 begins their introduction with a statement of the issue (Sentence 1), and the subsequent two sentences present the controversy (Many people…However some people). The final sentence describes what this essay will discuss. Although the introductions appear to follow a similar pattern, the main difference discernible at Time 2 is that the introduction has become more focused on one particular argument (use of animals for scientific research, see Sentence 4) instead of rephrasing the given topic.

Excerpt 2:

There is an issue about animal rights rising in the society today. Many people believe that animals should not be killed, even for scientific research. However, some also believe that humans have a lot of benefit from those scientific research. This essay will discuss about several reasons on whether or not animals should be killed even for the scientific research. [Respondent 1, T2].

At Time 1, only about one-third of the essays (n = 9) had a suitable, relevant conclusion that did not merely restate the given topic. At Time 2, about two-thirds of the essays (n = 17) had appropriate conclusions. The conclusions taken from the essays written by Respondent 2 illustrate this difference. Whereas at Time 1 (Excerpt 3), the conclusion is brief (one sentence) and merely restates the main arguments, at Time 2 (Excerpt 4), the conclusion summarises the main arguments and presents a forceful concluding statement.

Excerpt 3:

In conclusion, based on the supporting points above I strongly support the animal rights to be retained. [Respondent 2, T1]

Excerpt 4:

In conclusion, animals in our daily life have played an important role in assisting human being in carrying

Table 2: Comparison of Time 1 (T1) and Time 2 (T2) analytic and overall scores (paired-sample statistics)

Criterion	Mean T1	SD	Mean T2	SD	*t*	df	Sig. (2-tailed)
Fluency	5.60	0.50	6.32	0.75	–4.27*	24	0.00
Content	5.84	0.80	6.20	0.65	–2.38*	24	0.03
Form	5.52	0.51	6.04	0.73	–3.64*	24	0.001
Overall	5.60	0.50	6.20	0.58	–4.24*	24	0.00

* Significant at $p < 0.05$

out their job, such as, as a transport, and accompany human beings, such as pets. Thus, we should make consideration to their rights, and treat them as to respect their lives. [Respondent 2, T2]

The most noticeable improvement in terms of structure was in the structure of paragraphs and coherence in the flow of ideas. The excerpts below, taken from the essays by Respondent 3, illustrate the nature of this improvement. At Time 1, Respondent 3's arguments are difficult to follow as two arguments are merged into one paragraph. In the first sentence of this paragraph, Respondent 3 notes differences between animals and humans, and what the differences could mean for animal-based research results. This is followed by a restatement of this impact, rather than the development of the idea. In Sentence 3, the discussion of the evolutionary process is not clearly linked to the main idea. The last two sentences of this paragraph describe the differences between humans and animals.

Excerpt 5:

And as we know, animals should have some difference from people, which would lead to wrong results of researches. We should always suspect the explanations of animal based researches. The evolution of human from monkey is a long and even complex period. People was separated from other animals by certain reasons. As we have higher intelligence and speaking abilities, we are given to the priority of living in the world. [Respondent 3, T1]

In contrast, at Time 2 (Excerpt 6), despite faulty cohesion and some unnecessary repetition, there is a sense of a main argument focusing on the notion of intelligence as distinguishing between man and animal. This is presented in the second sentence, developed in the third sentence, repeated in the fourth, and further elaborated in the final two sentences.

Excerpt 6:

The most important point is that human being is different from any other animals. Ignoring this gap between distinct people, they all have much higher intelligence than the most clever animal, no matter it is monkey or dog. According to this feature, human has got other special abilities, such as using languages and tools. These abilities are the most significant differences between human and animals. In another word, human has some emotions, which are the lack of animals. Emotion is also a feature of high intelligence. [Respondent 3, T2]

Improvement in content was observed mainly in terms of development of arguments, rather than in the number of arguments presented, and thus seems to be linked to the noticeable improvement in coherence and cohesion. This close link between overall coherence and how well arguments are developed is evident in Excerpts 7 and 8 below (as well as in Excerpt 6). As Excerpt 7 shows, Respondent 4 presents a number of arguments as to why animals should not be given equal rights to humans: animals are useful in medical testing (Sentences 2, 7 and 8), humans' ability to convey feelings (Sentence 3), the notion of rights (Sentence 4), pragmatic considerations (Sentence 5), and animals' inability to speak (Sentence 6). The arguments are not well developed and there is a lack of coherence and cohesion. The overall impression is that of a string of unrelated arguments.

Excerpt 7:

*Animal testing is still needed in medical fields as well as cosmetics. With animal testing, many sickness and health problems such as diabetis and asthma can be cured (*Search *Volume 25, no. 9 1994). Humans can convey feelings and expression clearly showing likes and dislikes of each individual. For example, child rights; labour rights; women rights cannot be allocated to animals. To give animals equal rights as human is not only a joke but not practical as well. An animal cannot speak for themselves or protest against any dislike or mistreat. We can always stop animal testing in medical field, yet when in extreme need, no one protect against the use of animals for research. Certainly, one suffering from asthma would not object to being cured despite the medicine given went through animal testing!* [Respondent 4, T1]

At Time 2, Respondent 4 again presents a number of arguments in support of treating animals differently, but this time each argument is elaborated and is linked to an overarching main argument (Sentence 1). As Excerpt 8 shows, there is improved coherence and cohesion as the entire paragraph is devoted to supporting the main argument against the use of animals in medical research.

Excerpt 8:

Just because human are unique in the sense that they can speak and think intelligently does not justify the actions of using animals as research subjects. Why should animals have to endure lab testing of chemicals just for the benefit of human? if we need to know the effectiveness of a drug or the side effects of it, we, humans, should be the subject of experiment, not the animals. Animal's genes are after all different from human. Besides, we should be prepared to undergo risk and take responsibility in testing the consequences of our own medicines. Years of scientific experiment on animals could still produce long term side effect on a drug declared safe for human consumption. It is ironic that human should place so much importance on themselves and declare their uniqueness, and yet still depend on animals as experiment subjects. If we are so special and no animals are equal to us, then why the use of animals for testing? Humans should be the specimens instead. [Respondent 4, T2]

Quantitative analysis of language use

Fluency

There seemed to be no difference in fluency between Time 1 and Time 2 (see Table 3) when fluency was measured quantitatively rather than qualitatively (i.e. how fluently the essay communicates ideas). The mean number of words produced at Time 1 and Time 2 was almost identical, as was the mean number and length of T-units. On both occasions, the standard deviations (SDs) were large, showing a great deal of variation among the students.

Accuracy

Descriptive results for accuracy scores at Times 1 and 2 are summarised in Table 4. The higher mean ratio of errors per

words (E/W) and lower means for proportion of error-free T-units (EFT/T) and clauses (EFC/C) suggest in fact a slight decrease in accuracy at Time 2. However, paired-samples *t*-tests showed that the differences were not statistically significant for E/W ($t(24) = -1.45$, $p = 0.16$), for EFT/T ($t(24) = 0.15$, $p = 0.88$), and nor for EFC/C ($t(24) = 0.33$, $p = 0.74$). That is, grammatical accuracy did not change from Time 1 to Time 2 (12 weeks later).

The distribution of error types also did not alter greatly over this time period. Most errors at Time 1 (56.13%) and Time 2 (52.80%) were morphological, related mainly to the use of articles, singular/plural nouns and use of verbs. Lexical errors accounted for just under 30% of errors at Times 1 and 2. Errors in syntax formed less than 20% of total errors, and there were more syntactical errors at Time 2 (18.32%) compared to Time 1 (13.04%). These errors were mainly due to incomplete sentences, typically consisting only of a subordinate clause.

Grammatical and lexical complexity

Grammatical complexity measures at Times 1 and 2 (see Table 5) were almost identical.

Results for lexical complexity or sophistication as measured by the percentage of words used in the essays that appeared on the AWL showed no change over time. At both Times 1 and 2, only 5% of the total words appeared on the AWL. The only improvement in lexical complexity seemed to be the decrease in the frequency of colloquial and informal expressions (see Table 6).

At Time 2, the learners produced fewer informal expressions, particularly first person pronouns and contractions, and fewer vague terms. Announcements such as 'I'm going to write about…' were replaced with phrases such as 'This essay will discuss…' and personal anecdotes introduced by phrases such as 'I've seen lots of…' were less prevalent in the learners' writing at Time 2.

Use of sources and borrowed words

Although the essay instructions did not require students to refer to the given excerpts (sources), most students (21 at Time 1 and 20 at Time 2) made at least one use of the sources in their essays. There were very few direct quotes used at either Time 1 (a total of two quotes found in two separate essays) or Time 2 (a total of three quotes, all in the one essay produced by a student who used only one quote at Time 1). Table 7 shows the number of attempted paraphrases, distinguishing between the three types of attempted paraphrase. Table 8 presents the relevant descriptive statistics (mean, SD and range) for paraphrase types. It also shows the mean number of words borrowed from the sources, and what proportion these borrowed words form of the total number of words in the essays.

As Table 7 shows, there was more use of sources at Time 2 than at Time 1, but the difference was not statistically significant ($t(24) = -0.85$, $p = 0.405$). Attempted paraphrases, either with Moderate Revision or Substantial Revision, were the most common types of paraphrases at both Times 1 and 2. A greater proportion of the attempted paraphrases were Near Copies at Time 2 (22/90; 24.44%) compared to Time 1 (15/78; 19.23%), but the difference in the mean number of Near Copies between Times 1 and 2 was not statistically significant ($t(24) = -0.91$, $p = 0.37$).

The proportion of borrowed words with respect to total words seemed greater at Time 2 (525/9355; 5.61%) than at Time 1 (352/9381; 3.75%), but this difference was not statistically significant ($t(24) = -1.33$, $p = 0.196$). Furthermore, the large SD and range, shown in Table 8, shows that there was considerable variation between the learners in terms of the number of words borrowed from the texts.

Most attempted paraphrases (over 60%), regardless of type, were not accompanied by a reporting phrase or a citation, at either Time 1 or 2. Furthermore, most of the citations were inaccurate because they included far too much detail (e.g. the first name initial of the author). However, at Time 2 the proportion of accurate citations (35.89%) was greater than at Time 1 (18.18%).

Table 3: Results for fluency ($n = 25$)

Criterion	Mean	SD	Minimum	Maximum
Words (W)				
Time 1	375.24	88.74	165	560
Time 2	374.28	91.98	192	575
T-units (T)				
Time 1	24.96	5.30	18	36
Time 2	25.08	4.89	13	32
W/T				
Time 1	15.10	2.34	9.17	19.26
Time 2	14.93	2.23	12.35	18.88

Table 4: Results for accuracy ($n = 25$)

Criterion	Mean	SD	Minimum	Maximum
E/W				
Time 1	0.08	0.04	0.02	0.14
Time 2	2.09	0.04	0.02	0.23
EFT/T				
Time 1	1.37	0.17	0.04	0.64
Time 2	2.37	0.19	0.05	0.76
EFC/C				
Time 1	1.51	0.17	0.19	0.78
Time 2	2.50	0.16	0.14	0.84

Table 5: Results for grammatical complexity ($n = 25$)

Criterion	Mean	SD	Minimum	Maximum
C/T				
Time 1	1.55	0.17	1.23	1.86
Time 2	1.60	0.24	1.19	2.00
DC/C				
Time 1	0.35	0.08	0.14	0.46
Time 2	0.36	0.10	0.16	0.50

Table 6: Results for lexical complexity ($n = 25$)

Colloquial expressions	Mean	SD	Minimum	Maximum
Time 1	15.80	11.24	3	52
Time 2	12.60	6.29	2	27

Discussion

A comparison of the analytic scores given by the raters on each criterion used to assess the essays at Time 1 and Time 2 indicates an improvement in writing skills achieved over one semester of in-country study. The qualitative analysis of the essays revealed that the learners' writing showed improvement in terms of structure and content development; that is, in presenting more well-developed and coherent arguments, and more appropriate conclusions. These improvements are important. As noted earlier, well-structured essays and coherent arguments are among the distinguishing traits of 'good academic writing'. They are also the traits that seem to be valued by content teachers (see Leki, 2007; Storch and Tapper, 2000).

The other discernible improvement in the learners' writing was in terms of decreased informality, findings consonant with those of Shaw and Liu's (1998) study. Although these findings may seem contrary to those of Hinkel (2003), it should be noted that Hinkel compared the writing of ESL and native English speakers rather than investigating development in ESL writing over time. The decreased informality at Time 2 found in this study may be attributable to the greater exposure to the kind of formal academic texts learners are required to read for their assignments.

Whereas the scores on the criterion of form suggested improvement over the semester, the quantitative measures of language use (fluency, accuracy, and grammatical and lexical complexity) showed no change over time. This discrepancy between an analytic score and quantitative measures of language use is consistent with findings reported by other studies (e.g. Polio, 1997). The discrepancy may arise because, as mentioned earlier, analytic scores tend to collapse a number of different features into the one criterion. For example, the criterion 'form' on the analytic scale used in this study includes reference to accuracy (e.g. 'errors are few and unobtrusive'), complexity (e.g. 'a wide variety of sentence structures is used'), and vocabulary (e.g. 'vocabulary is extensive and used appropriately').

Furthermore, there is reference to both quantity ('few', 'wide range') and quality ('unobtrusive', 'appropriate'). However, this is not to say that quantitative measures provide a more accurate reflection of learners' language use. For example, the measures of grammatical accuracy used in this study measure the quantity but not the quality of errors. That is, they do not assess the impact that different grammatical errors may have on the reader. Perhaps the raters in this study gave a higher rating on form to the essays produced at Time 2 because they contained fewer instances of informal language.

The lack of improvement in fluency (length of the writing) may be related to the recommended word length (minimum of 300) and time limits imposed. Word count can be affected by the use of more complex grammatical structures (e.g. reduced clauses) or attributive rather than predicative adjectives. Although a more detailed analysis of clauses and adjective types was not undertaken, a rough count of reduced clauses at Time 1 and 2 suggested that this was not the case in this data.

Lack of improvement in grammatical accuracy and complexity, as well as in the use of academic vocabulary, could be attributable to the relatively short time period of the study (12 weeks). For example, Ortega's (2003) meta-analysis showed that grammatical complexity (measured via scores such as C/T and DC/C) may take up to 12 months of college-level instruction to develop. It could also be argued that these learners were already fairly advanced and thus improvement for them might be harder or take longer to achieve (see Green, 2004).

The informants' use of sources also showed little change over time. Descriptive statistics suggest a more frequent use of Near Copies and an even greater reliance on borrowed words at Time 2, but the differences were not statistically significant. Nevertheless, the fact that the learners continued to rely on Near Copies to such an extent (around 20% of all paraphrases) suggests that even after studying at the university for a semester, they may still not have the linguistic skills necessary to paraphrase texts appropriately (Howard, 1995; Shi, 2004). Alternatively, perhaps their experience in writing university assignments may have in fact encouraged such practices. Leki (2007) describes the kind of assignments that students are required to complete in their university content subjects, and which seem to tacitly encourage students to copy from published reports. Learners' use of sources is clearly an area that requires further investigation given the ongoing concerns about plagiarism in academic writing.

The finding that, despite some improvement, the majority of the citations were incorrect at Time 2 was somewhat surprising given the emphasis in course handbooks on correct citations (Pecorari, 2001). Pecorari (2006) suggests that because learners do not receive feedback on incorrect citations

Table 7: Number of attempted paraphrases classified under each type

Type of paraphrase	Time 1	Time 2	Total
Near Copy	15	22	37
Moderate Revision	32	36	68
Substantial Revision	31	32	63
Total	78		

Table 8: Results for attempted paraphrases and borrowed words

Criterion	Mean	SD	Minimum	Maximum
Attempted paraphrases				
Time 1 (n = 78)	3.20	2.90	0	10
Time 2 (n = 90)	3.72	3.49	0	12
Near Copy				
Time 1 (n = 15)	0.60	1.29	0	5
Time 2 (n = 22)	0.88	1.30	0	4
Moderate Revision				
Time 1 (n = 32)	1.28	1.70	0	7
Time 2 (n = 36)	1.50	1.98	0	7
Substantial Revision				
Time 1 (n = 31)	1.24	1.39	0	5
Time 2 (n = 32)	1.50	1.46	0	5
Borrowed words				
Time 1 (n = 352)	14.16	18.61	0	68
Time 2 (n = 525)	20.88	30.99	0	133
Proportion of borrowed words of total words				
Time 1	0.04	0.05	0	0.15
Time 2	0.06	0.08	0	0.31

from their instructors, they may simply assume that their citations are correct. She recommends that learners should be given explicit feedback on inappropriate use of sources.

Thus, the findings of the present study suggest that following a semester of study in an L2-medium university, students' writing showed improvement in terms of rhetorical organisation and content development. There was also some reduction in the use of informal language. These improvements may be attributable to the language the learners were exposed to; that is, the well-structured and formal academic texts they were required to read for their hotel school assignments and the practice in writing academic assignments. Information (declarative knowledge) about how to structure an academic essay, what to include in the introduction, and how to create links between ideas is taught in EAP preparatory courses. Evidence of the pre-existence of such knowledge was clearest in the students' similar introductions in essays produced at Time 1 and Time 2. It may be that extensive practice in writing assignments may be sufficient, even after one semester, to automatise (DeKeyser, 2007a) the learners' declarative knowledge of academic essay structure.

However, a one-semester immersion experience did not lead to improved language use in terms of greater grammatical accuracy and complexity or a greater range of academic vocabulary when measured quantitatively, nor did it lead to improved use of sources. Development of more accurate and complex language use may take longer than one semester to develop (Ortega, 2003). Another plausible explanation may be related to the nature of university assessment tasks. Although it is assumed that when studying in an L2-medium university, learners are required to produce lengthy assignments, Leki's (2007) longitudinal study suggests that this may not be the case for all learners. The use of group assignments and multiple-choice examination questions means that some learners have little need or opportunity for sustained writing practice.

Even when learners are required to produce lengthy assignments, this practice in extended writing may not necessarily lead to improved language use. As research in immersion programs in Canada has shown (e.g. Swain, 1991), development in linguistic accuracy requires that learners be asked not only to produce language, but to produce accurate language. Yet, the focus of university assignments is primarily on content knowledge, as evident in the nature of the assessment tasks (Leki, 2007; Leki and Carson, 1997), grade allocations, and feedback comments (Storch and Tapper, 2000). Faculty interviewed by Leki (2007) reported that they were not unduly concerned about the L2 errors in the writing of non-native writers.

Ferris (2003) argues that the single most important element for successful development in learners' writing, and specifically the accuracy of their writing, is the provision of feedback on writing. It is such feedback and revisions in response to the feedback, rather than the act of writing (Cumming, 1990), that may 'push' (Swain, 1985, 1993) learners to express their ideas more accurately and appropriately. Studies that have investigated the impact of feedback on learners' use of language on early drafts show that such feedback can lead to improved grammatical accuracy after one semester (e.g. Polio et al., 1998) as well as improvement in lexical complexity (e.g. Storch and Tapper, 2007). Leki's (2007) longitudinal study documents the minimal feedback provided to learners on their assignments in general education classes (unless the student actively seeks it out), and the fact that it is given on final drafts, which means that it is either ignored or not useful.

This study was largely exploratory and hence its results need to be interpreted cautiously. It was small scale and was conducted over a relatively short period of time (12 weeks), and analysis of the learners' improvements, or lack thereof, was based on one text produced under test conditions. As such, the study focused only on the 'product'. It did not consider the host of personal and academic experiences of the students in this context nor the students' goals (Cumming et al., 2002). It is these experiences and goals that may help explain development or lack thereof in academic writing (see Leki, 2007). Furthermore, the findings were discussed in terms of mean scores for the entire cohort. These means (and the relatively large SDs for some measures) disguise the fact that some students did improve even in terms of language use. The other potential limitations of the study were in terms of its design, the possible practice effect of using the same topic and the use of a low stakes test. The participants may not have been equally motivated to complete the test to the best of their ability on both occasions.

Nevertheless, the findings of this study suggest that mere immersion in the L2 context and incidental learning will not necessarily lead to improved language use, as measured in this study, particularly after only one semester of immersion. We still do not know whether longer periods of immersion would lead to improvements, although the existing research does not give us reason to be hopeful either. A report published in Australia (Birrell, 2006) has revealed that some learners, even after a 2- or 3-year program, have the same linguistic proficiency on exit as on entry.

Leki's (2007) longitudinal case studies show that for some learners, it may take the duration of their entire degree program (3 to 5 years) to develop their writing abilities. However, some students on study abroad programs come only for one semester. Clearly more research is needed to document the nature of L2 development over time, as well as the kind of opportunities for output that studying in an L2-medium university provides international students. Such research will then be able to inform us what is realistically achievable in the short and long term when studying in L2-medium universities, and the kind of advice and support we should be providing incoming international students.

References

Australian Education International. 2008. International student enrolments up 20.0% in October 2008. Available at http:// aei.gov.au/AEI/MIP/Statistics/StudentenrolmentAndVisastatistics/2008/Default.htm [accessed December 2011].

Bardovi-Harlig K, Bofman T. 1989. Attainment of syntactic and morphological accuracy by advanced language learners. *Studies in Second Language Acquisition* 11: 17–34.

Birrell B. 2006. Implications of low English standards among overseas students at Australian universities. *People and Place* 14(4): 51–63.

Candlin CN, Hyland K (eds). 1999. *Writing: texts, processes and practices*. New York: Longman.

Chandler J. 2003. The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of L2 student writing. *Journal of Second Language Writing* 12: 267–296.

Christie F. 1998. Learning the literacies of primary and secondary schooling. In: Christie F, Mission R (eds), *Literacy and schooling: new directions*. London: Routledge. pp 47–73.

Coxhead A. 2000. A new academic word list. *TESOL Quarterly* 34: 213–238.

Cumming A. 1990. Metalinguistic and ideational thinking in second language composing. *Written Communication* 7: 482–511.

Cumming A, Busch M, Zhou A. 2002. Investigating learners' goals in the context of adult second language writing. In: Ransdell S, Barbier M (eds), *New directions for research in L2 writing*. Dordrecht: Kluwer. pp 189–208.

Cumming A, Kantor R, Baba K, Erdosy U, Eoanzoui K, James M. 2005. Differences in writing discourse in independent and integrated prototype tasks for next generation TOEFL. *Assessing Writing* 10: 5–43.

DeKeyser RM. 1997. Beyond explicit rule learning: automatizing second language morphosyntax. *Studies in Second Language Acquisition* 19: 195–221.

DeKeyser RM. 2007a. Introduction: situating the concept of practice. In: DeKeyser RM (ed.), *Practice in a second language.* New York: Cambridge University Press. pp 1–18.

DeKeyser RM. 2007b. Study abroad as a foreign language practice. In: DeKeyser RM (ed.), *Practice in a second language.* New York: Cambridge University Press. pp 208–226.

Duran P, Malvern D, Richards B, Chipere N. 2004. Developmental trends in lexical diversity. *Applied Linguistics* 25: 220–242.

Elder C, O'Loughlin K. 2003. Investigating the relationship between intensive EAP training and band score gains on IELTS. In: Tulloh R (ed.), *IELTS research reports,* vol. 4. Canberra: IELTS Australia. pp 207–254.

Ellis R. 2003. *Task-based language learning and teaching*. Oxford: Oxford University Press.

Ferris D. 2003. *Response to student writing. Implications for second language students*. Mahwah, New Jersey: Lawrence Erlbaum.

Freed B. 1995. Language learning and study abroad. In: Freed B (ed.), *Second language acquisition in a study abroad context.* Amsterdam: John Benjamins. pp 3–34.

Gass S. 2003. Input and interaction. In: Doughty C, Long M (eds), *Handbook of second language acquisition*. Oxford: Blackwell. pp 224–255.

Green T. 2004. Making the grade: score gains on the IELTS Writing test. *Research Notes* 16: 9–13.

Green A, Weir C. 2003. Monitoring score gain on the IELTS academic writing module in EAP programmes of varying duration. Phase 2 report. Cambridge: UCLES.

Hinkel E. 2003. Simplicity without elegance: features of sentences in L1 and L2 academic texts. *TESOL Quarterly* 37: 275–301.

Hirsch D. 2007. English language, academic support and academic outcomes: a discussion paper. *University of Sydney Papers in TESOL* 2: 193–211.

Howard R. 1995. Plagiarisms, authorships, and the academic death penalty. *College English* 57: 788–805.

Hunt K. 1966. Recent measures in syntactic development. *Elementary English* 43: 732–739.

Hyland K. 2003. Genre-based pedagogies: a social response to process. *Journal of Second Language Writing* 12: 17–29.

Keck C. 2006. The use of paraphrase in summary writing: a comparison of L1 and L2 writers. *Journal of Second Language Acquisition* 15: 261–278.

Krashen SD. 1985. *The input hypothesis: issues and implications.* London: Longman.

Leki I. 2007. *Undergraduates in a second language: challenges and complexities of academic literacy development*. New York: Lawrence Erlbaum.

Leki I, Carson J. 1997. Completely different worlds: EAP and the writing experiences of ESL students in university courses. *TESOL Quarterly* 31: 39–69.

McCarthy PM, Jarvis S. 2007. Vocd: a theoretical and empirical evaluation. *Language Testing* 24: 459–488.

Melles G, Millar G, Morton J, Fegan S. 2005. Credit-based discipline specific English for academic purposes programmes in higher education. *Arts and Humanities in Higher Education* 4: 283–303.

Muranoi H. 2007. Output practice in the L2 classroom. In: DeKeyser RM (ed.), *Practice in a second language.* Cambridge: Cambridge University Press. pp 51–84.

Ortega L. 2003. Syntactic complexity measures and their relationship to L2 proficiency: a research synthesis of college-level L2 writing. *Applied Linguistics* 24: 492–518.

Pecorari D. 2001. Plagiarism and international students: how the English-speaking university responds. In: Belcher D, Hirvela A (eds), *Linking literacies: perspectives on L2 reading-writing connections.* Ann Arbor: University of Michigan Press. pp 229–245.

Pecorari D. 2003. Good and original writing: plagiarism and patchwriting in academic second-language writing. *Journal of Second Language Writing* 12: 317–345.

Pecorari D. 2006. Visible and occluded citation features in postgraduate second-language writing. *English for Specific Purposes* 25: 4–29.

Polio C. 1997. Measures of linguistic accuracy in second language writing research. *Language Learning* 47: 101–143.

Polio C, Fleck C, Leder N. 1998. If I only had more time: ESL learners' changes in linguistic accuracy on essay revisions. *Journal of Second Language Writing* 7: 43–68.

Read J, Hayes B. 2003. The impact of the IELTS test on preparation for academic study in New Zealand. In: Tulloh R (ed.), *IELTS research reports*, vol. 4. Canberra: IELTS Australia. pp 237–262.

Richards J, Platt J, Platt H. 1992. *Dictionary of language teaching and applied linguistics*. Harlow: Longman.

Shaw P, Liu ET-K. 1998. What develops in the development of second language writing. *Applied Linguistics* 19: 225–254.

Shi L. 2004. Textual borrowing in second-language writing. *Written Communication* 21: 171–200.

Silva T, Brice C. 2004. Research in teaching writing. *Annual Review of Applied Linguistics* 24: 70–106.

Storch N, Hill K. 2008. What happens to international students' English after one semester of university? *Australian Review of Applied Linguistics* 31(1): 04.1–04.17.

Storch N, Tapper J. 2000. Discipline specific academic writing: what content teachers comment on. *Higher Education Research and Development* 19: 337–356.

Storch N, Tapper J. 2007. Assessing postgraduate ESL learners' writing after one semester of study. Paper presented at the Language, Education and Diversity Conference, Hamilton, New Zealand, 21–24 November 2007.

Swain M. 1985. Communicative competence: some roles of comprehensible input and comprehensible output in its development. In: Gass SM, Madden CG (eds), *Input in second language acquisition.* Rowley: Newbury House. pp 235–256.

Swain M. 1991. Manipulating and complementing content teaching to maximise second language learning. In: Phillipson R, Kellerman E, Selinker L, Sharwood-Smith M, Swain M (eds), *Foreign/second language pedagogy research: a commemorative volume for Claus Faerch*. Clevedon: Multilingual Matters. pp 145–163.

Swain M. 1993. The output hypothesis: just speaking and writing aren't enough. *Canadian Modern Language Review* 52: 158–164.

Swain M, Lapkin S. 1995. Problems in output and the cognitive processes they generate: a step towards second language learning. *Applied Linguistics* 16: 371–391.

University of Melbourne. 2006. Healthy report card for international students. Uninews media release, Thursday 27 April, 2006. Available at http://uninews.unimelb.edu.au/view.php?articleID=3351 [accessed December 2011].

Weigle SC. 2002. *Assessing writing.* Cambridge: Cambridge University Press.

Wolfe-Quintero K, Inagaki S, Kim H. 1998. *Second language development in writing: measures of fluency, accuracy and complexity*. Honolulu: University of Hawaii at Manoa.