



Defining a Unified Model of Vocabulary Acquisition via Extensive Reading

Final report 2016

RMIT University

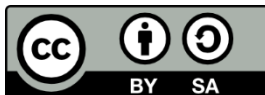
Alexandra Louise Uitdenbogerd

Sara Kablaoui

Aidan Martin (Research Assistant)

<http://titan.csit.rmit.edu.au/~e06876/olt2014/>

Support for the production of this report has been provided by the Australian Government Office for Learning and Teaching. The views expressed in this report do not necessarily reflect the views of the Australian Government Office for Learning and Teaching.



With the exception of the Commonwealth Coat of Arms, and where otherwise noted, all material presented in this document is provided under Creative Commons Attribution-ShareAlike 4.0 International License <http://creativecommons.org/licenses/by-sa/4.0/>.

The details of the relevant licence conditions are available on the Creative Commons website (accessible using the links provided) as is the full legal code for the Creative Commons Attribution-ShareAlike 4.0 International License <http://creativecommons.org/licenses/by-sa/4.0/legalcode>.

Requests and inquiries concerning these rights should be addressed to:

Learning and Teaching Support
Student Information and Learning Branch
Higher Education Group
Department of Education and Training

GPO Box 9880,
Location code C50MT7
Canberra ACT 2601

<learningandteaching@education.gov.au>

2017

ISBN 978-1-76051-060-2 [PRINT]
ISBN 978-1-76051-061- 9 [PDF]
ISBN 978-1-76051-062-6 [DOCX]

Acknowledgements

We thank RMIT English Worldwide Melbourne for providing access to electronic copies of the textbooks, a space for running experiments, allowing advertising of the study to students during class time, and generally supporting the study.

Thanks also to Patricia McLaughlin for advice at critical stages of the project.

List of acronyms used

EAP - English for Academic Purposes

ELICOS - English Language Intensive Courses for Overseas Students

ICCE - International Conference on Computers in Education

L1 - First language

L2 - Second language

NLP-TEA - Natural Language Processing Techniques for Educational Applications

MC –Multiple choice

REW - RMIT English Worldwide

SHEER - Science, Health and Engineering Educational Research

TELL - Technology Enhanced Language Learning

YN – Yes no

Executive summary

Project context

Tertiary students with a non-English-speaking background frequently struggle with tasks such as reading their textbooks, due to their substantially smaller operating vocabulary, as well as their somewhat more limited grammar competence compared to native speakers. Achieving the vocabulary knowledge required for comfortably reading textbooks isn't feasible within the timeframe of typical English programmes. Past research has shown that extensive reading, or reading large quantities of not-too-difficult text for pleasure improves language skill, including vocabulary. However, it is not certain how much reading is required to achieve a particular vocabulary acquisition goal.

Aim of the project

In this project, the aim was to investigate the relationship between various factors related to reading and the amount of vocabulary gained. The project goal was to be able to advise stakeholders about the feasibility of using extensive reading for vocabulary improvement, and how this can be achieved by students entering or in the Australian tertiary education system.

Project approach (in brief)

We completed a study with students studying English for Academic Purposes at RMIT English Worldwide Melbourne campus, in which students sat vocabulary tests, completed a participant questionnaire, and in which some students borrowed Oxford Bookworms easy readers and reported their reading.

Project outputs/deliverables/resources

The main outputs of the study are:

- a vocabulary test that uses minimal vocabulary knowledge to express the questions, and tests words that are likely to be encountered in English stories, and unlikely to occur often in academic English curriculum (to allow measurement of vocabulary gain likely to come from extra-curricular reading)
- analysis of a pilot extensive reading programme, in terms of participant reading habits, vocabulary size and vocabulary change
- models showing the relationship between vocabulary knowledge and reading
- estimates of effect size and required sample size for future studies of factors related to vocabulary and reading
- recommendations for how to organise an extensive reading programme as an extracurricular activity for students of English for Academic Purposes (EAP).
- several presentations of the project to different stakeholders:

- RMIT Science, Health and Engineering Educational Research (SHEER) Centre members meeting
- Researchers interested in Natural Language Processing Techniques for Educational Applications (NLP-TEA), including Technology Enhanced Language Learning (TELL) at the International Conference on Computers in Education (ICCE).
- Natural Language Processing research group at the University of Melbourne
- Poster at the Workshop of the Australian Language Technology Association, collocated with the Australian Linguistics Society at Confluence 2015 at University of Western Sydney, Parramatta
- RMIT English Worldwide Melbourne's Professional Development day for teaching staff
- North American Chapter of the Association for Computational Linguistics.

Impact of the project

- Participants in the project have seen the value of extensive reading in improving their English language skill, with some continuing their involvement well beyond the original study.
- Project talks have led some attendees to use graded readers for their own foreign language skill development.
- Papers are being submitted to journals and conferences related to language learning. These are yet to be accepted.

Key findings or recommendations

- Via the participant questionnaire and vocabulary test a relationship between reading habits and vocabulary knowledge was discovered. Those who regularly read in English performed better in the vocabulary test than those who didn't, and this relationship was stronger than that of the enrolled English course level and vocabulary knowledge (for vocabulary with a bias toward stories).
- When comparing reading participants with control group participants, there was a slight but apparently significant increase in the vocabulary test results of participants. However, it should be noted that, like many education-related studies, the study used a convenience sample of volunteers, making the results suggestive but not statistically valid.
- Another goal was to confirm the predicted model (based on an informal meta-study of 16 published studies) of an approximately linear relationship between quantity read, weighted by text difficulty, and vocabulary gain. The number of participants in the reading group that remained in the study was smaller than hoped, limiting the ability to show a relationship between the factors.

- The effect size determined from the study leads to the conclusion that 80 participants are required for a full study that can provide statistically significant conclusions. As drop-out of volunteers was about 65 per cent, it is recommended to recruit 120 participants at the start.
- Participation in an extracurricular reading programme with weekly meetings, an attractive set of books to borrow, and the opportunity to converse with other participants and the researcher, was greatly valued by a subset of participants, and led to greater continued participation than when extensive reading was merely recommended via a single introductory session.
- ***Based on the study, it is recommended that EAP providers introduce an extra-curricular extensive reading programme using the full range of available graded readers. This can be combined with existing conversation groups, to provide both a reading incentive and a topic of conversation. At later stages of reading skill, it is recommended that students select reading material of a more academic nature, to better prepare for the type of text to be tackled in their tertiary studies.***

While the project has been finalised, team members continue to explore the topic, present the results to different groups, and write publications.

Table of contents

Contents

Acknowledgements.....	2
List of acronyms used	3
Executive summary.....	4
Project context.....	4
Aim of the project	4
Project approach (in brief).....	4
Project outputs/deliverables/resources.....	4
Impact of the project	5
Key findings or recommendations.....	5
Table of contents	7
Contents.....	7
Tables and figures	9
Tables	9
Figures.....	9
Chapter 1: Introduction	10
Project Approach	10
Value/Need of the Project	12
Chapter 2: Models from a Meta-Study.....	13
Chapter 3: Factors Affecting Vocabulary Scores of EAP Students.....	16
Action Research Activity	16
Aims.....	16
Method	16
Sessions.....	17
Participant Questionnaire.....	17
Text.....	17

Test.....	18
Test Component.....	19
Number of Questions.....	19
Explanation	19
Results.....	19
Participants	19
Test Results	20
Discussion.....	23
Chapter 4: Change in Vocabulary after an Extensive Reading Programme	25
Chapter 5: Long-Term Participants	28
Chapter 6: Conclusions and Outputs	30
Findings	30
Interdisciplinary Linkages.....	30
Factors Affecting the Success of the Project	30
Dissemination and Project Impact.....	30
Implementing an Extensive Reading Programme at an ELICOS Provider.....	31
Links to Other OLT Projects	31
Future Work.....	31
References	32
Appendix A.....	33
Appendix B: Workshop Presentations	34
Appendix C: Impact Plan	35
Appendix D: Participant Questionnaire	37

Tables and figures

Tables

Table 1: Vocabulary Test Components19
Table 2: Least and most known words of the vocabulary test.....20

Figures

Figure 1: Reported vocabulary size for number of hours of instruction.....14
Figure 2: Reading habits Likert response vs MC score.....21
Figure 3: Reading habits adjusted to represent magnitude vs MC score.....22
Figure 4: Course level versus MC test score22
Figure 5: MC score versus YN score23
Figure 6: MC test scores before and after the 5 week programme for each group25
Figure 7: Changes in Yes/No word knowledge scores for each group. Error bars show standard error using t-distribution values due to the small number of participants.26
Figure 8: Difficulty and Change in Vocabulary Familiarity. The easiest difficulty is 1 and the most difficult is 5.....27

Chapter 1: Introduction

The International English Language Testing System (IELTS) that is used for assessing English language skills for entry into academic institutions, tests each of the four communication skills: reading, writing, speaking and listening. Vocabulary knowledge is an integral part of each of these skills. However, the vocabulary required for academic reading is quite large and time spent in English language courses is limited. The use of extensive reading as an adjunct to language courses or even integrated within a course can improve vocabulary as well as other language skills (Horst, 2005; Waring & Takaki 2003). This project involved trialling the use of extensive reading with RMIT English Worldwide (REW) students and testing the amount of vocabulary gained by participants beyond that expected from their normal coursework.

The project results:

- can inform both students and language teachers of the vocabulary improvement to expect from extensive reading, so that informed decisions can be made regarding its use, given the language goals of students with respect to their further education.
- can be used by REW and other providers of English Language courses to better prepare students for further study.

Project Approach

This project initially involved refining a model based on a meta-study of prior research into vocabulary acquisition via extensive reading. Preliminary work has determined that the main variables that affect the quantity of vocabulary acquired are the amount of reading completed and the difficulty of the text for the learner. Text difficulty has two major components: vocabulary and grammar (Uitdenbogerd, 2005). The type of vocabulary test used also affects the reported results. For example, where a learner needs to define a tested word their success is less than if they merely need to identify the correct definition amongst several choices (Waring & Takaki, 2003).

The model was to be verified by recruiting many REW students to participate in a study in which they read books for enjoyment over a period of several weeks.

The frequency of exposure to a word while reading is known to be another important factor in determining how likely it is to be learnt (Horst, Cobb & Meara, 1998; Webb, 2007). This is to be expected as it is consistent with the well-established principle of learning, repetition. It is the nature of word frequency distributions that there are an enormous number of words that occur only once in any given text, and a small number of words occur very frequently. For example, in the story *Alice in Wonderland*, there are 26,505 words, consisting of 2,651 distinct words, of which 1,176 occur exactly once. The most frequent word is “the”, which occurs 1,631 times (about 6 per cent). The word frequency distribution of ordinary text is sometimes described as a “power law” distribution. The study selected words of varying

frequency for testing, in order to more precisely model the pattern of vocabulary gains to be expected.

Horst reports many flaws in prior experiments into vocabulary acquisition via extensive reading (Horst, 2005), including not testing a sufficient number of words to demonstrate learning, using texts of an inappropriate level, such as unsimplified texts, and insufficient text read by participants. There is also a question of how much vocabulary acquisition over the course of a study can be attributed to the extensive reading, and how much to other language exposure. The main methods of measuring the contribution from extensive reading are:

1. selecting vocabulary that is virtually guaranteed to be unfamiliar to participants,
2. substituting real words with invented words in special versions of the texts to be read, and
3. measuring the difference in vocabulary gains between control groups and those who engage in extensive reading.

Option 1 would be very difficult to control, however, vocabulary test words can be selected based on a pre-test of likely candidates that rarely occur in the English language course material. The use of invented words (option 2), while a very good methodology, generally limits the amount of text available to participants for reading. It also limits the benefits to the participants, as they are deprived of the opportunity of genuine vocabulary growth. The third option was chosen, with careful selection of test vocabulary based on pre-tests and knowledge of the vocabulary of course materials.

Participants were placed into two groups: a reading group and a group that did not read extra books as part of their course. Both groups were tested before and after the study to determine their vocabulary knowledge. Both groups kept a reading diary in which they recorded any English text read during the weeks of the experiment. Any gains by the reading group above that of the control group would be from the reading. A third control group only took the pre- and post- tests. This was to measure potential changes in behaviour of participants who keep a reading diary, and any resulting changes of vocabulary acquisition. Coursework test scores were to be used to measure general language improvements.

Participants were expected to read different amounts during the course of the study, depending on their skill level and their enthusiasm. The target reading quantity was in the range 10,000-60,000 words, being the equivalent of approximately two graded readers at the lowest levels to two at the highest levels in the Oxford Bookworms series. This range would allow us to examine more closely under a controlled environment the effect of text quantity on vocabulary acquired.

Participants are likely to differ in their initial vocabulary knowledge. This is likely to mean that the reading material would be more difficult for some people than others, allowing us to test the effect of text difficulty on acquired vocabulary.

Value/Need of the Project

This project addresses one of the fundamental issues of students with a non-English speaking background in Australian education institutions. Once they have completed their English studies, students still struggle with reading textbooks due to their substantially smaller operating vocabulary, as well as their somewhat more limited grammar competence compared to native speakers. For example, a 20-year-old native English speaker can be expected to know about 20,000 word families (Nation & Waring, 1997), whereas a person with English as a second language will operate on a vocabulary of about 4,000 after more than 2,000 hours of instruction (Laufer, 2000).

We expect that some gains may be possible while students are studying English through their experience with extensive reading during their course, but that the main benefits will extend beyond their course, as they continue reading books in English for pleasure.

While it is known that extensive reading has many benefits for language learners, this study should enable students and staff to better predict the vocabulary gains to be expected from reading books of a specific level of difficulty at a specific rate, so that appropriate reading targets can be set. Behaviours associated with reading in foreign languages, such as guessing and/or looking up definitions of unfamiliar words, can be strategically chosen, while avoiding techniques more associated with “intensive” reading, in which text is studied rather than read fluently (Kablaoui & Khabbaz, 2012). In addition to the knowledge from the study that can be disseminated, the participants discovered that they can read for pleasure in English via graded readers of a level that is comfortable for them, and thus consolidate and improve their English language skills by reading more text than they may have otherwise attempted.

While this study focuses on the use of extensive reading for English language learners, since English is the language that is most needed for international students, the same principles can be applied to other languages. However, the precise vocabulary gains and vocabulary requirements may differ (Nation & Waring, 1997).

Chapter 2: Models from a Meta-Study

Past studies into vocabulary gain from extensive reading have provided little opportunity for a predictive model to be created. However, when the studies are examined as a whole, a relationship becomes visible. This chapter summarises the linear models¹ and discusses the implications for language learners.

A simple model based on text size in words (x), and the ratio of the change in test score divided by words tested (y), had an R-squared value² of 0.458.

$$\text{Equation (1)} \quad y = 3 \times 10^{-6}x + 0.193$$

This suggests that the proportion of initially unknown words that are learnt through reading increases with the amount of text read. The trend is probably due to the repetition likely to occur as more words are read.

A second linear model used logarithms for the text size and achieved a similar but slightly higher R-squared result (0.464).

$$\text{Equation (2)} \quad y = 0.2036 \log x - 0.5368$$

However, text length is not necessarily the only large factor in determining vocabulary gain. Additional models include estimates of the participants' vocabulary coverage of the text. Equation 3, using coverage only resulted in an R-squared of 0.667.

$$\text{Equation (3)} \quad y = 0.0578x - 5.0897$$

This demonstrates that it is easier to pick up unknown words when the bulk of the text is well understood.

Combining coverage and text length into a single model gave an R-squared value of 0.63.

$$\text{Equation (4)} \quad y = 3.19 \times (v/100 x + k \times l/200,000) - 2.7$$

where v is the coverage percentage, k is a constant set to 0.1 and l is the text length in words.

The fifth model uses the log of the text length instead of the raw length. In this case the log length is scaled by 6 and k is set to 2. The resulting R-squared value was larger again.

These models all suggest that **higher coverage and higher quantities of reading lead to the highest vocabulary gains.**

¹ The list of papers from which source data were acquired is located on the website: <http://titan.csit.rmit.edu.au/~e06876/olt2014/>

² Coefficient of determination. A value of 1 means that y is completely explained by x , and 0 represents no relationship between the two variables.

Another model was built based on vocabulary gain from coursework, using the data provided by Laufer (2000) (See Figure 2). The resulting equation with R-squared of 0.8392 is shown below:

$$\text{Equation (5) } y = 1.7x + 341.4$$

where x is the number of hours of instruction and y the resulting vocabulary size. This very roughly shows that for the normal range of instruction times, the vocabulary size is about twice the number of hours of instruction. For an additional 50 hours, the vocabulary increases by 85 words.

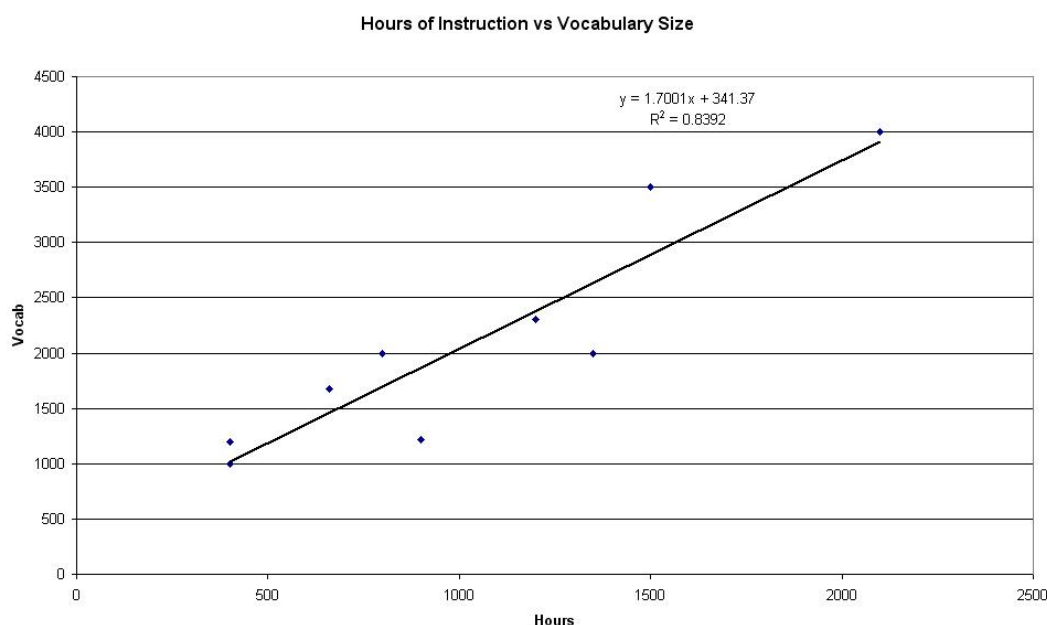


Figure 1: Reported vocabulary size for number of hours of instruction

There is at least one study that states that an extensive reading programme led to greater language gains than a more traditional class. Language learner reading speeds range from 50 to 170 words per minute, so conservatively, in 50 hours a learner could read 150,000 words. The vocabulary gains from this reading are not easy to quantify precisely, as word frequency distributions are not stable with text size or text type. However, the models suggest a gain of about 50 per cent of words initially unknown to the reader - assuming the statistical properties of the words are similar to those selected for testing in the studies used to create these models, and that the reading level is not too challenging.

As an example, the first 20 Bookworms books in this study had a combined total number of words of 70,630 and a vocabulary of 2,510 words (where names, plurals and all verb forms are included in the count). With 95 per cent coverage, it would be expected that the student knows at least 1,086 of the vocabulary words occurring in their reading (the most frequently occurring). The remaining 1,424 items of vocabulary are initially unknown, but it can be expected that half of these are learnt during reading, making a total of 712 vocabulary

items. The reading would take less than 25 hours. This is, of course, based on many assumptions, and the standard deviation of the vocabulary gain is very large. It also doesn't take into consideration that vocabulary is also often forgotten when it is not encountered for a period of time, and that some words are more memorable than others. The number is considerably larger than the likely vocabulary gain, and this is probably because the distribution of test words is less skewed than vocabulary of a large text. It is an ongoing research question how to validly compare word frequency distributions.

Despite the inability to accurately estimate the number of vocabulary words likely to be learnt when reading a text, the models do demonstrate that extensive reading can be a useful method for students to continue their vocabulary growth both during and after they have completed their EAP studies.

Chapter 3: Factors Affecting Vocabulary Scores of EAP Students

Action Research Activity

Aims

To refine and verify a model of the amount of vocabulary gained from extensive reading, given the quantity, difficulty level, interest level of the text and the linguistic background of the student. The *a-priori* hypotheses are:

1. Extensive reading contributes to vocabulary knowledge
2. The more a learner reads, the greater the vocabulary gain
3. There is more vocabulary gain within a set period of time from reading easier texts than more difficult ones (assuming there is at least *some* vocabulary that is unfamiliar in the easy texts).
4. There is more vocabulary gain from interesting texts than less interesting ones.

Method

To meet with ethics requirements and constraints of coursework, participants were invited to take part in the study, which would run outside of normal class hours. As this would mean fewer participants than if entire classes could be studied within their normal courses, recruitment was from successive intakes of classes. It should be noted that the nature of the recruitment means that it is a convenience sample, rather than a random sample of the population of students learning Academic English at REW, in Melbourne, Australia.

Participants are those that are probably more likely to choose additional English study beyond the basic coursework.

Participants were recruited from levels 3-6 of the REW academic English programme. One of the research team members visited classes to announce and encourage participation in the study. Interested students came to the first session, and most of these joined the study.

To match with REW's course structure, a five-week extra-curricular reading programme was chosen. Students completed a participant questionnaire, sat vocabulary tests, borrowed books and reported their reading. Two additional groups of students were treated as controls. After five weeks, participants were invited to sit a second vocabulary test, so that change in vocabulary could be measured.

The resulting data were analysed for relationships between the factors mentioned above. The assumption in this study is that participants will have similar exposure to the English language during their course. They all use the same textbook series, and would have similar levels of extra reading requirements. It is acknowledged that the English exposure is not likely to be completely the same. In particular, at the higher course levels, students are

required to research topics and write about them, which tends to lead to different resources being read.

Sessions

Participants who attended the first session were:

1. given an information and consent form in their native language
2. asked to fill out the participant questionnaire on-line
3. asked to complete the vocabulary test

This completed the activity of control group participants. Participants who attended sessions for the diary study or intervention, were given a short presentation (about 5 minutes), about extensive reading, shown the form that they needed to fill out when they had done some reading in English beyond their coursework, and shown an “easy reader”.

Participants in the intervention group were then invited to borrow a book from the Oxford Bookworms series, kept in a special box apart from the library collection. These participants were also asked to attend every week for about half an hour to return books, borrow new books, report some reading, and spend a few minutes reading.

All participants were asked to complete a second vocabulary test five weeks after the first one.

Participant Questionnaire

Participants were presented with a questionnaire (see Appendix D: Participant Questionnaire) that captured general demographics, as well as languages known, reading habits in their native language, and reading habits in English. It was designed to capture likely confounding variables, in addition to providing an opportunity to compare long-term English reading habits to vocabulary knowledge.

Text

Source Text

RMIT English Worldwide has its own set of textbooks, entitled ‘Passport’. They are published for levels 2 to 7 of the English course. These textbooks were used to model the language exposure and knowledge of the participants.

The selection of story books

Due to the large body of extensive reading work previously published using the Oxford Bookworms stories, these were selected as the reading material for the study.

A “starter pack” of 10 stories was acquired for the starter stage and stage 2. Five stories with high ratings on the Goodreads website were acquired for stage 1. In addition, the

starter stage story Star Reporter was purchased due to its high rating on Goodreads. This gave a set of 26 unique stories ranging from starter to stage 2.

As some of the participants found the initial set of books too easy, extra books were added later on from stages 3 to 6. These were again chosen based on Goodreads ratings (at least 3.5 average rating over at least 10 ratings).

Processing of Source Text

Story books were scanned in colour at 300 dpi with a Ricoh Aficio MP C3002. Scan Tailor was used to remove headers and footers from scanned pages. (Headers and footers consisted of page numbers and chapter and volume titles). Scan Tailor was also used to de-skew the scanned text. OCR was performed on the de-skewed, colour scans with tesseract version 3.03 using default settings. Output from tesseract was visually inspected for errors and manually corrected as required. The Passport textbooks were available in pdf format. Text was extracted from pdf using pdftotxt.

Raw word counts were generated for each story book and textbook. Portions of text unlikely to be read by subjects were not included when generating raw word counts. For the Passport textbooks, the licensing agreement and acknowledgements text was discarded. For the story books, copyright information (front of book) and comprehension exercises (rear of books) were discarded. Prior to generating word counts, the following text pre-processing was performed: 1) Unicode characters were decomposed to their individual characters and combining characters, and the combining characters were dropped. 2) Punctuation was removed. 3) Uppercase characters were transformed to lowercase. Individual words were identified by splitting text on whitespace. Words containing digits were discarded.

Test

The vocabulary test was developed on the principle that words that occurred frequently in stories but rarely in the textbook would best reflect vocabulary gain from reading. To ensure the test questions were understood, vocabulary for the answers and distractors was restricted to that occurring in the lowest level of text book. The initial test was shorter than the final one due to additional books and more difficult words being added later. Both versions can be found on the website. The test questions that were common to all participants were constructed in the following way:

1. Raw word counts were calculated for each reading book and for the entire set of student textbooks (levels 2-7), with different word forms counted as separate words.
2. The set of words that occurred at least 5 times in the textbooks were subtracted from the sets of words occurring at least 5 times in each story book.
3. Words that occurred at least 5 times in at least 3 story books (out of books 1-20) and less than 5 times in the textbooks were included as test words: dead, smiled, stood, quietly, replied.

All except “quietly” were used in a match word to meaning question. “Quietly” was excluded from this type of question due to its part of speech, and was tested via a multiple choice question.

4. The most frequently occurring term in a story book, where the meaning is apparent in the story, and with fewer than 5 occurrences in the textbook, was included. One word was chosen per book, but culturally insensitive words were excluded (“heroin”). These words were tested via multiple choice.
5. Words that occurred at least 5 times in at least 2 story books and less than 5 times in the textbooks were included as checklist words. The following words were excluded due to cultural sensitivity (“whisky”), and where the meaning was not clear (“papers”).
6. The second most frequently occurring term in one book, excluding terms already listed, terms where the meaning is not clear, or culturally insensitive terms, was included in the checklist.
7. Where there were fewer than 5 test words for a story, additional words were added to the checklist.
8. For ‘Star Reporter’, it was necessary to subtract level 2 textbook words only (instead of levels 2 to 7) for there to be any test words at all. ‘Sally’s Phone’ also required using a lower level to find a second test word. All other stories had some suitable words to choose from.
9. Multiple choice questions were written, providing distractors and answers that used the vocabulary of the level 2 textbook as much as possible. The most challenging answer to create within this constraint was “Someone who ends another person’s life”, as “ends” had only been encountered as a verb in the sense “ends with”.

Table 1 summarises the test sections.

Table 1: Vocabulary Test Components

Test Component	Number of Questions	Explanation
MM	4	4 common test words
MC.1	20	Words from initial test list that remained MC questions throughout the experiment
YN.1	37	First set of checklist words created as described above, excluding words that later became MC words.

Results

Participants

There were 63 participants in total, of which 32 were Chinese, 16 Arabic speakers, and no more than 4 from 10 other specified languages. There were 36 males, 25 females, and 2 unspecified. The most common age range was 21-25 (24), with only one participant being over 40. Five participants reported knowing two languages other than their native language and English. A further six reported knowing one other language. The other languages

mentioned were Japanese (5), Cantonese (2), Korean (2), French (2), Hindi (1), Dutch (1), Portuguese (1), Nepali (1).

Thirty four participants had completed a bachelor degree, 15 had completed high school, and six a master’s level qualification. The majority of participants were in levels 3-5 of the EAP course, with level 5 having the most participants (27).

For those who had studied beyond high school, the areas of study were fairly evenly distributed between business (10), arts (11), engineering (9) and medicine (7), with a smaller number in science (3) and other (4) fields.

Twelve participants stated that they don’t read books in English, 35 that they read less than one book per year, and 11 read 1-5 books per year, with four reading more than 5 books (and one unspecified). On average participants read one category higher in their native language than in English. However, there was a full spread of native reading habits for each of the lowest categories of English reading habits. The one person who read more than 20 books in English per year, also did so in their native language.

Test Results

The following test analysis is for the test questions that remained the same for all participants. The number of participants completing the additional questions is substantially lower than 63.

Test Difficulty

The median test score was 15/20 for the multiple choice component (minimum 6, lower quartile 13, median 15, upper quartile 17, maximum 20). Despite this, there was only one word (“knife”) that every participant got correct in their first test.

The initial “mix and match” question was answered correctly by almost all participants (51), but was misunderstood by a few (7).

Table 2: Least and most known words of the vocabulary test

Most Known Words	Known by	Least Known Words	Known by
knife	63	attic	16
skirt	61	sobbed	26
killer	61	astronauts	30
truck	58	sheriff	33
soldier	58	costume	36
wolf	56	guard	36

Reading Habits and Test Score

We found a weak positive relationship between the reported level of reading and the multiple choice score. This analysis ignored the slightly non-linear scale used for the reported number of books read per year. A post-hoc power analysis showed that, if the underlying statistical assumptions were true, then the effect size is large enough to have significance for the number of participants (assuming Power = 0.8, and alpha = 0.05).

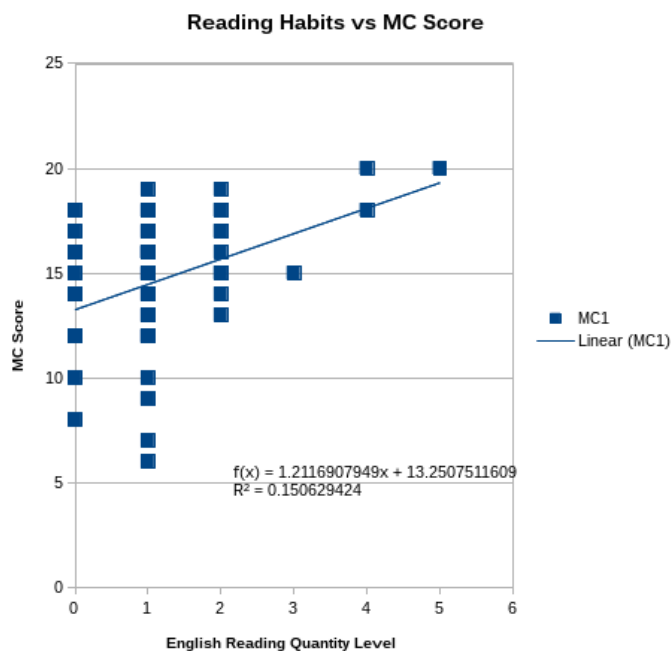


Figure 2: Reading habits Likert response vs MC score

When the Likert values were replaced with midpoint values of ranges (and 25 for the final “more” category), a clearly non-linear relationship is shown. Note that axes have been inverted to allow a polynomial to be fitted to the data.

As can be seen in the graph, the only students to get 100 per cent on the MC test were those who read at least 11 books (graphed as 15.5) per year. While it was possible for one participant to achieve 18/20 on the test with a reported English reading level of 0, the lowest score associated with reading at least one book per year was 13/20. For reading at least 6 books, a score of 15/20 was the minimum found.

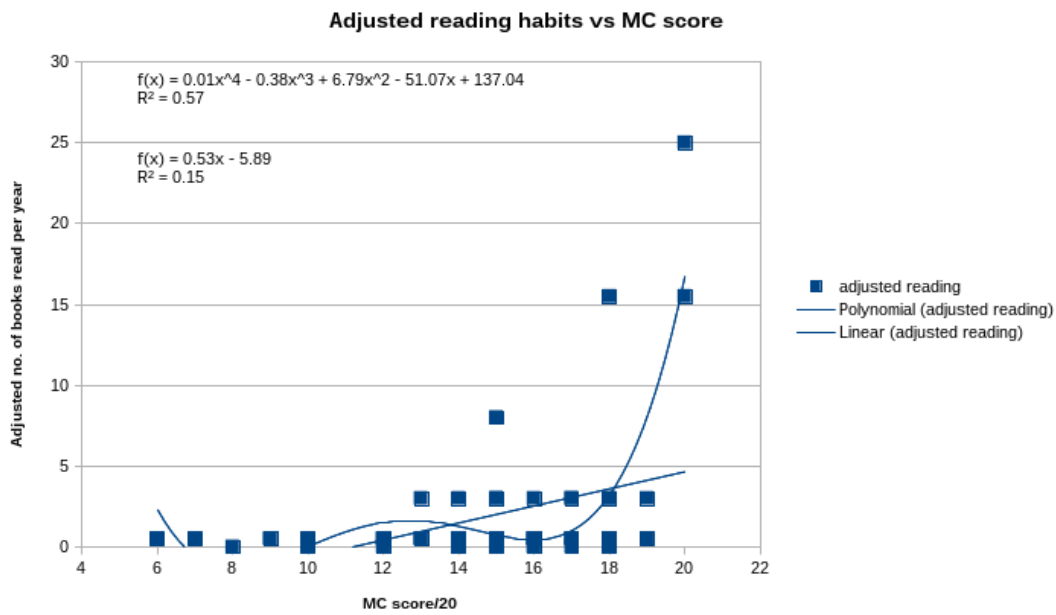


Figure 3: Reading habits adjusted to represent magnitude vs MC score

Course Level and Test Score

Course level had a slightly lower correlation with the test score than was found for reading habits.

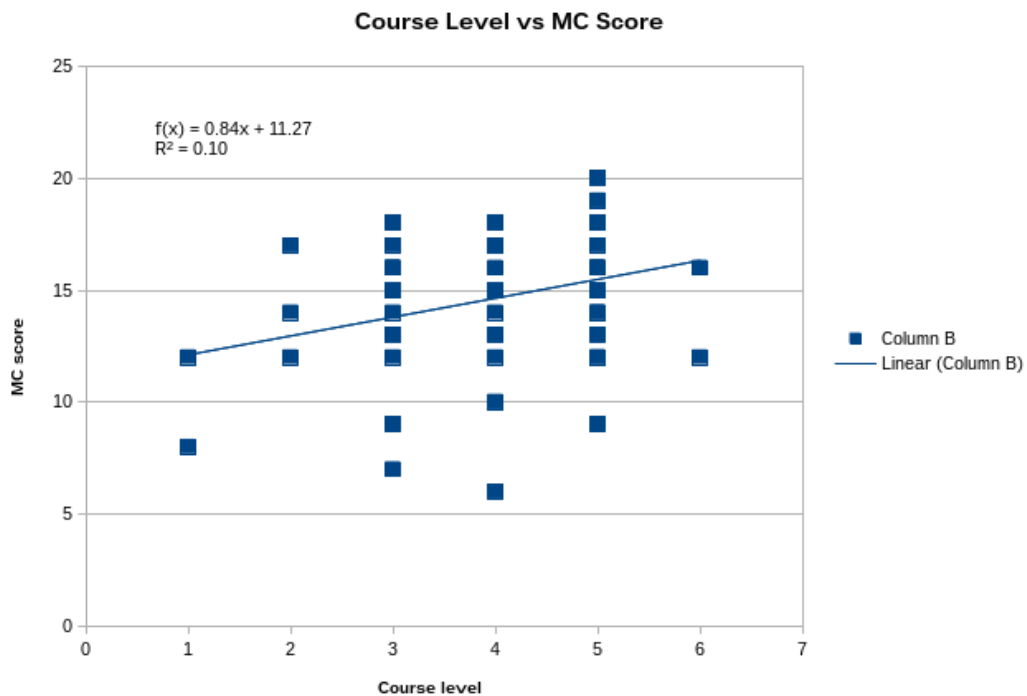


Figure 4: Course level versus MC test score

Multiple Choice vs Yes/No Answers

We didn't control for biases in self-reporting knowledge of words for the Yes/No word knowledge questions. As the test was not for coursework or accreditation, there was little motivation for false reporting. There was a reasonably strong correlation between the MC test and the YN score, possibly with a slight Dunning-Kruger effect occurring for those with the lowest knowledge as measured by the MC test. The MC test score is possibly less accurate than the YN score in measuring vocabulary knowledge, due to the 25 per cent chance of guessing correctly when a word is unknown.

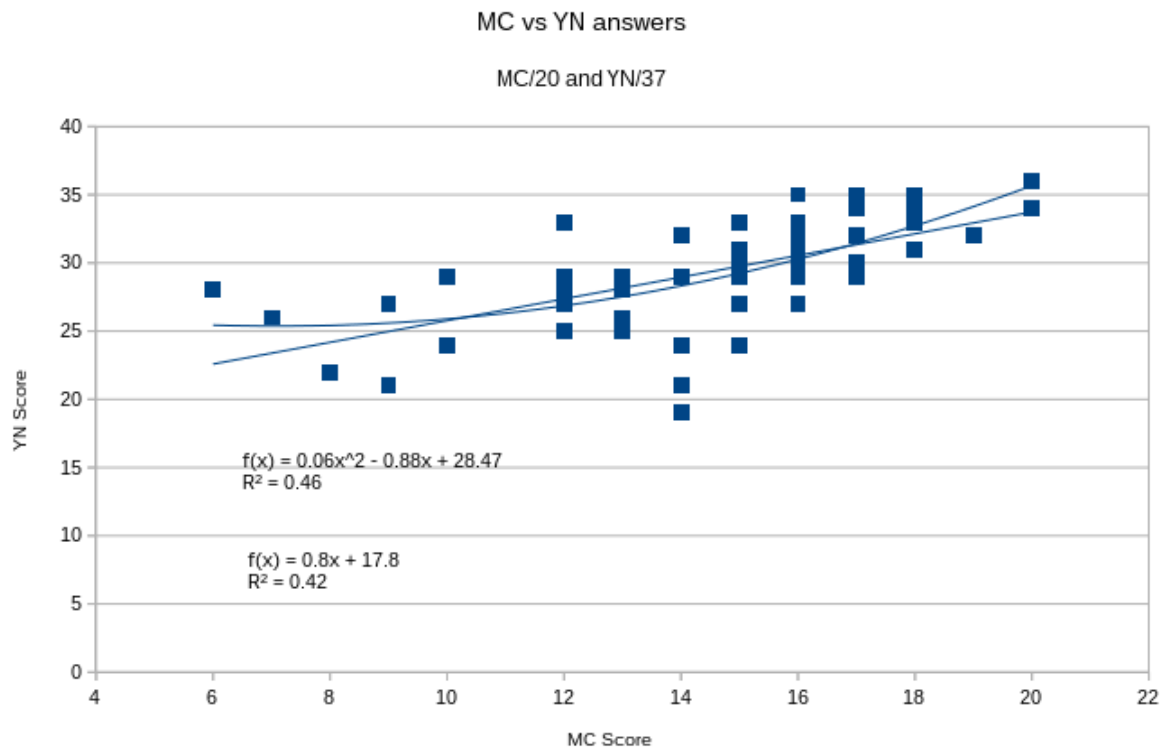


Figure 5: MC score versus YN score

Discussion

The initial (and therefore common to all participants) test was easier than ideal for the participants in the study, however, it still captured change in vocabulary knowledge over the 5 weeks of the course.

The lower correlation between test scores and course level compared to that of reading habits might appear to suggest that participants haven't been placed into the correct course level. However, the test focuses on words that occur in stories, rather than general or academic vocabulary, so it may be that participants have good academic vocabularies and low story vocabularies. Further, while there is much evidence that vocabulary knowledge correlates with reading, listening and writing skills, it has little relationship with

conversation skills, which would also be measured for determining entry level into an EAP course.

Regarding the relationship between reading habits in English and vocabulary scores, naturally it is impossible to say that one directly leads to the other. It is likely that those who scored less than 13/20 found reading books in English too difficult and therefore not enjoyable. Those reading more than five per year probably find some reading in English enjoyable, and 15/20 on the test is potentially an indicator of the level required before that is likely to happen. What was not captured by the questionnaire was whether participants habitually read *graded* readers.

Chapter 4: Change in Vocabulary after an Extensive Reading Programme

Participants were invited to take a second vocabulary test five weeks after the first one. Twenty two of the participants did so, being 11 intervention group participants, five diary only group participants, and six control group participants. Figure 6 shows the changes in MC scores for these participants, and Figure 7 shows the score changes for YN questions.

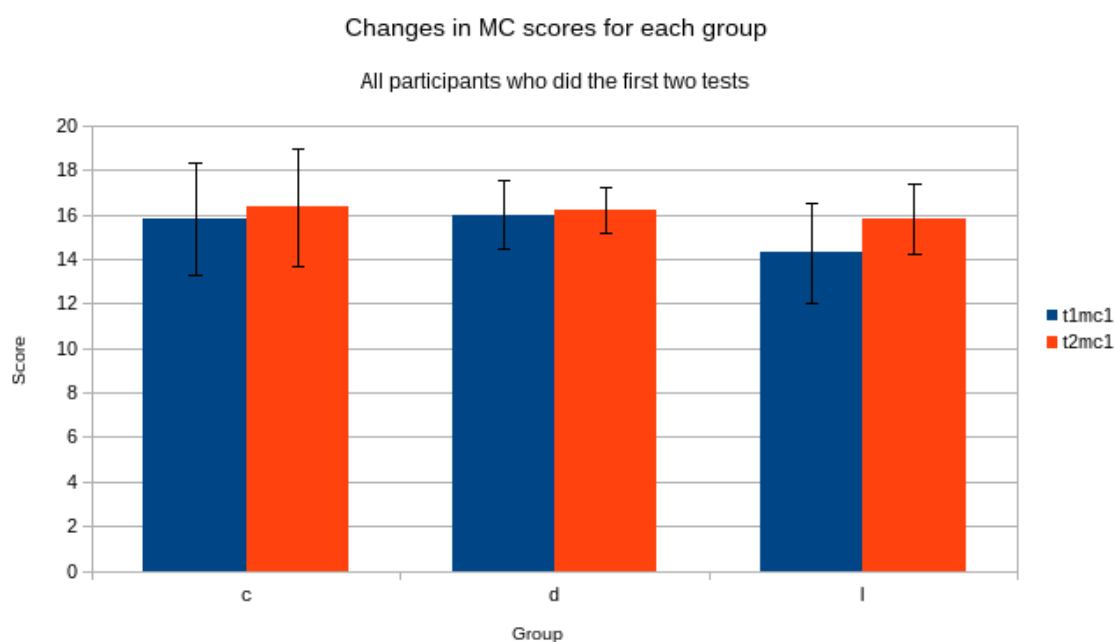


Figure 6: MC test scores before and after the 5 week programme for each group (c = six control participants, d = 5 diary only participants, I = 11 intervention group participants)

As can be seen from the 95 per cent confidence interval error bars in the graph, the small number of participants prevents any strong conclusions being drawn. Each group showed incremental improvement in their scores, and the intervention group showed a slightly larger increase on average than the other groups. The YN questions were less suggestive of a change from the intervention. It may just be that those with smaller vocabularies to start with had greater gains through whatever English language activities they engaged in. Some evidence for this is the reduction in the standard deviation for the second test in all cases except the diary only group.

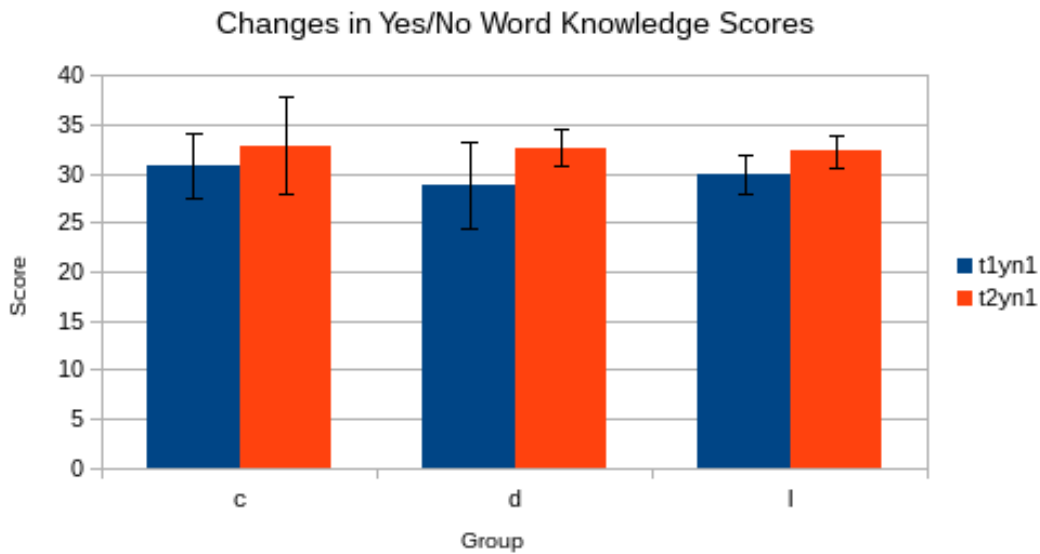


Figure 7: Changes in Yes/No word knowledge scores for each group. Error bars show standard error using *t*-distribution values due to the small number of participants. (*c* = six control participants, *d* = 5 diary only participants, *l* = 11 intervention group participants)

When examining the change in MC test scores to determine the effect size, it was found that when comparing the intervention group to the control the effect size was 0.63, requiring 60-80 participants to achieve significance at 95 per cent confidence, depending on the assumptions made. When the diary only and control groups were combined, the effect size was 0.74.

The relationship between reported reading and test scores was also examined, but this only involved eight participants and was dominated by two outliers. The strongest relationship found was that vocabulary familiarity increased with the difficulty of the text read (shown in Figure 8). Note that the range of difficulty is from “very easy” to halfway between “easy but I did not understand some words” and “not easy but I understood the story”. This result suggests that the text being read needs to be a little challenging for optimal vocabulary gain. However, what is not known is the amount of time taken to read the more challenging texts, so it isn’t clear whether there is greater gain per unit of time spent on reading.

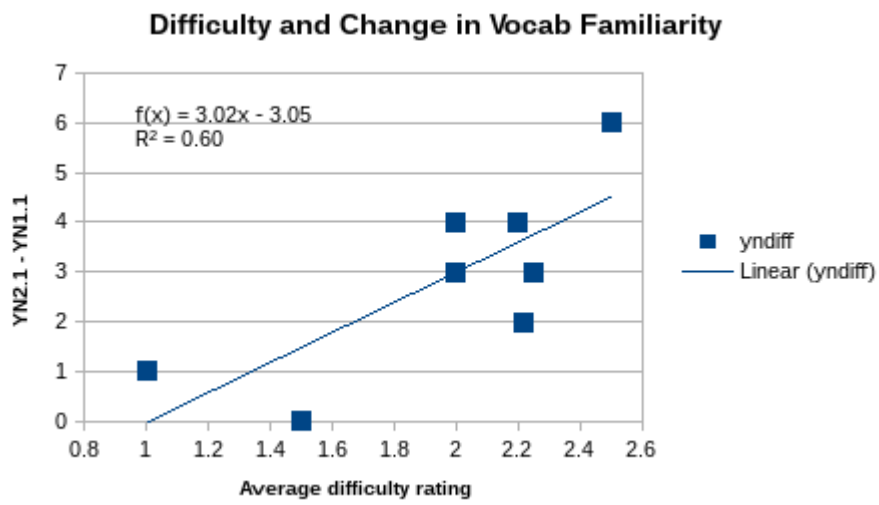


Figure 8: Difficulty and Change in Vocabulary Familiarity. The easiest difficulty is 1 and the most difficult is 5

Chapter 5: Long-Term Participants

Five participants chose to stay with the programme for more than 5 weeks, with two participants from the first group remaining with the study for the entire year. This chapter reports observations regarding these long term participants.

The long-term participants came from five different countries, each with a different native language (Chinese, Vietnamese, Arabic, Spanish and Thai). Participants were from either the intermediate (3) or advanced (2) levels of the English course. Their initial MC scores ranged from 13 to 17 out of 20, covering the lower (13) to upper quartile (17) range of initial MC scores of the entire cohort of 63 participants. Their initial YN scores ranged from 29 to 35, falling slightly higher than the interquartile range (27:32) of the entire cohort. All long-term participants reported that they read less than one book per year in English prior to the study, and either 1-5 books (2) or 6-10 books (3) per year in their native language. The countries of origin and language have been excluded from the descriptions to maintain participant privacy.

Participant 20 commenced the study while studying Advanced academic English. His initial MC and YN scores were 13 and 29. He remained in the study for the entire year (35 weeks), diligently attending all eight tests and reporting his reading, even after completing his EAP course and starting in his tertiary degree. He showed interest in the study itself, asking questions about the methodology. He believed that he had little aptitude for language. To him, the tests were like a game. He borrowed eight books from the collection, ranging from Stage 2 to Stage 5, and also read additional material. His final scores were 17 and 33.

Participant 22 commenced the study while enrolled in Advanced English. He remained in the study for 10 weeks. In conversations during the sessions he reported that the extensive reading programme improved his writing. He borrowed 17 books, ranging from Starter to Stage 4, during the 10 weeks he participated in the study, and also read additional material. His initial MC and YN scores were 16 and 32 respectively. His final scores were 17 and 36 respectively.

Participant 25 commenced the study while enrolled in Intermediate English. His initial scores were 15 and 29 respectively. He continued with the study for the entire year (35 weeks), with some gaps. Toward the end of the study he decided to read academic texts instead of the story books, as this would be more like the reading he would be required to do in his degree. In conversations during sessions he reported that the extensive reading programme improved his reading test marks for his EAP course. He also reported that it improved his grammar. He borrowed nine books from the collection, ranging from Stage 1 to Stage 5. His final MC and YN scores were 18 and 35 respectively.

Participant 50 stayed until the end of the study (25 weeks). She joined the study while enrolled in the Intermediate level of the English programme. Her initial scores were 17 and 35. She expressed a desire to read the entire library of books, and enjoyed all those that she read. She borrowed 17 books from the collection, ranging from Stage 1 to Stage 3, during the study. During conversations she mentioned that the programme helped her do well in

the reading assessments of her course. Her final scores were 17 and 37. She also sat the general vocabulary size test, achieving an initial score of 44 per cent on the initial B size test and 56 per cent on the final one.

Participant 54 commenced the programme while enrolled in the Intermediate English level. She stayed in the programme until the fourth test (20 weeks). She borrowed three books during the study. Her four reading entries into the reading diary had an average enjoyment rating of 4.5 and ease level of 3.25. Her MC and YN scores commenced at 16 and 27 respectively, and ended with 16 and 33 respectively. Her initial general vocabulary size test B score was 28 per cent and her final test A score was 36 per cent.

In summary, all long-term participants enjoyed reading the books and most commented that they observed an improvement in English language skills, the reported skills being writing, grammar and reading. They all had improvements in their final test scores, although some had no change in their final MC result, presumably due to not encountering the unknown words frequently enough during their reading, forgetting due to the time between reading and the test, or not being able to infer the exact meaning from the reading.

Chapter 6: Conclusions and Outputs

Findings

This project explored several methods of determining vocabulary gain from extensive reading. It was found that:

- A pilot meta-study of past research into vocabulary acquisition from reading revealed an approximately linear relationship between quantity read, and vocabulary gain, with additional variables being text difficulty for the person reading it. Maximum gain was found by reading large quantities of text at an easy level. This reflects the extensive reading recommendations from organisations such as the Extensive Reading Foundation. A rough guide is that if about 95 per cent of the words being read are known, then half of the unknown words will be learnt to some extent.
- The more an EAP student reads in English, the greater their receptive vocabulary, as measured with a vocabulary test that focuses on words more likely to be found in fiction than in EAP course materials. The relationship between these two factors is not necessarily causal in one direction only (one tends to read more if a text is interesting and easy enough to read), and there may be additional factors at play.
- There is measurable vocabulary gain from participation of EAP students in an extracurricular extensive reading programme. Having a set weekly time to meet for exchanging books, and particularly having an attractive, engaging set of books, increased participation rates. Merely providing a 10 minute introductory talk about extensive reading was insufficient for student engagement to occur.

Interdisciplinary Linkages

The project allowed the establishment of a working relationship between RMIT English Worldwide Melbourne and the School of Computer Science and IT (now part of the School of Science). It is expected that the relationship will continue well beyond the life of the project.

Factors Affecting the Success of the Project

The main positive factor was the enthusiasm and support of REW Melbourne management and staff for the project. This enabled the project to proceed with recruitment in classes, and the use of the Independent Learning Centre for sessions with participants.

The main hindrances were a very slow ethics approval process, and reduced availability of key team members. Additional staff were added to the project to increase its chance of success. Dissemination is awaiting acceptance of abstracts and papers by various venues.

Dissemination and Project Impact

There are three main cohorts to disseminate results to: language teachers, language students and applied linguistics researchers. The dissemination plan reflects these three

cohorts. Other than the project report, the plan is to disseminate the results via conferences, journals, web-sites and talks.

The current impact level of the project, on the IMPEL scale is at level 4. Participants in the project have seen the value of extensive reading in improving their English language skill. Various talks have been given related to extensive reading and vocabulary (see Appendix B). Some attendees of talks were new to the concept and availability of graded readers, and have chosen to use graded readers for their own foreign language skill development. Papers are being submitted to journals and conferences related to language learning. These are yet to be accepted.

The project was evaluated in various ways at different stages. Before submission feedback was provided to improve the grant application. Discussion between team members increased the robustness of the study design. Feedback from seminar attendees provided new insights and confirmed issues. Reviewer reports from conference submissions led to stronger arguments for the study design. The final report also benefited from feedback on a draft.

Implementing an Extensive Reading Programme at an ELICOS Provider

EAP courses can be short of spare class time, making the incorporation of an extensive reading programme within the course difficult. However, where there are extra-curricular activities already available for students, it would be simple to incorporate an extensive reading element to them.

For example, conversation classes can be converted partially or wholly into book clubs, where students discuss the books that they've been reading during the week. Where there are no libraries of graded readers in the institution itself, there are often nearby libraries where students can borrow appropriate books. In this project, a box of over 40 books at seven levels kept keen readers happy for many weeks. Some participants read two per week, whereas others took five weeks to finish one reading book.

Links to Other OLT Projects

There have been several OLT projects related to language skills of tertiary students. The OLT Good Practice Report - English Language Proficiency summarises these and provides recommendations for Universities for achieving a suitable English Language Proficiency standard for graduates (Arkoudis and Doughney, 2014). In light of the recommendations in that report, the best approach for extensive reading would be to target EAP students and instil good reading habits prior to attending university.

Future Work

Due to the constraints of the environment in which the study was held, only 63 participants took part, with just 22 remaining in the study for a second test. This limited the strength of evidence regarding vocabulary and extensive reading. Power analysis suggests needing about 80 participants, which means recruiting about 120 to allow for drop-out. Students in

this pilot study sat tests on paper, leading to the need for data entry prior to analysis. The intention is to study a larger number of students via on-line tests, and integrate the programme into REW's normal practice. This should allow more detailed numerical analysis of the various factors originally proposed in this study. The manageability of a larger scale study will be enhanced by software that handles book tracking, communication and testing.

References

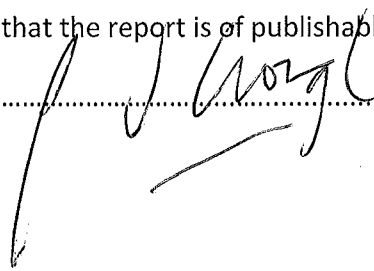
- Arkoudis, S. and Doughney, L. (2014). *Good Practice Report: English Language Proficiency*. Sydney, Australia: Office for Learning and Teaching.
- Horst, M. (2005). Learning L2 vocabulary through extensive reading: A measurement study. *The Canadian Modern Language Review*, 61(3), 355–382.
- Horst, M., Cobb, T., and Meara, P. (1997). Beyond a Clockwork Orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language*, 11(2), 207–223.
- Kablaoui, S. and Khabbaz, A. (2012). Developing reading skills of learners from Arabic-speaking backgrounds. *Cambridge ESOL Research Notes*, 48, 7-10.
- Laufer, B. (2000). Task effect on instructed vocabulary learning: the hypothesis of involvement. *Selected Papers from AILA '99 Tokyo*, 47–62. Tokyo, Japan: Waseda University Press.
- Nation, P. and Waring, R. (1997). Vocabulary size, text coverage and word lists, in N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Design, Acquisition, Pedagogy* (pp.6–19). Cambridge, UK: Cambridge University Press.
- Uitdenbogerd, A. L. (2005). Readability of French as a foreign language and its uses. In J. Kay, A. Turpin & R. Wilkinson (eds.), *ADCS 2005: Proceedings of the tenth Australasian Document Computing Symposium*, 19-25. Sydney, Australia: School of Information Technologies, The University of Sydney.
- Waring, R. and Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language*, 15(2), 130-163.
- Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, 28(1), 46–65.

Appendix A

Certification by Deputy Vice-Chancellor (or equivalent)

I certify that all parts of the final report for this OLT grant/fellowship (remove as appropriate) provide an accurate representation of the implementation, impact and findings of the project, and that the report is of publishable quality.

Name: Date:

 23 iii 16

Appendix B: Workshop Presentations

Date	Title	Venue
5/11/2014	Extensive Reading Research	RMIT Science, Health and Engineering Educational Research (SHEER) Centre members meeting
30/11/2014	Tools for Supporting Language Acquisition via Extensive Reading	Paper presented at Natural Language Processing Techniques for Educational Applications (NLP-TEA), a workshop associated with the International Conference on Computers in Education (ICCE) and including the ICCE Conference on Technology Enhanced Language Learning (TELL), Nara, Japan. (Partially funded by the OLT grant.)
15/09/2015	A Study of Extensive Reading and Vocabulary: Some Preliminary Results	Guest talk presented at the University of Melbourne Department of Computing and Information Systems, Natural Language Processing research group seminar series.
9/12/2015	Word Transformation Heuristics Against Lexicons for Cognate Detection	Poster at the Workshop of the Australian Language Technology Association, collocated with Australian Linguistics Society at Confluence 2015 at University of Western Sydney, Parramatta.
15/04/2016	Extensive Reading and Vocabulary Acquisition in the EAP Context	Guest presentation at RMIT English Worldwide Melbourne's Professional Development day for teaching staff.
June 2016	Classifying Type-level Word Complexity using Random Forests with Corpus and Word List Features	Semeval 2016 Task 11, to be held at the 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (attendance not funded by the OLT grant)

Appendix C: Impact Plan

	Anticipated changes at:			
	Project completion	Six months post-completion	Twelve months post-completion	Twenty-four months post-completion
Team members	A collaborative relationship between RMIT Computer Science and IT, and RMIT English Worldwide	Two journal and two conference articles submitted	Published conference papers	Published journal papers Strong case for promotion
Immediate students	Improved language skills via reading, An extensive reading habit that will extend beyond their EAP course	Greater reading skill in their degree courses	Better grades in their degree	
Spreading the word	Local talks and website	Talks further afield	Conference talks	Published journal articles
Narrow opportunistic adoption	Continued extensive reading programme at REW, as part of further study	Conversation groups that are run as reading clubs at REW	Continued conversation groups without the researchers' direct involvement	
Narrow systemic adoption				Integration into REW structure
Broad opportunistic adoption			Trials at other EAP course providers	
Broad systemic adoption				National guidelines for extensive reading

The programme will continue after this project is finished. Further funding will be sought to enable an on-line system that supports a larger scale project, possibly extending beyond REW. This will allow coordination of reading book club programmes nationally, and possibly internationally.

Appendix D: Participant Questionnaire

The following pages show a paper version of the Google Form used for gathering participant details for the study.

Participant Questionnaire

*Required

1. **Name ***

2. **Gender**

Mark only one oval.

Male

Female

3. **How old are you?**

Mark only one oval.

18-20

21-25

26-30

31-40

older than 40

4. **Where are you from?**

Country. Example: Indonesia

5. **What level of study have you finished?**

Mark only one oval.

High School Skip to question 7.

Foundation Studies Skip to question 7.

TAFE Skip to question 6.

Bachelor Degree Skip to question 6.

Master Degree Skip to question 6.

PhD Skip to question 6.

Other: _____ Skip to question 6.

Study Details

6. What did you study?

(for TAFE, Bachelor, Master or PhD study)

Mark only one oval.

- Arts
- Science
- Engineering
- Law
- Medicine
- Business
- Other: _____

English Language Knowledge

7. What class are you studying in? *

Mark only one oval.

- Elementary
- Pre Intermediate
- Intermediate
- Upper Intermediate
- Advanced
- Advanced Plus

8. Did you study English before? *

Mark only one oval.

- Yes *Skip to question 9.*
- No *Skip to question 11.*

English Study

9. How long did you study English in a non-English speaking country? *

(your home country)

Mark only one oval.

- less than 20 hours
- 20-100 hours
- 5-10 weeks
- 10-20 weeks
- 20 weeks to 1 year
- 1-2 years
- Other: _____

10. How long did you study English in an English speaking country? *

(Australia)

Mark only one oval.

- less than 20 hours
- 20-100 hours
- 5-10 weeks
- 10-20 weeks
- 20 weeks to 1 year
- 1-2 years
- Other: _____

Other Languages

11. What is your native language? *

12. Do you know any other languages? *

Mark only one oval.

- Yes
- No

Known Language 1

13. Which other language do you know?

14. How long did you study the language?

Mark only one oval.

- less than 20 hours
- 20-100 hours
- 5-10 weeks
- 10-20 weeks
- 20 weeks to 1 year
- 1-2 years
- Other: _____

15. Do you know any other languages?

Mark only one oval.

- Yes
- No

Known Language 2

16. Which other language do you know?

17. How long did you study the language?

Mark only one oval.

- less than 20 hours
- 20-100 hours
- 5-10 weeks
- 10-20 weeks
- 20 weeks to 1 year
- 1-2 years
- Other: _____

18. Do you know any other languages?

Mark only one oval.

- Yes
- No

Known Language 3

19. Which other language do you know?

20. How long did you study the language?

Mark only one oval.

- less than 20 hours
- 20-100 hours
- 5-10 weeks
- 10-20 weeks
- 20 weeks to 1 year
- 1-2 years
- Other: _____

Reading

21. In your native language, how much do you read?

Mark only one oval.

- I do not read books in my native language
- Less than one book per year
- 1-5 books per year
- 6-10 books per year
- 11-20 books per year
- More than 20 books per year

22. In English, how much do you read?

Mark only one oval.

- I do not read books in English
- Less than one book per year
- 1-5 books per year
- 6-10 books per year
- 11-20 books per year
- More than 20 books per year

23. In other languages, how much do you read?

Mark only one oval.

- I do not read books in other languages
- Less than one book per year
- 1-5 books per year
- 6-10 books per year
- 11-20 books per year
- More than 20 books per year