

# THE TESOLANZ JOURNAL

VOLUME 22 2014

ISSN 1172-9694



The TESOLANZ Journal is published once a year by TESOLANZ, Teachers of English to Speakers of Other Languages, Aotearoa/New Zealand (Incorporated), Te Ropu Kaiwhakaako Reo Ingarihi ki Iwi Reo Ke. The journal is double blind peer reviewed. A subscription to the journal is included in the annual membership fee of the Association. Enquiries about membership and subscription to the journal should be addressed to:

TESOLANZ  
Linguistics and Applied Language Studies  
Victoria University of Wellington  
PO Box 600  
Tel: (04) 463 5600  
Fax: (04) 463 5604

Those interested in submitting an article or book review should consult the Notes for Contributors printed at the back of this volume and be a current member of TESOLANZ.

**CO-EDITOR**

Dr Susan Gray  
Faculty of Education  
University of Auckland

**CO-EDITOR**

Dr Angela Joe  
Linguistics & Applied Language Studies  
Victoria University of Wellington

**EDITORIAL ASSISTANT**

Dr Patrick Coelho  
Linguistics & Applied Language Studies  
Victoria University of Wellington

**REVIEWS EDITOR**

Dr Rosemary Wette  
Applied Language Studies & Linguistics  
University of Auckland

**EDITORIAL BOARD**

Associate Professor Gary Barkhuizen  
Applied Language Studies & Linguistics  
University of Auckland

Associate Professor David Crabbe  
Faculty of Education  
Victoria University of Wellington

Dr Margaret Kitchen  
Faculty of Education  
University of Auckland

Dr Jonathan Newton  
Linguistics & Applied Language Studies  
Victoria University of Wellington

Associate Professor John Read  
Applied Language Studies & Linguistics  
University of Auckland

Dr Corinne Seals  
Linguistics & Applied Language Studies  
Victoria University of Wellington

Dr Gillian Skyrme  
School of Languages  
Massey University

Professor Cynthia White  
School of Languages  
Massey University



---

# THE TESOLANZ JOURNAL

---

## CONTENTS

<b>EDITORIAL .....</b>	<b>i</b>
 <b>ARTICLES</b>	
Coming to grips with quality in language assessment..... <i>John Read</i>	1
Creating and trialling six versions of the vocabulary size test ..... <i>Averil Coxhead, Paul Nation &amp; Dalice Slim</i>	13
Developing speaking fluency with the 4/3/2 technique: An exploratory study..... <i>John Macalister</i>	28
Teaching and learning academic writing: Narratives of future destination..... <i>Martin Andrew &amp; Zina Romova</i>	43
 <b>BOOK REVIEWS</b>	
Codeswitching in university English-medium classes: Asian perspectives <i>Reviewed by Lynn Grant</i> .....	56
The strategy factor in successful language learning <i>Reviewed by Sara Amani</i> .....	57
Teaching materials and the roles of EFL/ESL teachers: Practice and theory <i>Reviewed by Patrick Coleman</i> .....	59
What English language teachers need to know. Volume III: Designing curriculum <i>Reviewed by Rosemary Wette</i> .....	61
1) Progressive skills in English level 4 course book, 2) Progressive skills in English level 4 teacher's book and 3) Progressive skills in English level 4 workbook <i>Reviewed by Margaret Bade</i> .....	62
Developing materials for language teaching <i>Reviewed by Mhairi Mackay</i> .....	64
English for academic study: Grammar for writing <i>Reviewed by Martin White</i> .....	65
 <b>NOTES FOR CONTRIBUTORS .....</b>	 <b>68</b>

## EDITORIAL

An ongoing issue for secondary and tertiary language educators focuses on how to assess language use, what tools or constructs to use and how to promote academic literacy. The articles selected for publication in this year's journal provide practical suggestions addressing these concerns. With regard to language assessment, we are delighted to include Associate Professor John Read's invited article based on his plenary presentation at the recent CLESOL conference in Wellington. The other three articles relate to the development of six new versions of Paul Nation's Vocabulary Size Test, language learning opportunities afforded by the 4/3/2 technique and the use of narratives in academic writing portfolios to engage students with discourses needed in their future disciplines.

In our first article Read discusses how to conceptualise quality in language assessment from the viewpoint of "standards", in particular the contrast between standardised tests and standards-based assessment. He then introduces the concepts of validity and fairness and how they apply to three different measures of academic writing ability.

In the second article Coxhead, Nation and Sim report on the development and trialling of six versions of Nation's Vocabulary Size Test (VST). Forty-six participants took all six versions of the test and the results suggest that these versions fall into two sets of three parallel tests. The results are discussed in relation to variables including first language, gender, status as a university student, age, and level of education.

While earlier research into the 4/3/2 activity has focused on measurable markers of fluency, such as words or syllables per minute, Macalister wanted to see what changes in the quality of speech occurred. This exploratory study was carried out with university students in an intact class in an EFL context, and found that there were lexical improvements and changes in the ways in which content and organisation were being signalled. This provides support for the use of the 4/3/2 as a language learning opportunity.

In the final article, Andrew and Romova argue that providers of tertiary academic writing courses need to consider the destinations and 'future imagined communities' of their students. Their study extracts narratives of experience from raw data consisting in transcripts of pre- and post-programme focus group interviews and triangulated by students' reflective self-analyses of their improvements to multi-drafted academic texts. The study maintains that since students are highly invested in their future destinations, they need to be able to create and recreate texts characteristic of those future communities in their work during their academic writing course. This suggests a strong need for providers to keep a text bank of artefacts belonging to the discourse communities of such disciplines as Early Childhood Education, Nursing and Business, three disciplines for which many students in the cohort were bound.

The book reviews that follow have been selected to cover a range of areas relevant to language teaching and research and to highlight current issues being explored in the literature.

In conclusion, it is important to thank all the contributors, those who submitted manuscripts for consideration in this year's volume of the journal. Part of the process involved in preparing a manuscript for publication involves responding to questions and guidance from experienced peers. In this respect, we are indebted to members of the Editorial Board for their perspicacity and generosity of spirit that characterize the reviews.

We encourage the many readers of the TESOLANZ Journal who have not yet contributed to the publication to consider doing so in the following year – either individually, or, collaboratively. You will find Notes for Contributors at the end of the journal, but always feel free to contact the corresponding Editor by email ([angela.joe@vuw.ac.nz](mailto:angela.joe@vuw.ac.nz)) if you require any additional information. The closing date for receiving manuscripts will be Monday, 24 August 2015.

---

## ARTICLES

---





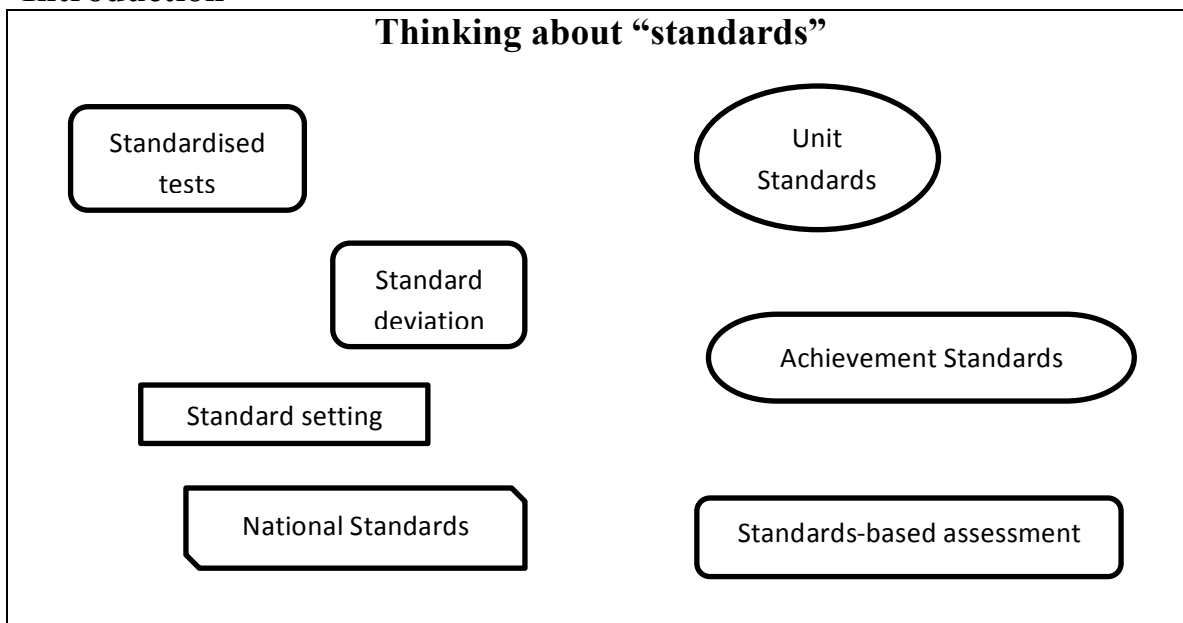
# COMING TO GRIPS WITH QUALITY IN LANGUAGE ASSESSMENT

John Read  
University of Auckland

## Abstract

*This article is a written version of a plenary presentation at the CLESOL Conference in Wellington in July 2014. It considers how we can conceptualise quality in language assessment through exploring different meanings of the term “standard”, and in particular the contrast between standardised tests and standards-based assessment. The criteria of validity and fairness are introduced and then applied to three measures of academic writing ability: Task 2 of the IELTS Academic Writing Module, the integrated writing task in the internet-based TOEFL, and the NZQA EAP unit standard on writing a crafted and researched text. Each has its strengths and weaknesses in terms of validity and fairness, so that there is no single best way of assessing the construct.*

## Introduction



*Figure 1: Uses of the term “standard” in educational assessment*

One word which comes to mind when we think about quality – in language assessment as in other fields – is “standard”. As Figure 1 highlights, the term is employed in a number of different ways in educational assessment. It is useful to try to unpick some of the meanings of the word, as the basis for approaching an understanding of what represents quality in language assessment. In this article, I will focus in particular on the contrast between standardised tests and standards-based assessment, as it applies to the assessment of academic writing ability.

## Standardised tests

Let us start with the concept of a standardised test. This term in itself may have various meanings in different countries, but according to the *Dictionary of language testing* (Davies et al. 1999, p. 187), a standardised test has the following set of features:

- A rigorous process of development, trialling and revision to establish the test's measurement quality.
- Standard content in all forms of the test, based on an explicit set of specifications.
- A set of norms based on the distribution of scores in a given population.
- Standard procedures for administering and scoring the test.

In the US, the critical feature is the third one. For nearly 100 years the American education system has been under the dominance of the psychometric paradigm, with the multiple-choice item as the basic building block for a complex set of statistical models and analyses designed to discriminate levels of ability or achievement in a large population of students. This norm-referenced approach is often referred to as “grading on the curve”. The other three features apply more widely to large-scale tests and exams, including those like the International English Language Testing System (IELTS), which follow more criterion-referenced principles, especially in the assessment of speaking and writing skills.

Thus, with standardised tests and examinations the key element is *consistency*: in the content, the administrative conditions and procedures, the scoring and the interpretation of the results. If high-stakes decisions are being made about learners on the basis of their test performance, it is important that their scores should not depend on where or when they took the test, or how their responses were marked. It can be argued that the consistency which is achieved by standardised testing procedures also promotes *fairness*. In language assessment, the concept of fairness has not had the same currency as reliability, validity or practicality, but it has been adopted by some authors, notably Antony Kunnan (2014; see also Xi, 2010), in the context of promoting ethical standards in the field. Although Kunnan acknowledges that other scholars see fairness as essentially one aspect of test validity, he argues that the term gives a distinctive emphasis to the idea that all learners should have an equal chance of doing their best in an assessment, regardless of gender, ethnicity, language background, type of school, geographical region, and so on.

The question then is whether the interests of fairness are best served by a standardised test. It is not difficult to see ways in which such a test may be quite unfair:

- Some learners have more opportunities to prepare for the test than others if, for example, their parents can afford to pay for extra tuition or test preparation materials.
- This reflects the fact that standardised test items and tasks tend to lend themselves to intensive coaching in test-taking techniques.

- Certain items, tasks or topics may favour some groups of test-takers over others.
- The testing conditions and administrative procedures may vary from one test site to another.
- The performance of some learners is badly affected by test anxiety and other psychological states.
- These tests tend to give priority to reliability, at the expense of validity. Thus, standardised language tests have traditionally neglected listening and speaking skills, concentrating on what can be assessed with a high degree of consistency through a written exam.

### Standards-based assessment

Increasing recognition of the limitations of standardised tests in this broad sense has given rise to an alternative paradigm: standards-based assessment (SBA). This approach has become familiar in the New Zealand secondary school context through the introduction of the National Certificate of Educational Achievement (NCEA). A similar approach, at least on an informal basis, was already well established in New Zealand primary and intermediate schools, although the promotion of National Standards by the current Government represents a controversial initiative both to formalise the assessment and to disseminate the results publicly.

Obviously, the word “standard” is being used in a different sense in standards-based assessment from its meaning in standardised testing. Here it refers to a statement about the outcomes of learning which students are expected to demonstrate, through a variety of assessment procedures. As implemented in secondary schools, the NCEA is a hybrid system in the sense that the mainstream subjects in the curriculum are assessed through a combination of external (exam-based) and internal (classroom-based) achievement standards. It is the internal standards which reflect more of the essence of standards-based assessment, as stated in this text from the Ministry of Education’s teacher resource website, Te Kete Ipurangi (TKI):

- Standards-based assessment allows us to make **judgments** about the level of an individual’s learning with respect to **shared benchmarks of expected performance**, supported by **exemplars**.
- Each standard has a number of components that students need to bring together to achieve the standard. **Teachers’ judgments [focus] on the work as a whole**, as opposed to the result from a single snapshot assessment.
- Teachers **improve the consistency of their judgments through engaging in moderation practices**, [which enable them] to develop a shared understanding of the meaning of standards and how to apply them in a range of cases (based on TKI, n.d.).

I have highlighted in bold font some of the key elements. Teachers make “judgements” on the work of individual learners in their classes, rather than engaging in marking or grading in the traditional sense. This introduces more subjectivity into the assessment, although the teachers’ judgements are guided by specifications of the

learning outcomes and samples of learner performance that meet the standard, as well as collegial activities to “develop a shared understanding” of the standards and how to apply them. The phrase “improve the consistency” is interesting because it implies that initially teachers are not very reliable assessors of their students’ work in relation to the national standards, but the hope is that over time they will become more proficient at it.

This leads us to the point that SBA is not free of the potential for unfairness, especially if it is used for a relatively high-stakes purpose such as the award of a national qualification:

- Teachers will vary in their experience and expertise in assessment, that is to say, some will be more “assessment-literate” than others.
- There will inevitably be some inconsistency in the way the standard is interpreted and assessment tasks are designed by different teachers.
- It is difficult to avoid differences in assessment conditions and the amount of support provided to learners, especially since teachers combine the teaching and assessing roles.
- For the same reason, steps need to be taken to avoid bias or favouritism towards certain learners.
- Although procedures are put in place to moderate teacher judgements, at both local schools and on a regional or national basis, only a sample of learners can have their work moderated; the NZQA guideline for external moderation is 10 per cent (NZQA, n.d.).
- This certainly creates the potential for unfair assessments to go undetected or uncorrected.

So the basic point I want to make here is that both standardised tests and standards-based assessment have their sources of unfairness. The best we can hope for is that threats to fairness are reduced as much as possible in any individual assessment, in keeping with the stakes involved for the student.

### **Validity and fairness**

In the rest of the article, I would like to explore this point by reference to the assessment of academic writing ability, but first I need to add the concept of validity into the discussion. Validity theory has become a complex area of modern educational assessment and I cannot do justice to it in the limited space I have available here.

We can integrate the concepts of validity and fairness in this way:

- Conceptually, an assessment is valid if we can meaningfully interpret the results as representing the level of knowledge, skills or ability of a population of learners.
- The relevant knowledge, skills or ability are defined as one or more constructs.
- An assessment is fair if it gives valid results for all the learners and does not disadvantage particular sub-groups.

To take the concept of validity a step further, let me introduce what the guru of modern validity theory, Samuel Messick, argues are the two basic threats to the validity of an assessment: construct under-representation and construct-irrelevant variance (Messick, 1996). Let us look at each one in turn.

Construct under-representation means that the assessment does not adequately cover the range of knowledge or skills defined by the construct. In the case of academic writing ability, particularly at the university level, writing is not a one-off event but a process which includes obtaining relevant information, planning, drafting, revising, formatting and so on. Writing tests typically represent this process only to a very limited extent.

Construct-irrelevant variance means that the assessment is affected by the needs for skills that are not relevant to the targeted construct. Again using academic writing ability as an example, correct spelling, good handwriting and keyboard skills may have an influence on how students' writing is assessed, but we need to ask whether each of these attributes is an essential element of the ability to produce good academic writing.

We will revisit these concepts in the discussion which follows.

### **Assessing academic writing ability**

I now want to explore issues of validity and fairness by considering three ways of assessing the construct of academic writing ability, as represented by these three particular assessments:

- Task 2 in the Academic Writing Module of the International English Language Testing System (IELTS) ([www.ielts.org](http://www.ielts.org))
- The integrated task in the writing section in the internet-based Test of English as a Foreign Language (TOEFL iBT) ([www.ets.org/toefl](http://www.ets.org/toefl))
- One of the unit standards in the domain of English for Academic Purposes on the NZ Qualifications Framework, specifically US 22750: Write a crafted text using researched material in English for an academic purpose ([www.nzqa.govt.nz](http://www.nzqa.govt.nz))

### **IELTS academic writing task**

Figure 2 shows a practice example of Task 2 of the IELTS Academic Writing Module:

You should spend about 40 minutes on this task.

Write about the following topic:

***The threat of nuclear weapons maintains world peace. Nuclear power provides cheap and clean energy.***

***The benefits of nuclear technology far outweigh the disadvantages. To what extent do you agree or disagree?***

Give reasons for your answer and include any relevant examples from your knowledge or experience.

Write at least 250 words.

*Figure 2: Sample IELTS Academic Writing, Task 2*

This is a good example of what is called in the writing assessment literature a “timed impromptu” task (Weigle, 2002). The candidates are given a limited amount of time to write on a topic that is not revealed to them in advance but which they are assumed to be able to write about on the basis of some background knowledge. On the other hand, understanding the writing prompt crucially depends on knowing the meaning of “nuclear” and the words it collocates with: “weapons”, “power” and “technology”. This may be an issue with less proficient candidates.

Along with the other writing task, candidate scripts are rated at local test centres on the standard IELTS scoring scale from Band 9 to Band 0 according to four analytic criteria: Task response, Coherence and cohesion, Lexical resource, and Grammatical range and accuracy. The examiners are typically experienced ESOL teachers who, in the words of the IELTS website, “undergo intensive face to face training and standardisation to ensure that they can apply the descriptors in a valid and reliable manner”. This includes re-certification every two years and regular monitoring of their ratings in between.

If we look at this task from the perspective of validity and fairness, the following points arise:

- It clearly under-represents the construct of academic writing ability. At best, it simulates the writing of essay-type answers in a university exam, although even then students write on a topic that they have studied during the course.
- This type of task is vulnerable to memorised responses by candidates who have engaged in intensive test preparation. Learners in China are renowned for their ability to memorise large numbers of whole texts – or at least adaptable templates – based on typical IELTS writing topics.
- For any particular topic, although candidates are expected to have sufficient ideas to compose a 250-word argument essay, inevitably some will have more interest in, and information about, it than others.
- Until recently, for most candidates a single examiner has rated their writing script (both Tasks 1 and 2), whereas the accepted practice in high-stakes writing assessment is to have at least two raters. However, as of 2014, IELTS has adopted “distributed marking”, which requires that each of a candidate’s writing tasks is rated by a different examiner independently.
- Since IELTS is a paper-based test, the influence of poor handwriting on examiner judgements cannot be discounted, even though examiners are trained to do their best to decipher such scripts.

### **TOEFL integrated writing task**

The second example of an academic writing test is found in the internet-based TOEFL (iBT). Introduced worldwide from 2005, this is a computer-based test which includes two writing tasks. One is quite similar to the IELTS Task 2 we have just looked at, in that the candidates write a short argument essay on a topic of general interest. In the iBT, it is known as an “independent” task because the amount of reading involved in understanding the prompt is kept to a minimum and the focus is on writing the essay. The other task is “integrative”, in the sense that it requires the candidates first to attend to both written and spoken input material on a given topic and then to draw on information from both sources in writing their own response. A sample task is summarised in Figure 3.

1. Candidates have three minutes to read a passage (292 words) arguing for the adoption of computer-based voting systems in the US.
2. Candidates listen to a mini-lecture (286 words) giving reasons to be cautious about computer-based voting.
3. Writing Task (20 minutes):  
Summarize the points made in the lecture, being sure to explain how they oppose specific points made in the reading passage.  
  
Typically, an effective response will be 150 to 225 words.

*Figure 3: Outline of a sample iBT integrated writing task*



In contrast to IELTS, the iBT writing tasks are scored on a 5-0 holistic scale. There is no space to reproduce the whole scale here, but the descriptor for Level 5 is as follows:

A response at this level successfully selects the important information from the lecture and coherently and accurately presents this information in relation to the relevant information presented in the reading. The response is well organized, and occasional language errors that are present do not result in inaccurate or imprecise presentation of content or connections.

Student responses to the tasks are sent through the internet to the central server for the Online Network for Evaluation (ONE) at the Educational Testing Service (ETS) in the US. The electronic scripts are first assigned to human raters working remotely at home across North America, under continuous monitoring by a scoring leader and with a requirement to recalibrate their judgements on a daily basis. Then an automated system scores the script and, if there is a discrepancy between the two ratings, a second human score is obtained.

Again, if we apply the concepts of validity and fairness, the following points can be made:

- An integrated iBT task goes a little further towards representing the construct of academic writing by providing source material, allowing for note-taking, and requiring some synthesis of the ideas from the two sources; however, it is very much a micro-version of a real researched essay.
- Since TOEFL is a computer-based test, facility with using a keyboard in a stressful testing situation is an issue. Although we accept the QWERTY keyboard as the norm, several other layouts are in widespread use in particular countries, and there is some unpublished evidence that this may have at least a small effect on the writing performance of test-takers who are unfamiliar with the QWERTY layout.
- The remote scoring system means that the candidates are completely anonymous to the raters, who in turn can be assigned scripts by learners from a wide range of countries and language backgrounds.
- Automated scoring of writing is somewhat difficult to accept, especially by those who have not kept up with the impressive advances in the technology in recent years (see Carr, 2014 and Xi, 2012 for updates). There are still obvious limits on what aspects of writing the computer can evaluate, but ETS argues that their combination of human and automated rating gives the best of both worlds: “Combining human judgment for content and meaning, and automated scoring for linguistic features, ensures consistent, quality scores” (ETS, 2014).
- This high-tech approach to writing assessment can realistically be implemented only by very large testing organisations like ETS.

Thus, the integrated writing task used in the iBT has a number of attractive features, but it retains the limitations of a large-scale standardised test.

### **NZQA unit standard for academic writing**

The third approach to assessing academic writing ability takes us back to standards-based assessment within the New Zealand Qualifications Framework and, in particular, a unit standard designed for students preparing to study at the tertiary level. It is one of a suite of five standards in the domain of English for Academic Purposes (EAP). Two of the EAP standards, including this one, are accepted by the universities as meeting the literacy requirement for University Entrance in the case of domestic students. In addition, all five EAP standards form the core of the recently introduced New Zealand Certificate in English Language (NZCEL), Level 4, which is intended to certify that international students and others have the academic language proficiency needed for tertiary study. However, the universities have yet to accept NZCEL-4 as fulfilling the English language requirement for admission.

The present EAP standard, with the reference number 22750, has the learning outcome: Write a crafted text using researched material in English for an academic purpose. The full specification of the standard can be found on the NZQA website ([www.nzqa.govt.nz/nqfdocs/units/pdf/22750.pdf](http://www.nzqa.govt.nz/nqfdocs/units/pdf/22750.pdf)), but the basic task is to produce a text of 800 words on a topic chosen by the class teacher. In preparation for the task, the students study relevant source material; in fact, this preparatory reading can be planned to meet the requirements for another unit standard on academic reading. The students receive a checklist to guide their preparation and can submit one draft for general feedback; they can also consult a model text.

Here is a sample task:

You have been reading about migration issues. Write a text which addresses the following research questions:

- What are the main issues faced by teenage migrants into New Zealand?
- Discuss a range of possible solutions to these problems.

The other main component of the unit standard is a set of “evidence requirements”, all of which must be met in order for the student to be credited with the standard:

- 1.1 Writing addresses and develops the topic in a manner appropriate to audience and academic purpose.
- 1.2 Ideas are developed and display a broad knowledge base to achieve the purpose of the discussion.
- 1.3 Text structure is clear, cohesive and coherent, with a logical progression.
- 1.4 Writing uses a formal style appropriate to the academic context.
- 1.5 Writing makes consistent use of appropriate lexical and grammatical forms throughout the text.
- 1.6 Writing integrates source material.
- 1.7 Source material is acknowledged.

Some of the issues that arise from this method of assessing academic writing ability have already been foreshadowed in the earlier discussion of standards-based assessment, but let us review several considerations:

- It can be argued that this standards-based assessment represents the construct much more adequately by requiring the students to engage in a variety of preparatory activities similar to those involved in producing a researched essay before they present the final version for assessment. They also have the opportunity to resubmit if their first effort falls short of the evidence requirements.
- Teachers have access to detailed resource materials to supplement the unit standard specification and are expected to have their assessments moderated by colleagues within the school (NZQA, n.d.). They are able to choose a topic and readings which are suitable for the background and interests of their particular students.
- Nevertheless, from a national perspective, there will inevitably be variation across schools in the way that the assessment is designed and implemented, particularly in the amount of support individual teachers provide, as well as their judgements as to whether evidence requirements have been met.
- Although schools are required to have internal moderation procedures, moderation at the national level is based on a small sample of work selected by the school and is primarily concerned with whether the school's assessment decisions are nationally consistent.
- External moderation does not lead to any adjustment in the results for individual students and thus there is the potential for unfair treatment of individuals, even if the moderator finds issues with the school's assessments.

## **Conclusion**

From a consideration of these three approaches to the assessment of academic writing ability, it should be clear that there is no one best way of ensuring the quality of language assessments. Each approach aspires to high standards within its own terms, but there are clearly trade-offs involved. In a high-stakes international proficiency test, the time available for testing is limited and a maximum level of reliability is paramount, so that test users can have confidence in the consistency and security of the results. This means that the writing tasks in such tests tend to have reduced validity in the sense of adequately representing the construct which they set out to assess. With the two tests described here, there is also the contrast between the paper-based IELTS and the computer-based TOEFL, which throws up particular concerns for validity and fairness. Although the longer term trend is surely towards computer-based assessments, in the meantime there are pros and cons associated with the two modes of delivery.

The standards-based assessment for the NCEA and other NZQA qualifications is quite a different paradigm, and of course operates on a much smaller scale than an

international proficiency test. It offers the opportunity to embed the assessment into schools' teaching and learning activities, and the extended timeframe allows for a more valid simulation of the process of academic writing. Clearly, this form of writing assessment cannot reach the level of reliability achieved by the carefully monitored and even regimented scoring system employed for the iBT, but it can be argued that at its best SBA offers compensating advantages in terms of the quality of the students' preparation for the demands of tertiary writing and the richness of the information available to teachers about what individual learners are capable of.

It is useful to highlight the different conceptions of fairness involved in the two approaches to assessment. In standardised tests, the assumption is that the fairest way of assessing students for a high-stakes purpose is to ensure that, to the greatest extent possible, all members of a student cohort respond to exactly the same task under identical conditions, and that their teachers should be excluded from the process. By contrast, standards-based assessment puts more faith in the judgement of teachers to tailor the content and procedures to address the needs and interest of their learners, particularly those who might be disadvantaged by a standardised testing regime, while at the same time adhering to consistent national standards. As we saw previously, achieving consistency is an ongoing process of improvement on the part of teachers, which may involve some unfairness for individual learners as teachers get up to speed.

The attitude of New Zealand universities towards the EAP unit standards such as 22750 is somewhat ambivalent. Two of the standards are accepted as meeting the literacy requirement for University Entrance but on the other hand, university representatives have to date been reluctant to consider recognising NZCEL-4 as meeting their English language requirements, despite the fact that the unit standards it is composed of target a demonstrably higher level of academic proficiency than IELTS Band 6 or iBT 80, which are the typical scores needed by undergraduate university students. There is undoubted concern in the universities about the consistency of standards among the public and private providers offering the NZCEL-4 nationally, especially since NZQA was at the time of writing (towards the end of the first year of the NZCEL programme) still developing the consistency arrangements for the qualification (NZQA, 2014). In a high-stakes environment, whatever claims may be made for its validity, a new qualification must establish its credibility by delivering a consistent standard of assessment, which also meets the expectation of fairness to all the students involved.

## References

- Carr, N.T. (2014). Computer-automated scoring of written responses. In A.J. Kunnan (Ed.), *The companion to language assessment* (pp. 1063-1078). Chichester, UK: Wiley Blackwell.
- Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T., & McNamara, T. (1999). *Dictionary of language testing*. Cambridge: Cambridge University Press.

- ETS (2014). TOEFL iBT® test scores. Princeton, NJ: Educational Testing Service. Retrieved October, 30, 2014 from: <http://www.ets.org/toefl/ibt/scores/>
- Kunnan, A.J. (2014). Fairness and justice in language assessment. In A.J. Kunnan (Ed.), *The companion to language assessment* (pp. 1098-1114). Chichester, UK: Wiley Blackwell.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13, 241-256.
- NZQA (n.d.). *NCEA resource kit. Managing internal moderation*. Wellington: New Zealand Qualifications Authority. Retrieved October, 28, 2014 from: <http://www.nzqa.govt.nz/assets/qualifications-and-standards/qualifications/ncea/NCEA-Resource-kit/managing-internal-assessment.pdf>.
- NZQA (2014). Draft consistency arrangements for New Zealand Certificates in English Language (NZCEL) – September 2014. Wellington: New Zealand Qualifications Authority. Retrieved October 29, 2014, from: [www.nzqa.govt.nz/assets/qualifications-and-standards/qualifications/EL-quals-review/Draft-NZCEL-Consistency-Arrangements-August-2014.pdf](http://www.nzqa.govt.nz/assets/qualifications-and-standards/qualifications/EL-quals-review/Draft-NZCEL-Consistency-Arrangements-August-2014.pdf)
- TKI (n.d.). Standards-based assessment. Te Kete Ipurangi Assessment Online website. Retrieved October 31, 2014 from: <http://assessment.tki.org.nz/Using-evidence-for-learning/Concepts/Concept/Standards-based-assessment>
- Weigle, S.C. (2002). *Assessing writing*. Cambridge: Cambridge University Press.
- Xi, X. (2010). How do we go about investigating test fairness? *Language Testing*, 27, 147-170.
- Xi, X. (2012). Automated scoring and feedback systems: Where are we and where are we heading? *Language Testing*, 27, 291-300.

# CREATING AND TRIALLING SIX VERSIONS OF THE VOCABULARY SIZE TEST

Averil Coxhead, Paul Nation and Dalice Sim  
Victoria University of Wellington

## Abstract

*This paper reports on research in progress on the creation and trialling of six versions of a 20,000 version of the Vocabulary Size Test (VST) (Nation & Beglar, 2007). The VST is beginning to be used by language teachers in various contexts. Six 100 item versions of the VST were developed and trialled with 46 test-takers who sat all six versions of the test. The results indicated that there were two sets of three parallel versions with generally no statistical differences between each set. The equivalence of the tests was checked across a range of variables including first language, gender, status as a university student, age and level of education. The paper suggests limitations of the VST, some cautious implications, and further research.*

## Introduction

Knowing the vocabulary size of language learners is important for setting goals in a classroom programme (Nation & Webb, 2011) and possibly providing a diagnosis for learners who have problems with reading and writing (Nguyen & Nation, 2011, p. 87). The Vocabulary Size Test (VST) (Nation & Beglar, 2007) was designed to measure both first and second language learners' written receptive vocabulary size in English. The test measures knowledge of written word form, the form-meaning connection, and to a smaller degree concept knowledge at the item level. Analysed using Read and Chapelle's (2001) framework, the VST is a discrete, selective, relatively context-independent vocabulary test. At the test level, it provides a rough estimate of total vocabulary size where vocabulary knowledge is considered as including only single words (not multiword units) and vocabulary size does not include proper nouns, transparent compounds, marginal words like *um*, *er*, *gee gosh*, and abbreviations. The VST does not measure the ability to distinguish homonyms and homographs.

The original version of the VST tested up to the 14,000 level and was developed by Paul Nation. This means the test starts by testing words at the 1,000 frequency level, then the 2,000, then the 3,000 and so on up to 14,000. This version has 140 multiple-choice items, with 10 items from each 1000 word family level from the most frequent 14,000 word families of English. A learner's total score needs to be multiplied by 100 to get their total receptive vocabulary size (Nation & Beglar, 2007). Test-takers select the best definition of each word from four choices. Here is an example item:

16. strap: He broke the <strap>.
  - a. promise
  - b. top cover
  - c. shallow dish for food
  - d. strip of strong material

Beglar's (2010) examination of the 140 item VST showed it can be used with learners with a very wide range of proficiency levels and it clearly measures a single factor (written receptive vocabulary knowledge). He also found the test has a range of item difficulties related to the frequency level of the tested words. Beglar compared the performance of male participants with female participants, versions of the test with different numbers of items, and learners of various proficiency levels. Rasch reliability measures were around .96.

Nguyen and Nation (2011) showed that it is important to sit all levels of the test because some words at the lower frequency levels will be known. This may be because they are loan words or cognates, they relate to learners' hobbies and interests, they are technical words in fields the learners are familiar with, or the learners just happened to meet and learn them.

The issue of cognates is important in the VST. Removing the loanwords or cognates in the learner's first language from the test would distort the measurement of vocabulary size because they are a legitimate part of a learner's second language vocabulary size. Because cognates are so influential, it may be necessary to ensure that when the test is used with learners with the same first language, the proportion of cognates in the test reflects the proportion of cognates in the language (Elgort, 2013; see also Elgort & Coxhead, in press).

The 20,000 word family versions of the VST were developed to reduce the ceiling effects of the 14,000 level test and because the test should measure frequency levels beyond the test-takers' likely vocabulary size. The larger version means that the test could be used with adult native-speakers as well as high proficiency non-native speakers. For more on the vocabulary size testing research from Victoria University of Wellington, see Nation and Coxhead (2014).

### **The advantages of having parallel versions of a test**

One practical advantage of parallel versions is test security. It lessens the chance that a learner who has just sat the test can inform others who are yet to sit the test. It also reduces the effect of an earlier test on a later test. Having parallel versions also makes longitudinal research on vocabulary size and growth much more manageable because the same version of a test does not have to be used over and over again. Finally, comparing the same learners' results on parallel versions of a test is also a way of assessing the reliability of a test. Carmines and Zeller (1970, p. 40) note that,

The alternative-form method for assessing reliability is obviously superior to the simple retest method, primarily because it reduces the extent to which individuals' memory can inflate the reliability estimate.

The basic limitation of the alternative-form/parallel-form method is the practical difficulty of constructing alternative versions that are parallel. The Vocabulary Levels

Test (Nation, 1993; Schmitt, Schmitt & Clapham, 2001) is an example of a test which has parallel versions.

### **Criticisms of the VST**

The multiple-choice format provides opportunities for guessing, which might be done by elimination of choices. Because each item represents 200 word families, random guessing can inflate scores. The amount of random guessing will depend on the way the test is administered (one-on-one versus group administration), learners' attitudes to the test, and learners' vocabulary size (learners with larger vocabulary sizes have fewer items that they truly don't know). When interpreting the results of the test, it needs to be remembered that multiple-choice tests (recognition tests) give higher scores than translation or interview tests (recall tests) (Laufer & Goldstein, 2004). Recall tests tend to underestimate vocabulary knowledge, while recognition tests overestimate vocabulary knowledge.

The VST is designed to give credit for partial knowledge because the distractors are not closely related in meaning to the correct choice (Nagy, Herman, & Anderson, 1985). Partial knowledge may be sufficient to cope with a word and learn more about it, when it is met in context while reading. Another criticism of the test might be that it is based on word families and there is no guarantee that knowing a headword of a word family implies knowledge of the other words in a family. Keeping these criticisms in mind, this paper reports on an analysis of the six versions of the VST, through an analysis of the results from 46 participants.

### **Research questions**

1. To what extent can the six versions of the test be considered parallel or equivalent across the 46 individual test takers?
2. What effect might the test order have on the test results?
3. What effect might first language, gender, current university study, age, and level of education have on the results of the tests?

### **Methodology**

#### ***Developing six versions of the VST***

The words in the Vocabulary Size Test were sampled from word family lists originally created from data from the British National Corpus (Nation, 2006). Sampling from frequency-based word lists avoids the severe sampling biases that occur when sampling from dictionaries (Nation, 1993). Distractors were definitions of words chosen from the same 1000 word level as the tested word. Care was taken with the length of the options. Finally each test was run through the Range program (Heatley, Nation & Coxhead, 2002) to double check that distractors were the same level as their test items. This procedure was used for five versions (B-F). Version A consists of the original 14,000 version with six new levels added (15,000-20,000), using the procedure as for the other versions.



### ***Item sampling and test format***

Each test contains a total of 100 items, five from each of twenty 1,000 word level bands. The margin of error of a sample is primarily determined by the size of the sample not the sampling rate. For the VST, 100 items reliably represent the combined 20 bands because 100 items are a large enough sample to represent 20,000 words. However, the test cannot be used to see what proportion of words is known at each 1000 word family band because five items are not enough to represent a band. During initial trialling, some of these tests were combined into 200 item versions, but Beglar's (2010) findings and our own piloting showed that a 100 item test could be sat in a reasonably short period of time, and that the scores were consistent with 100 and 200 item tests (Coxhead, unreported data). Because the sampling rate from the BNC lists is one in 200, scores on the 100 item versions of the VST need to be multiplied by 200 to estimate total vocabulary size.

### ***Participants***

Almost all the 46 participants were university students who ranged in age from 16 years old to two people over 60, with most in their twenties. 28 were native speakers of English and the rest were high proficiency non-native speakers. There were 34 females and 12 males. The participants are a convenience sample. It was difficult to find test-takers who were willing and could spare the time (around three hours) to sit six tests.

### ***Administration of the tests***

The tests were administered on computer with a researcher present. Computer scoring ensured reliability of scoring. About one third of the test-takers sat all six tests at once with rest breaks, and the remainder sat three tests at a time in two sessions. Each test took between 20 to 40 minutes. The participants sat the tests individually, not in groups.

## **Results and discussion**

### ***Research question one: To what extent can the six versions of the test be considered parallel or equivalent across the 46 individual test takers?***

To decide if the six tests were equivalent (parallel versions), we needed to see if the mean scores and variances on the different versions were significantly different from each other, and how much an individual's score would differ when sitting two different versions. We compared each test to all other tests using the methods of Bland and Altman (1986). Here equivalence of two tests is assessed using the mean difference ( $\bar{d}$ ), and its standard deviation,  $s$ . If differences within  $\bar{d} \pm 2s$  would not lead to differences in interpretation of the result, then the two tests could be used interchangeably. Table 1 provides the descriptive statistics for the six tests, ordered by mean score.

Table 1:

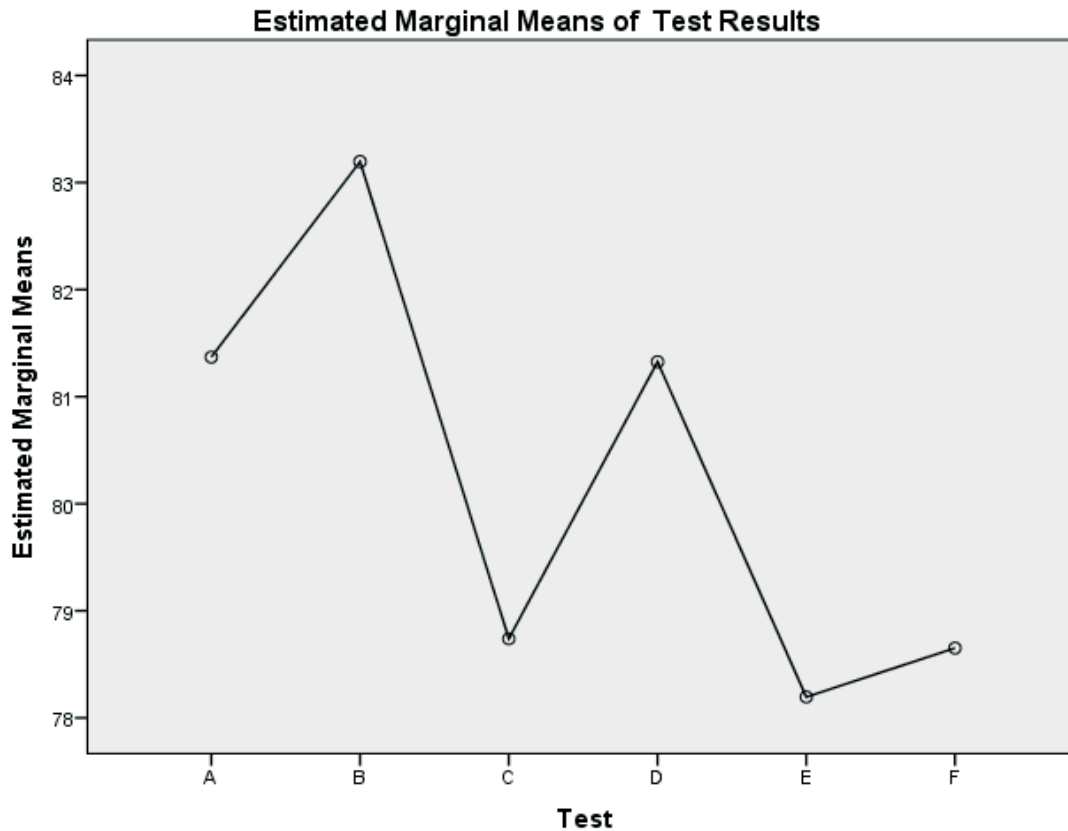
*Descriptive statistics for the six versions of the Vocabulary Size Test*

<b>Test version</b>	<b>Mean</b>	<b>Std. Deviation</b>	<b>N</b>
B	83.20	13.982	46
A	81.37	16.662	46
D	81.33	14.592	46
C	78.74	15.221	46
F	78.65	14.439	46
E	78.20	13.341	46

The first question was whether the test results (total number correct out of 100) varied significantly by test (A – F). All the participants took all 6 tests, so repeated measures analysis of variance was used to compare the mean results, while taking into account the correlated responses (that is, the same person's results on two different tests would be expected to be correlated). To compare different subgroups of participants (e.g., native and non-native speakers, see research question 3 below), the subgroup was used as a between group factor and test as a 'within' factor in the repeated measures design.

Since, by Mauchly's test, we cannot assume sphericity with these data, the results of the Greenhouse-Geisser statistic was used because it adjusts for lack of sphericity in the data to test the overall hypothesis that the results vary by test. We concluded ( $F(4.061, 182.759) = 14.573, p < 0.0005$ ) that the mean test results differed significantly by test.

We then used the Bonferroni correction for the pairwise comparisons to see which tests were different from which other tests. From these results, we can conclude that tests A, B and D have significantly ( $p < 0.05$ ) higher mean results than tests C, E and F. We cannot statistically differentiate between A, B and D, nor between C, E and F. Figure 1 is the plot of the predicted mean results. Note that the scale on the vertical axis covers a small range of scores.



*Figure 1:*  
Means of six versions of the Vocabulary Size Test

In measures of total vocabulary size, we would be satisfied if two or more tests placed the same person roughly within the same 1000 level band. If scores for the same person differed by less than 5 out of 100, that would be satisfactory equivalence. The highest mean score (Test B: 83.20) and the lowest mean score (Test E: 78.20) differ by five points, which is on the margins of being too far apart. Within each grouping of three tests, (A, B, D) (C, E, F) the mean differences between the highest and lowest are much smaller (1.87, and 0.54).

Table 2:  
*Mean differences and 95% confidence intervals for the six test versions*

Comparison	Mean Diff	SD Diff	Lo Limit 95%	Hi Limit 95%
A vs B	-1.8261	4.29110	-3.07	-.59
A vs C	2.6304	5.11798	1.15	4.11
A vs D	.0435	4.29965	-1.20	1.29
A vs E	3.1739	6.05400	1.42	4.92
A vs F	2.7174	5.98392	.99	4.45
B vs C	4.4565	4.88402	3.05	5.87
B vs D	1.8696	4.42020	.59	3.15

B vs E	5.0000	4.47710	3.71	6.29
B vs F	4.5435	5.82602	2.86	6.23
C vs D	-2.5870	4.84688	-3.99	-1.19
C vs E	.5435	4.34975	-.71	1.80
C vs F	.0870	4.95682	-1.35	1.52
D vs E	3.1304	5.22647	1.62	4.64
D vs F	2.6739	4.97127	1.24	4.11
E vs F	-.4565	5.44755	-2.03	1.12

From Table 2, we see that the 95% confidence intervals for the mean differences are no more than 4 points (in either direction) for comparisons between A, B and D. The same is true for comparisons between C, E and F. Comparing between these two groups of three test versions, the 95% confidence limits are almost all 4 (A vs C, A vs E, A vs F, D vs E, D vs F) or more (B vs C, B vs E, B vs F). For assessing vocabulary size, these data support the conclusion from the repeated measures analysis of variance that the six test versions fall into two groups of three equivalent versions: (A, B, D) and (C, E, F).

If two tests are equivalent, they give the same information about the test participants, so we would expect the tests to have the same or similar variances in the same sample of participants. This would indicate that both tests find the same spread of responses within the same sample. We did Fisher's test to compare variances between each pair of tests (Snedecor & Cochran, 1980, pp. 98-99). None of the variances was statistically different from any other, the smallest p-value being 0.0699 for the comparison of A with E. Therefore, comparing the variances of the six tests provides no evidence to support or refute the two sets of three parallel tests.

It is important to know if different tests give the same score or close to the same score for each of the individuals for each pair of comparisons (15 comparisons, A vs B, A vs C, A vs D, A vs E, A vs F, B vs C, and so on) between the six versions of the tests. Looking at this involved a total of  $46 \times 15 = 690$  comparisons. The aim was to see how many of the 690 comparisons were identical scores, or differed by 1 point out of 100, 2 points, 3 points and so on. The comparisons were done for each of the two groups of three tests, and for all six tests. If most of the comparisons were identical scores or within five or less points of each other, this would give us greater confidence in using the tests as parallel versions for looking at individuals.

Table 3:

*Percentage of participants with pairwise differences within 3 or 5 points for the parallel versions*

<b>Comparison</b>	<b>Percent 0</b>	<b>Percent &lt; 3</b>	<b>Percent &lt; 5</b>
A vs B	15.2	60.9	73.9
A vs D	10.9	67.4	80.4
B vs D	6.5	63.0	87.0

C vs E	2.2	56.5	73.9
C vs F	8.7	41.3	69.6
E vs F	6.5	52.2	71.7
Average	8.3	56.9	76.1

The percentages in each row in Table 3 are cumulative. Table 3 shows that 15.2% of the 46 test-takers got exactly the same score on Tests A and B, 60.9% got identical scores or scores differing by 3 or less on Tests A and B, and 73.9% of the test takers got scores differing by 5 or less. On average, just over 76% of the test-takers got scores within 5 points of each other on the parallel versions. However, approximately one quarter of the test-takers had scores differing by six points or more.

Table 3 also shows that tests A, B and D (the first three comparisons) are very consistent, with a high percentage of participants (73.9% - 87.0%) scoring within 5 points on these three tests. Tests C, E and F are less consistent (the second three comparisons) with 69.6% - 73.9% of participants scoring within 5 points on these three tests. Not all of these differences will be the fault of the tests themselves, because the differences can also come from the learners and the care with which they sat the tests. Test-retest data on exactly the same versions is needed to act as a comparison control for these variables.

Let's now look at pairs of tests between the two groups of ABD and CEF. So, A is compared with C for each person, A with E and so on (nine comparisons). See Table 4 below.

Table 4:

*Pairwise comparisons of differences in test scores for each test-taker expressed in percentages of test-takers for the non-parallel versions*

<b>Comparison</b>	<b>Percent 0</b>	<b>Percent &lt; 3</b>	<b>Percent &lt; 5</b>
A vs C	4.3	45.7	60.9
A vs E	4.3	34.8	54.3
A vs F	4.3	32.6	63.0
B vs C	4.3	45.7	58.7
B vs E	2.2	23.1	56.5
B vs F	2.2	39.1	58.7
D vs C	15.2	43.5	69.6
D vs E	8.7	41.3	54.3
D vs F	6.5	45.7	69.6
Average	5.8	39.1	60.6

If we compare the averages in the bottom row of Table 3 (the two sets of three parallel versions comparisons) with those in Table 4 (the nine non-parallel versions comparisons), we see a greater likelihood of closer scores when sitting two parallel versions. 60.6% of the test-takers had scores within five points or less of each other on

the non-parallel versions compared with 76.1% of the test-takers on the parallel versions.

This data supports the two groupings of parallel versions but shows very clearly that even parallel versions are unlikely to provide identical scores, and for a significant group of test-takers (around 20-30%) the scores on two tests taken by the same person are likely to be several points apart.

***Research question two: What effect might the test order have on the test results?***

The order of the tests was varied for different learners so there was no one set order for taking the six tests. However, scores on tests sat later could have benefited from the test-takers' experience of sitting the previous tests. Alternatively, tests sat later could have been affected by test-taking fatigue. The overall means and standard deviations of test results by order are shown in Table 5. In column 1 of Table 5, the number 1 refers to tests sat first, 2 to tests sat 2<sup>nd</sup> and so on. If order of sitting has an effect on the results, we would expect to see either a rise in mean scores as we move down Table 5 as a result of improvements in test-taking skill through practice, or a drop as a result of fatigue or declining commitment. There is no evidence of such changes.

Table 5:  
*Order, means and standard deviations*

Order	Mean (sd)	n
1	80.54 (15.49)	48
2	83.26 (13.66)	47
3	79.55 (15.20)	49
4	78.55 (14.21)	42
5	77.31 (15.89)	55
6	83.43 (12.55)	35
Total	80.25 (14.72)	276

The number of people in column 3 is sometimes higher than 46 because some tests were sat in 200 item versions, and they were counted as being sat simultaneously. So, two people sat 200 item versions as their first test and the two 100 tests in this 200 version were both counted as being sat first, raising the n from 46 to 48.

***Research question three: What effect might first language, gender, current university study, age, and level of education have on the results of the tests?***

We next wanted to see whether or not this difference between the tests was maintained in different subgroups of the participants. The subgroups were native speakers (n=28) and non-native speakers (n=18), gender (male n=12; female n=34), status as a university student (n=31) versus not studying (n=15), different age levels, highest level of education, and first language group (English - n=28; European - n=7; Other - n=11). Table 6 below contains the results divided into these subgroups. The order of

the tests in column 1 of Table 6 is based on the ranked mean scores from Table 1, so B was the test with the highest mean and E was the lowest.

Table 6:

*Means and standard deviations for the six tests comparing native speakers and non-native speakers, females and males, and status as a university student*

<b>Test</b>	<b>Native</b>	<b>Non-native</b>	<b>Female</b>	<b>Male</b>	<b>Studying</b>	<b>Not studying</b>
B	90.79	71.39	85.67	76.92	80.00	89.80
(sd)	(5.61)	(15.00)	(13.05)	(14.83)	(14.55)	(10.28)
A	89.93	68.06	84.79	72.69	77.35	89.67
(sd)	(6.90)	(18.73)	(15.18)	(17.68)	(17.30)	(11.90)
D	88.96	69.44	84.33	73.69	77.71	88.80
(sd)	(6.60)	(15.78)	(12.91)	(16.31)	(15.46)	(9.12)
C	86.64	66.44	81.27	72.31	74.68	87.13
(sd)	(7.29)	(16.31)	(14.87)	(14.72)	(15.88)	(9.61)
F	86.18	66.94	81.64	71.08	74.68	86.87
(sd)	(6.01)	(16.00)	(13.54)	(14.36)	(14.81)	(9.65)
E	84.96	67.67	80.55	72.23	74.48	85.87
(sd)	(6.39)	(14.61)	(13.19)	(12.42)	(13.84)	(8.26)

The table shows that scores within each group of three tests are very close to each other. Repeated measures ANOVAs showed no significant difference within each of the two groups of three. Mauchly's test of sphericity was significant for all comparisons, and so the Greenhouse-Geisser statistic was used, and showed that both groups in each comparison followed the same pattern in their test scores. For all the people tested, native speakers predictably scored higher than non-native speakers. Table 6 shows large standard deviations for the non-native speakers indicating a wide range of English proficiency levels. Females had larger vocabulary sizes than males and those not currently studying at university had higher scores than those who were studying at university. Because the people tested were not randomly chosen and are unlikely to be representative of the general population, not too much can be generalized from these scores.

Comparisons of the means on the six tests across age groups, level of education and first language generally supported the groupings of the two sets of three tests. Table 7 shows the effect of first language on performance on the test. The speakers of European languages which have cognate relations with English were in one group (n=7). This included speakers of German, Italian, Dutch and Afrikaans. Speakers of other languages without cognate relations to English (n=11) were in another group and included speakers of Indonesian, Malay, Vietnamese and Chinese.

Table 7:

*Effect of first language on mean scores and standard error of the six versions*

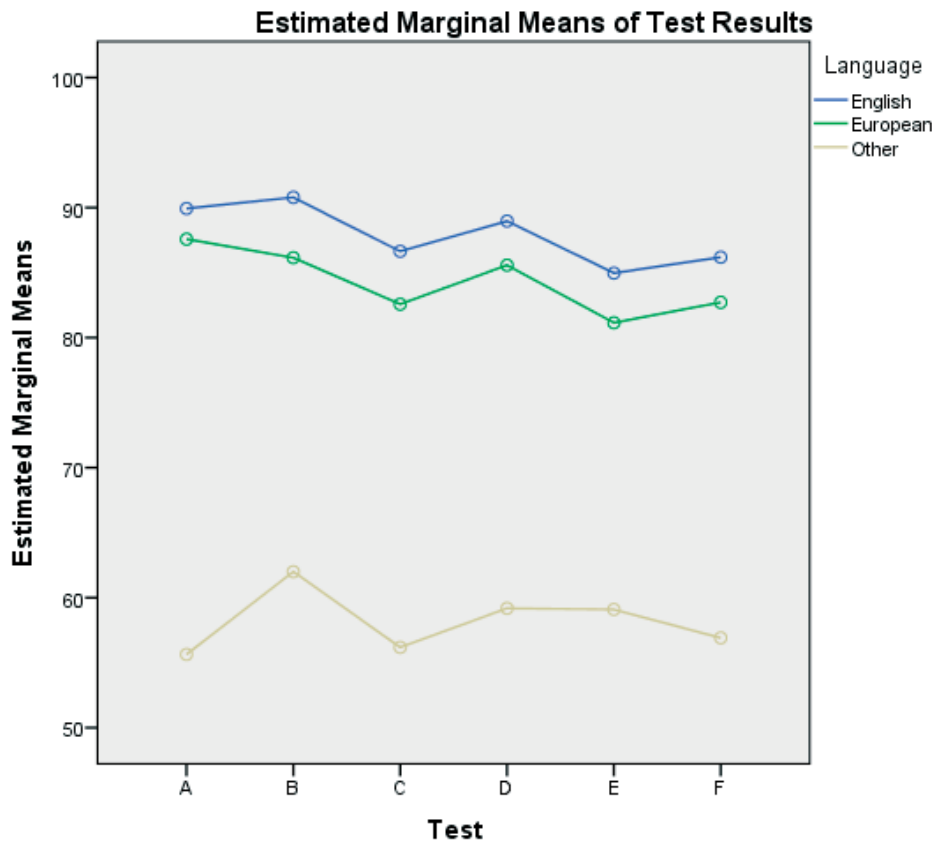
Test	English (n=28)	European (n=7)	Other (n=11)
	<b>Mean (sd)</b>	<b>Mean (sd)</b>	<b>Mean (sd)</b>
B	90.79 (1.35)	86.14 (2.69)	62.00 (2.15)
A	89.93 (1.55)	87.57 (3.10)	<b>55.64</b> (2.47)
D	88.96 (1.42)	85.57 (2.84)	59.18 (2.27)
C	86.64 (1.57)	82.57 (3.14)	56.18 (2.51)
F	86.18 (1.44)	82.71 (2.87)	56.91 (2.29)
E	84.96 (1.48)	81.14 (2.97)	<b>59.09</b> (2.37)

A ceiling effect is operating for both native-speakers (note the low standard errors) and speakers of European languages. Table 7 illustrates how close the first language speakers of European languages are to the scores of the native English speakers. The two groupings of three tests are once again maintained in the European language speakers' scores. Only the score for European A is out of order within that group. The group of speakers of other languages do not follow the previous patterns so well (see the bolded results in column 4 of Table 7). The score for A is lower than it should be, and the score for E is higher than it should be, perhaps a result of cognates or loan words.

The statistical tests from the repeated measures ANOVA verify these results. The correlations between the test results do not meet the criterion for sphericity, so the Greenhouse-Geisser adjustment was used. Overall, there was a significant difference between tests,  $F(3.952, 169.932) = 10.465$ ,  $p < 0.0005$ . The pattern of results across tests was different for the different language groups,  $F(7.904, 169.932) = 3.034$ ,  $p = 0.003$ . That is, the plots of results by test are not parallel for the three language groups. Averaged over all six tests, there was a difference in score by language group,  $F(2,43) = 70.143$ ,  $p < 0.0005$ . Multiple comparisons indicate that group 3 (Other) is significantly different from groups 1 and 2, which are statistically similar. Learners of English as a foreign language who are native speakers of a European language typically achieve much higher proficiency than native speakers of other languages.



The plots are:



*Figure 2:*

Means on the six versions of the Vocabulary Size Test for native speakers of English, speakers of European languages, and speakers of other languages

Those with higher education had higher scores (PhD 1, High school 1, degree 44) and, generally, older test-takers had higher scores than younger test-takers. However, some of the groupings had very small numbers, for example, only 2 people in the 61+ age group and 5 in the 16-20 age group, so we cannot depend on these findings.

Based on our data, we can conclude that tests A, B and D score consistently higher than tests C, E and F although this is on average with a difference of a few points out of 100. This pattern was repeated in all subgroups, except for the first language subgroups, suggesting that the differences seen between tests will be consistent across most subgroups. The choice of test to be used may, however, potentially advantage or disadvantage different groups of participants according to their first language, A and E being out of order (Table 4, column 4). Although the difference between the two sets of three tests is rather small, researchers using two or three versions should use the tests within one group of three.

To work out what a score on the tests means in terms of language use, we need to look at the vocabulary size needed to gain a text coverage of 98% in various kinds of texts. Table 8 provides such data.

Table 8:

*Vocabulary sizes needed to get 98% coverage (including proper nouns) of various kinds of texts (Nation, 2006)*

<b>Texts</b>	<b>98% coverage</b>	<b>Proper nouns</b>
Novels (Nation, 2006)	9,000 word families	1-2%
Newspapers (Nation, 2006)	8,000 word families	5-6%
Children's movies (Nation, 2006)	6,000 word families	1.5%
General Spoken English (Nation, 2006)	7,000 word families	1.3%
Spoken Academic English (Dang & Webb, 2014)	4,000 word families plus proper nouns and marginal words	0.37-1.69%
TED Talks (Coxhead & Walls, 2012)	9,000 word families plus proper nouns (1.44%)	1.44%

Note that Dang & Webb (2014) report 96.05% coverage over the British Academic Spoken English corpus at 4,000 word families plus proper nouns and marginal words, and 98.00% coverage at 8,000 word families plus proper nouns. Note that the range of proper nouns in this study differs across academic disciplines. Coxhead & Walls (2012) found the vocabulary load of 9,000 plus proper nouns over a corpus of TED Talks.

The goal of around 8,000-9,000 word families is an important one for learners who wish to deal with a range of unsimplified spoken and written texts. It is helpful to know how close learners are to this critical goal. Initial studies using the test indicate that undergraduate non-native speakers of non-European backgrounds successfully coping with study at an English speaking university have a vocabulary size around 5,000-6,000 word families. Non-native speaking PhD students have around a 9,000 word vocabulary.

### **Limitations**

The small number of participants is an important limitation. It is a matter of debate whether a cut-down version of the test, for example a 50 item test going up to the 10th 1000 words, is better for intermediate learners of English as a foreign language. Limiting the size of the test like this will have the negative effect of not allowing the learners to show knowledge of the low frequency words that they happen to know. The positive effects will be to reduce the time to sit the test, the elimination of a large number of items that learners do not know, and the subsequent reduction of the effect of random guessing.

Note that an item analysis for the new versions is needed to establish whether all of the items in the new versions are working correctly. These results will be reported in another paper.

### **Implications, future research, and conclusion**

This research suggests there are two sets within the six versions of the VST, and that we need to be cautious with releasing these versions until further validation work has been carried out. It is clear from the data that tests (B, A, D) have equivalent means and variances, and are more likely to provide roughly similar scores. C, F, and E can also have equivalent means, but they are not as “consistent” as the ABD grouping, in that a lower percent of participants had scores fewer than 5 points apart. All six tests give average scores within 5 points of each other which would roughly place test-takers within the same 1000 level band. In order to achieve greater reliability, it may be wise to follow Diack’s (1975) guidelines. That is, to get test-takers to sit more than one test and calculate the average.

Future work on the VST will include more validation research, investigating the VST alongside other measures of vocabulary development (Elgort & Coxhead, in press for more), and a VST for listening is under development by researchers in Japan. More work also needs to be done on the effect of test-taking strategies and the VST, to probe the problem of guessing with the multiple choice format (Elgort & Coxhead, in press).

Teachers need to be cautious with administering the test to large groups because of the need for engagement with the test. In this study, participants sat the tests in one-on-one conditions with a researcher, which ensured that the test-takers remained focused on the task. We also caution against using the larger versions of the test with beginner or even intermediate level learners of English, considering the possible motivational impact of encountering words in a test which even adult native speakers of the language might not know. Students (native and non-native speakers) might want to test their vocabulary independently or teachers might want to administer the test to classes at the beginning of a course to help set or modify learning goals. As more bilingual versions of the test become available, the VST will be able to be used to find out more about the vocabulary size of lower proficiency learners of different first languages. Finally, it is important to resist the urge to consider this test in any way a levels test for the BNC 1000 lists. Versions A and B are on Paul Nation’s website (<http://www.victoria.ac.nz/lals/about/staff/paul-nation>). Two versions will be kept in-house and the final two versions can be obtained for research purposes by contacting the authors.

### **Acknowledgements**

This research was supported by a grant from the Faculty of Humanities and Social Sciences, Victoria University of Wellington. Many thanks to Alex Heatley for his IT support, Brian King for test development, and our highly skilled research assistants. Finally, thank you to all the participants.

## References

- Beglar, D. (2010). A Rasch-based validation of the Vocabulary Size Test. *Language Testing*, 27(1), 101-118.
- Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical assessment. *Lancet*, i, 307-310.
- Carmines, E. G., & Zeller, R. A. (1970). Reliability and validity assessment. *Sage University Paper*, 17.
- Coxhead, A. & Walls, R. (2012). TED Talks, vocabulary, and listening for EAP. *TESOLANZ Journal*, 20: 55-67.
- Dang, Y. & Webb, S. (2014). The lexical profile of academic spoken English. *English for Specific Purposes*, 33: 66-76.
- Diack, H. (1975). *Test your own wordpower*. St. Albans: Paladin.
- Elgort, I. (2011). Deliberate learning and vocabulary acquisition in a second language. *Language Learning*, 61(2), 367-413.
- Elgort, I. (2013). Effects of L1 definitions and cognate status of test items on the Vocabulary Size Test. *Language Testing*, 30(2), 253-272.
- Elgort, I., & Coxhead, A. (in press). An introduction to the Vocabulary Size Test: Description, application and evaluation. In J. Fox & V. Aryadoust (Eds.) *Current trends in language testing in the Pacific Rim and the Middle East: Policies, analysis, and diagnosis*. Cambridge: Cambridge Scholars.
- Heatley, A., Nation, P., & Coxhead, A. (2001). The RANGE Programme. Available at <http://www.victoria.ac.nz/lals/about/staff/paul-nation>.
- Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, 54(3), 399-436.
- Nagy, W. E., Herman, P., & Anderson, R. C. (1985). Learning words from context. *Reading Research Quarterly*, 20(2), 233-253.
- Nation, I. S. P. (1993). Using dictionaries to estimate vocabulary size: Essential, but rarely followed, procedures. *Language Testing*, 10(1), 27-40.
- Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, 63(1), 59-82.
- Nation, P. (1983). Testing and teaching vocabulary. *Guidelines* 5, 12-25.
- Nation, P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, 31(7), 9-13.
- Nation, P. and Coxhead, A. (2014). Vocabulary size research at Victoria University of Wellington, New Zealand. *Language Teaching*, 47(3): 398 – 403.
- Nguyen, L. T. C., & Nation, I. S. P. (2011). A bilingual vocabulary size test of English for Vietnamese learners. *RELC Journal*, 42(1), 86-99.
- Read, J., & Chapelle, C. (2001). A framework for second language vocabulary assessment. *Language Testing*, 18(1), 3-32.
- Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, 18(1), 55-88.
- Snedecor, G. W., & Cochran, W. G. (1980). *Statistical Methods* (7th ed.). Ames, Iowa: Iowa State University Press.

# DEVELOPING SPEAKING FLUENCY WITH THE 4/3/2 TECHNIQUE: AN EXPLORATORY STUDY

John Macalister  
Victoria University of Wellington

## Abstract

*For over thirty years the 4/3/2 technique has been promoted as a means of developing speaking fluency in the language classroom, but it is a technique that has received surprisingly little research attention. Existing research focuses on the ESL context and quantifiable measures of fluency such as speed and hesitation phenomena. This paper adds to the existing literature by reporting on an exploratory, small-scale study in an EFL context that looks at the effects of the technique on learners' spoken production. It identifies changes in the quality of speech, specifically lexical repair and changes in the signalling of content and organisation. These findings suggest the 4/3/2 offers second language learners important opportunities to consolidate their existing linguistic knowledge.*

## Introduction

Fluency development in the language classroom has been receiving increased attention in recent years. This is particularly so in the receptive skills, with recent research in both EFL and ESL contexts into timed or speed reading (Chang, 2010; Chung & Nation, 2006; Macalister, 2008, 2010; Tran, 2012) and repeated reading (Chang, 2012; Gorsuch & Taguchi, 2008; Taguchi & Gorsuch, 2002; Taguchi, Takayasu-Maass, & Gorsuch, 2004), and developing fluent listening (Renandya & Farrell, 2011). The promotion of 'quick listens' (Millett, 2014) also adds listening to a daily fluency programme that already includes reading, writing, and speaking (Millett, 2008).

One catalyst for the renewed focus on activities that promote fluency development may be the four strands framework (Nation, 2007), that suggests a balanced language course will devote roughly equal time to each of four strands – meaning focused input, language focused learning, meaning focused output, and fluency development. For learning to occur in any strand, certain conditions need to be met and for the fluency development strand these are:

- the learners are largely familiar with all they are listening to, reading, speaking or writing
- the learners' focus is on receiving or conveying meaning
- there is some pressure or encouragement to perform at a faster than usual speed
- there is a large amount of input or output.

One activity that is promoted as developing fluency in speaking is the 4/3/2 technique (Nation, 2013, p. 37), in which students speak for reducing amounts of time (four minutes, then three, then two) on the same topic, each time to a different partner. The 4/3/2 meets the conditions for fluency development because:

- learners are using language and content that they already know to convey meaning to a listener;
- the reduction of time, and the encouragement to convey the same message in the time allowed, provide the challenge;
- and as each speaker speaks continuously for nine minutes, output is considerable.

Yet, despite its promotion as an effective activity, such promotion has for long been supported by surprisingly little research (Arevart & Nation, 1991; Nation, 1989). The earlier of these studies involved six advanced-level adult participants and found gains for fluency, as measured by words per minute and forms of hesitation; the later study included 20 intermediate-level adult participants in an ESL context and found similar fluency gains using the same measures. On the strength of these studies 4/3/2 has gained a reputation as a fluency development technique (incorporated into the fluency workshop described by Wood, 2001, for example), but one limitation of those studies is that they consider the effects within one iteration of the 4/3/2 only. It is perhaps not surprising that the priming effect of repeating the task should lead to short-term fluency gains. As de Jong and Perfetti (2011, p. 540) have pointed out, “when students repeat their speech, they do not have to generate content (semantic, grammatical, lexical), which frees up cognitive resources” that can be applied to other markers of fluency, such as speed and forms of hesitation. With an interest in investigating any long-term increase in fluency through the use of the 4/3/2 technique, de Jong and Perfetti (2011) reported on 24 students in an ESL setting, with some students speaking on the same topic in three 4/3/2 sessions over a two week period, and others on different topics. Long-term fluency gains as measured on a two minute monologue in pre- and post-tests were recorded for those who repeated their topics in 4/3/2, not for those who spoke on different topics. One difference between this experimental study and earlier studies was that participants spoke to and were recorded by a computer, rather than to another student. It is perhaps worth noting with respect to this study that the lack of a human audience may have reduced the learner’s focus on conveying meaning, one of the conditions Nation (2007) has proposed as necessary for fluency development.

De Jong and Perfetti (2011) are among a new wave of researchers looking at support for the use of 4/3/2 as a means for developing fluency in speaking. Boers (2014) also offers a reappraisal of the activity, comparing the performance of ten adult ESL learners in two different conditions, the 4/3/2 and a constant time condition where each of the three talks is for three minutes. He found improvements for fluency, as measured by words per minute, but not for accuracy or complexity of speech.

In sum, research with language learners in ESL settings (although de Jong and Perfetti (2011, p. 563) argue their “results are expected to be generalizable to nonimmersion classroom settings”) shows short-term fluency gains within one iteration of the 4/3/2 technique, but no long-term fluency gains on a new topic unless learners have

repeatedly spoken on a single topic beforehand. Further, this existing work on fluency development has focused on quantitative indicators. As de Jong and Perfetti (2011, p. 563) have mentioned, there remains a need for “deeper, qualitative analysis of the students’ production”. This could include a focus on vocabulary for, as Boers (2014, p. 3) points out, “Potential effects at the level of lexis...have yet to be examined”.

The purpose of this paper, then, is to contribute to the growing body of research on oral fluency development by, first, investigating the use of the 4/3/2 technique in an EFL rather than an ESL context; the distinction between EFL and ESL contexts is relevant because long-term fluency gains are less likely to be influenced by language use opportunities beyond the classroom in the former than the latter, and to date the transferability of current research to EFL contexts has only been assumed. It also contributes to the growing body of research by focusing on the quality of speech production as well as quantifiable measures of speaking fluency in order to address the following research questions:

- How does speaking fluency develop from turn to turn within one iteration?
- How does speaking fluency develop over multiple iterations of the 4-3-2 technique?

### **What is fluency?**

In language teaching, fluency is often regarded in general, qualitative terms and has been “loosely described as the ability to produce and process the target language with native speaker-like ease” (Macalister, 2008, p. 23). For language teachers, as Chambers noted (1997, p. 537), fluency is usually only considered in relation to speech and “restricted to speech flow and speech rate”. This view of fluency is reflected in the public version of IELTS assessment criteria where hesitation phenomena and speed are among the identified characteristics; at band 3 a speaker “speaks with long pauses”, at band 5 “uses repetition, self-correction and /or slow speech to keep going”, while at band 7 “may demonstrate language-related hesitation at times, or some repetition and/or self-correction” (IELTS, 2012, p. 18).

In language teaching, then, the emphasis tends to be on perceived fluency – that is, the subjective impression of the listener, who may be a rater or judge. Segalowitz (2010) has proposed two other forms of fluency, utterance and cognitive. Utterance fluency can be objectively measured from temporal aspects which can in turn be sub-divided into breakdown fluency (e.g. pauses), speed fluency (e.g. words per minute), and repair fluency (e.g. corrections). Cognitive fluency, on the other hand, is speaker-internal; it is determined by the extent to which the speaker has developed procedural knowledge of the target language. Wood (2001, p. 579, following Schmidt (1992)) discusses fluency in terms of automatic and controlled processing and usefully summarises it in a table reproduced below.

Table 1:  
*Automatic versus controlled processing*

<b>Automatic processing</b>	<b>Controlled processing</b>
fast and efficient	slow and inefficient
effortless	requires effort
not limited by short-term memory	limited by short-term memory capacity
not under voluntary control	under subject control
inflexible	flexible
inaccessible to introspection	at least partly accessible to introspection

Recent work (e.g. de Jong, Steinel, Florijn, Schoonen, & Hulstijn, 2012) has sought to tease out the relationship between aspects of fluency, and in particular the aspects of utterance fluency that act as indicators of L2 cognitive fluency. This work has also taken into account the fact that the use of pauses is related to individual speaking style and shown that pauses may not be correlated strongly to cognitive fluency (although hesitation phenomena have been claimed to reduce in the 4/3/2 technique; see Arevart & Nation, 1991). It is perhaps worth pointing out at this juncture that the present study is focused on the second language learning opportunities afforded by the 4/3/2 technique rather than probing the nature of fluency.

## **Methodology**

### ***Participants and setting***

The research was conducted over a two-week period at the Royal University of Phnom Penh, Cambodia, with volunteer participants in a speaking elective. At the university, electives are only open to students who have reached an intermediate level of English, as measured either by placement tests administered at the beginning of a student's course of study or by having successfully completed 300-level English courses taught at the university. The participants were, therefore, among the most proficient English speakers in the student population.

Although sixteen students signed consent forms, as required as part of the ethics approval granted for this project, to participate in the study, the number of participants was restricted to six because only three digital voice recorders were available. The six participants were chosen by ballot. Issues with attendance and recording quality meant that complete data for only four of the six participants is reported in the Findings section.

### ***Procedure***

Engagement with the class for the purposes of this study unfolded in the following way. On the first encounter I was present as an observer in order to gain some



understanding of the class dynamic, and to be introduced to the class as a researcher. Field notes record that there had been a “range of activities and interaction patterns, all having opportunities to speak either in pairs, as individual to whole class, or in small groups” and that “I enjoyed the class”. In other words, this was not a case of current classroom practice being viewed through a deficit lens. The following day I taught the class using the 4/3/2 activity and followed this activity with an opportunity for the students to reflect on the experience. The comments were generally positive although observation had suggested that at least one student had struggled to speak. Following this discussion I explained that I would be teaching the class on subsequent occasions, and invited student participation in determining the content focus. The students indicated that they would like to speak on familiar rather than unfamiliar topics in the next lessons, and made suggestions. The agreed topics, and their order, were Cambodian culture, sport, and education.

The structure of the three data collection lessons followed the same pattern of input followed by modelling, then preparation leading to performance. At the initial input stage a key word or phrase was written on the board, and students discussed this in pairs. The whole class then constructed a mind map of ideas on the board. Students then read a short text related to the day’s topic before discussing in pairs, focussing on new ideas or anything that was not understood. Students then reported back, and had the opportunity to add to the mind map that remained on the board. The main purpose of these related steps was to activate background knowledge. Following this, I modelled a four minute speaking turn and then students had five minutes to think about and prepare the content for their own speaking turn. In this preparation the emphasis was on generating ideas in note form, not speech writing. The emphasis was also on talking about personal experience in relation to the topic. Thus, for instance, on the sport topic one male speaker spoke exclusively about football while one female speaker talked about tennis, dancing at weddings, and being hit in the head by a football. Once this preparation stage was completed, performance of the 4/3/2 occurred.

### *Analysis*

The authentic classroom setting did have an effect on the quality of the recordings, particularly in the amount of background noise (e.g. other speakers, overhead fans) captured. However, speakers were clearly audible and subsequent transcription was performed manually, at times using the slow playback function to check what had been said. A research assistant subsequently verified the accuracy of the transcriptions and identified instances of uncertainty, which were again listened to and a consensus reached.

In the analysis, reported in the next section, the focus was on aspects of utterance fluency (Segalowitz, 2010) – speed, breakdown, and repair. In the analysis, speed was measured as syllables per second, determined by the total number of syllables other than filled pauses uttered by the speaker divided by the total time available. Both

filled and unfilled pauses were considered as measures of breakdown fluency. Filled pauses (*ah*, *er*, *um*, as examples) were counted and calculated as number per minute. Unfilled pauses of one second or more were also counted and calculated as seconds per minute. Changes in the quality of speech, or accuracy, were investigated by identifying language-related episodes, which typically were moments in which a speaker “simply solved [a language problem] (again, either correctly or incorrectly) without having explicitly identified it as a problem” (Swain & Lapkin, 1995, p. 378). All such episodes were treated as instances of repair fluency although not all occurred immediately; they may have occurred in another of the following talks within the 4/3/2 activity.

It is perhaps worth noting that the approach taken here was more inductive than deductive. While a typical analysis of speech quality might look at pre-determined accuracy phenomena such as error-free clauses and correct verb forms per minute, or complexity features relating to syntax and lexical sophistication, the approach here was to focus initially on corrections or improvements made by the speakers, then to analyse them.

## Findings

This section begins with quantitative data relating to speed and breakdown fluency, and then considers evidence for repair fluency. Tables 2, 3, and 4 show results for rate (syllables per second), filled pauses (number per minute), and unfilled pauses (seconds per minute) respectively for each of the four speakers in the three iterations of the 4/3/2. Each table is organised in the same way; columns show speakers and rows show the speaking turn. The first two columns (M1, F2) are for the speakers who appeared to be more fluent, i.e. they achieved higher speed fluency and generally had fewer instances of breakdown fluency, than the other two speakers. In the rows, the first number refers to the iteration of the 4/3/2 activity and the second to the speaking turn within that iteration; thus, 2.2 is the second speaking turn of three minutes in the second iteration. In order to differentiate between iterations, each column has been presented with a different text alignment (i.e. left, centre, right) for each iteration.

Table 2:

*Speed fluency (syllables per second)*

Iteration/turn	M1	F2	F1	M2
<b>1.1</b>	2.63	3.05	2.21	1.97
<b>1.2</b>	3.07	3.33	2.44	2.1
<b>1.3</b>	3.04	3.33	2.33	2.57
<b>2.1</b>	2.4	2.92	2.39	1.46
<b>2.2</b>	2.62	3.11	2.33	2.01
<b>2.3</b>	2.86	3.42	2.65	2.15
<b>3.1</b>	2.45	2.77	2.99	2.08
<b>3.2</b>	3.07	3.06	2.96	2.51
<b>3.3</b>	<b>3.33</b>	<b>3.56</b>	<b>3.01</b>	<b>2.67</b>

Table 3:

*Breakdown fluency (number of filled pauses per minute)*

Iteration/turn	M1	F2	F1	M2
<b>1.1</b>	2.25	0.5	10	9
<b>1.2</b>	1.67	0.3	7.3	8.3
<b>1.3</b>	3.5	1	8	9.5
<b>2.1</b>	4.25	1.5	9	11.5
<b>2.2</b>	3.67	2	7	9
<b>2.3</b>	3	0	6.5	11.5
<b>3.1</b>	3.5	2.5	9.25	6.75
<b>3.2</b>	2.67	2	9.17	7
<b>3.3</b>	3.5	1	5.5	7.5

Table 4:

*Breakdown fluency (unfilled pauses in seconds per minute)*

Iteration/turn	M1	F2	F1	M2
<b>1.1</b>	0.5	0	1.5	0
<b>1.2</b>	0	0	2.3	0
<b>1.3</b>	0	0	5	0
<b>2.1</b>	3.5	0	0.5	2.75
<b>2.2</b>	0	0	4	1.3
<b>2.3</b>	0	0	2	6
<b>3.1</b>	6.5	0	2.25	3
<b>3.2</b>	0	0	1	0.67
<b>3.3</b>	0	0	0	0

In terms of speed fluency (Table 2) the same pattern is observed across the three turns in each iteration – rate is faster on the third turn than the first, although faster on the second turn than the third for three of the speakers on the first iteration. Of particular interest is the fact that the fastest rate (in bold) for each speaker was the third turn on the third iteration.

For breakdown fluency, however, the picture is more variable. Two speakers (F1, M2) have a large number of filled pauses per minute (Table 3), and their higher breakdown fluency does appear to relate to lower speed fluency (Table 2). Only one speaker (F1) shows a reduction of filled pauses within each iteration of the 4/3/2, although only one (M2) shows no reduction from turn one to turn three in any of the three iterations. Similar variability is found for unfilled pauses (Table 4). Two points in relation to this table deserve comment. First, the relatively large amount of unfilled pausing for M1 at turns 2.1 and 3.1 is largely explained by the student not speaking for the full time. Second, one feature of F1's speaking that is not captured in Table 4 is moments of laughter, as in “and I want to cry [laughs] because I'm very hurt” from turn 2.1.

As well as quantitative indicators of change in performance, the transcripts also

revealed examples of changes in the quality of speech. These changes related to two broad areas – lexical repair (which here includes both form-based and lexis-based language related episodes), and organisation and clarification. In terms of lexical repair, Extracts 1 to 3 show three different patterns; in the first the speaker generates an immediate on-line repair such as the non-word *lateenth* being replaced by *nineteenth*. In the second pattern the speaker uses an incorrect word form on the second turn, but the correct form on the final turn. There is, however, no opportunity for retention to be demonstrated, unlike Extract 3 which is potentially the most interesting. Here the speaker has used the incorrect form of the word (*classify*) on the first turn, retained the meaning and introduced a more complex and accurate word on the second turn (*classification*), and retained that repair on the third turn.

Extract 1:

*One off lexical repair*

---

F2 – 2.2 *stay/live*

- that if we want to stay longer to live longer in this world

F2 – 2.3 *lateenth/nineteenth*

- in the lateenth century nineteenth century
- 

Extract 2:

*Delayed lexical repair*

---

**M3 – Buddhism/Buddhist**

- 1.2: all the people in Cambodia believe the Buddhism religion
  - 1.3: the people in Cambodia believe to the Buddhist the Buddhist religion
- 

Extract 3:

*Retained lexical repair*

---

**M1 – classify/classification**

- 3.1: [and] I achieve a high score in all my subjects and UH classify as a good student [and]
  - 3.2: I achieve a high score and never remove from the smart student classification [and] ... [and] I'm still in classification of smart student but I'm not so good student [because]
  - 3.3: [and] I never remove from the AH sched- UH the smart student classification and but I'm not so good [because]
-

The second broad area of change relates to organisation and clarification of content and is illustrated in Extracts 4 and 5. In the first example, the speaker makes changes to signal the structure of the talk by introducing the phrase *I will tell you step by step* in the second turn, which is then retained in the third turn, similar to the example in Extract 3. Like Extract 4, Extract 5 comes from the introductory moves and shows the speaker using increasingly sophisticated language to signal the content of her talk, from *about womans* in the first turn to *especially related to the woman* in the third turn. Other aspects of the improvement demonstrated in Extract 5 are two instances of retained lexical repair, i.e. the change from *womans* to *woman* from 3.1 to 3.2, and the retention of *I just want to talk about* in 3.2 and 3.3 after the immediate on-line repair in 3.1. In both Extracts 4 and 5, improvements in the quality of speech are evident through a reduction of errors.

Extract 4:

*Signalling structure*

---

**M2**

- 3.1: Hello UH today we will talk about the education and the future yes I have my own experience about education first
  - 3.2: Yes hello AH today we will we will talk about AH education and future yeah I have my own experience about my education **I will tell you step by step** when I about my education
  - 3.3: Yeah hello today we will talk about AH education and the and future yeah I have own my own experience **I will tell you step by step** about my education yeah
- 

Extract 5:

*Signalling content*

---

**F2**

- 3.1: See you again yeah I want just want to talk about the education in the future about womans
  - 3.2: Hello yeah today I just want to talk about the education in the future and if the **the important thing is related to** the woman
  - 3.3: Hi yeah nice to see you today I just want to talk about the education in the future **especially related to** the woman
- 

## Discussion

The first research question focused on speaking fluency development from turn to turn within one iteration of the 4/3/2 technique. The technique has been promoted as a means to develop speaking fluency for over thirty years and the findings from this

small-scale investigation in an EFL context would suggest that quantifiable aspects of speech, particularly speed fluency, improve from turn to turn within an iteration of the 4/3/2. These results regarding short term fluency gains are not surprising. They confirm findings from earlier studies (Arevart & Nation, 1991; Boers, 2014; de Jong & Perfetti, 2011; Nation, 1989), but add to the literature regarding the use of this technique as these findings involve language learner participants in an EFL rather than an ESL context. They also reinforce doubts raised about pauses as a measurement of fluency, given the lack of any clear pattern of reduction in Tables 3 and 4. Pauses may be a feature of an individual's speech style, and thus not a reliable indicator of cognitive fluency, even if they affect judgements of perceived and utterance fluency.

As well as measurable dimensions of fluency, the data also provided indications of improvements in the quality of speech from turn to turn within one iteration. As Extracts 1 to 5 illustrated, there was evidence of repair fluency among these speakers. Particularly interesting are the instances of self-generated delayed repair, both lexical (Extracts 2 and 3) and in terms of signalling (Extracts 4 and 5). These raise the question of the impetus for the change. It is not possible to determine whether the change was a result of noticing, or whether it was achieved with a lack of awareness or attention; in either case, however, it was the 4/3/2 format that provided an opportunity for the change to occur. Thus, especially where there is evidence of retention, these instances suggest a language learning opportunity. Indeed, echoing Swain and Lapkin's suggestion, "What goes on between the first output and the second output ... is part of the process of second language learning" (1995, p. 386).

However, one shortcoming of the 4/3/2 is that the two minutes of the third turn was sometimes not enough to allow retention to occur. This can be seen in the following example where the speaker introduces new information in the second turn of the third iteration, and performs a one-off lexical repair (cf. Extract 1), from *over* to *skip*. On the third turn, however, the time ended before he had reached this point in his story, so he did not have the opportunity to retain the repair which may, in turn, have minimised the learning potential of this repair. This was also the case for the first example in Extract 1.

Extract 6:

*Lost opportunity*

---

**M3**

3.2 ... so I over grade eight I over I skip grade eight

---

The more interesting question that this study sought to investigate, however, is whether repeated use of 4/3/2 leads to improved fluency on later iterations. It is tempting to suggest that the fact that all four participants achieved their fastest speed fluency (Table 2) on the third turn of the third iteration indicates the cumulative benefit for fluency development from repeated use of 4/3/2 on at least one measure, but this would be over-claiming from the results. It may simply be that participants

were more familiar with that day's topic, providing faster access to lexical and semantic content. Perhaps all that can be said as regards this small-scale investigation is that the demonstration of cumulative benefits for fluency development measures (rate, filled and unfilled pauses) in repeated iterations of the technique is inconclusive.

Given the change in topic each time, meaning that there were likely to be few if any opportunities to re-use topic-related language across different iterations of the 4/3/2, the most obvious area in which to observe improvement in the quality of speech from one iteration to the next might be expected to be in signalling (Extracts 4 and 5). There is, for example, some suggestion that speakers are settling upon fixed expressions as a way of signalling content and this does provide some support for cumulative improvement. Speaker M2, for example, began in the first iteration with "we are talking about" before using "today we will talk about" throughout the second and third iterations. Similarly, speaker F1 settles on "I want to tell you about" in the second and third iterations after beginning with "today we study about" in her first, and M1 moves from "I'll tell you something what I have to say about" in the first to "I would like to tell you" throughout the second and third iterations.

As a final comment on the data, it is worth noting that learners seem to treat the 4/3/2 seriously as a communicative act. Extract 1, for example, clearly suggests on-line attention to meaning; similarly, the laughter that punctuated F1's speech in the second and third iterations (Table 4) would also suggest she is conscious of what she is saying. This is important to note in the context of the conditions for fluency development – learners' focussing on conveying meaning (Nation, 2007). At the same time, however, the repeated turns mean that learners draw on their existing linguistic resources to improve the quality of the message. It may be that this is an opportunity to consolidate existing knowledge (Swain & Lapkin, 1995, p. 384), in which case it is similar to another teaching activity, the dictogloss (Wajnryb, 1990), of which one teacher-practitioner has remarked, it "provides learners an opportunity to apply their understanding of grammar in the task of text creation" (Dung, 2002, p. 29).

### **Pedagogical and research implications**

As was noted earlier, this study was focused on the second language learning opportunities afforded by the 4/3/2 technique. The technique is designed to provide opportunities to develop speaking fluency, and it should typically be seen as forming one part of an integrated unit of work or activity cycle. In such a cycle, learners will have gradually developed familiarity with the language and ideas of a topic through a series of activities that will have provided opportunities for meaning focused input, language focused learning, and meaning focused output – in other words, the 4/3/2 technique is one way of incorporating the fluency development strand in the four strands framework (Nation, 2007). This, it should be noted, is slightly different from the way in which the 4/3/2 was carried out here, although the choice of familiar topics and the preparatory work carried out each time were intended to mimic such a context.

One issue with the elegant division of speaking turns into four, then three, then two minute blocks is that it may not allow sufficient time in the final turn for learners to reproduce an earlier repair, as discussed above. Furthermore, an impression gained from listening to the recordings was that participants seemed to lose control of their material in the final turn; in other words, the pressure to perform faster may have come at the expense of accuracy, an observation also made by Boers (2014). As a result, it may be worth allowing more than two minutes for the final turn. A worthwhile research project could compare the quality of speech in two minute and two and a half minute final turns.

A concern for accuracy leads to a related issue. As can be seen in Extract 2, the speaker introduced an error into his talk (*believe to*) at the same time as using the correct word form *Buddhist*. On this occasion the error occurred in just one turn of that iteration of the 4/3/2, but repeated talks do allow the possibility of “consolidating (or ‘fossilizing’) erroneous word strings and patterns in the learners’ interlanguage” (Boers, 2014). Boers discusses a number of possible interventions or modifications that could be made to the 4/3/2, and the effect of these could also be investigated in future research. However, it is worth emphasising that when used for fluency development within the four strands framework there should already have been time for relevant language-focused learning prior to the 4/3/2.

A third area for further research is whether fluency development is retained in repeated iterations of the 4/3/2 and whether improved speaking fluency on the 4/3/2 transfers to other speaking tasks. This would extend the work already done by de Jong and Perfetti (2011) on this issue and is similar to questions that have been asked about reading fluency development activities such as speed or timed reading courses, where transfer effects have now been found (Macalister, 2008, 2010; Tran, 2012). It is all very well to improve performance on a controlled task, but that task is a means to an end, not the end in itself.

This question does, however, again call into consideration the purpose of the 4/3/2 technique. It could be argued that, within the four strands framework at least, its purpose is primarily to develop fluency with the language and ideas of a defined content area – Wood’s automatic processing (Table 1) – rather than to address speaking proficiency globally. If consolidation of content- or thematically-related language learning is the goal, transfer effects may be of secondary importance.

A final suggestion for further research could be an investigation of whether learner perceptions of themselves as language users change as a result of repeated iterations of the 4/3/2. While Boers (2014) has provided some learner comments after a single experience of the technique, these do not relate to learners’ L2 selves and their sense of self-efficacy.



## Limitations

The obvious limitation on the findings of any small-scale study such as this is that of its small size. While it is reassuring that the results particularly in terms of speed fluency are consistent with those from earlier studies, they do represent the performance of only four participants. Thus this study should be regarded as exploratory in nature.

The quality of the recordings also limited what was able to be measured. Better quality recordings may have enabled more sophisticated analyses of the participants' performance, such as the mean run between pauses. This should be considered for any future study, although it is likely that data gathered from an intact class in a language learning context such as this in the future will benefit from the constant improvements in recording technology.

Any future study would also benefit from consideration of a post-test, such as an opportunity to repeat speaking on an earlier topic (as in de Jong & Perfetti, 2011). This would provide another way of investigating speaking fluency development over repeated iterations of the 4/3/2 technique.

## Conclusion

Despite the limitations mentioned above, this study does provide fresh support for the use of the 4/3/2 activity for speaking fluency development. Furthermore, it does so in an EFL environment rather than the ESL setting found in earlier studies. The most obvious improvement is in speed fluency within one iteration of a 4/3/2. However, it is also an activity that provides multiple opportunities for second language learning as shown by the evidence of lexical repair and improvement in signalling structure and content in this study. These improvements in the quality of speech performance through use of the 4/3/2 have not previously been identified. Given these benefits, the 4/3/2 activity deserves a place in the language teaching classroom. At the same time, however, it is unclear to what extent performance on one iteration of the 4/3/2 transfers to subsequent iterations, and whether performance on the 4/3/2 transfers to other speaking tasks. These remain areas for future investigation.

## References

- Arevart, S., & Nation, P. (1991). Fluency improvement in a second language. *RELC Journal*, 22(1), 84 - 94.
- Boers, F. (2014). A reappraisal of the 4/3/2 activity. *RELC Journal* 10.1177/0033688214546964
- Chambers, F. (1997). What do we mean by fluency? *System*, 25(4), 535-544.
- Chang, A. C-S. (2010). The effect of a timed reading activity on EFL learners: Speed, comprehension, and perceptions. *Reading in a Foreign Language*, 22(2), 284 - 303.

- Chang, A. C-S. (2012). Improving reading rate activities for EFL students: Timed reading and repeated oral reading. *Reading in a Foreign Language*, 24(1), 56 - 83.
- Chung, M., & Nation, P. (2006). The effect of a speed reading course. *English Teaching* 61(4), 181-204.
- de Jong, N., & Perfetti, C. A. (2011). Fluency training in the EFL classroom: An experimental study of fluency development and proceduralization. *Language Learning*, 61(2), 533-568. 10.1111/j.1467-9922.2010.00620.x
- de Jong, N. H., Steinel, M. P., Florijn, A., Schoonen, R., & Hulstijn, J. H. (2012). Linguistic skills and speaking fluency in a second language. *Applied Psycholinguistics, FirstView*, 1-24. doi:10.1017/S0142716412000069
- Dung, K. N. (2002). Dictogloss: Using dictation to teach grammar. *Guidelines*, 24(1), 29-32.
- Gorsuch, G. J., & Taguchi, E. (2008). Repeated reading for developing reading fluency and reading comprehension: The case of EFL learners in Vietnam. *System*, 36(2), 253 - 278.
- IELTS. (2012). *Guide for teachers*. Manchester: IELTS.
- Macalister, J. (2008). The effect of a speed reading course in an English as a Second Language environment. *The TESOLANZ Journal*, 16, 23 - 33.
- Macalister, J. (2010). Speed reading courses and their effect on reading authentic texts: A preliminary investigation. *Reading in a Foreign Language*, 22(1), 104 - 116.
- Millett, S. (2008). A daily fluency programme. *Modern English Teacher*, 17(2), 21 - 28.
- Millett, S. (2014). Quicklistens: Using what they already know. *Modern English Teacher*, 23(4), 64-65.
- Nation, I. S. P. (2013). *What should every EFL teacher know?* Seoul: Compass Publishing.
- Nation, P. (1989). Improving speaking fluency. *System*, 17(3), 377-384.
- Nation, P. (2007). The four strands. *Innovation in Language Learning and Teaching*, 1(1), 1-12.
- Renandya, W. A., & Farrell, T. S. C. (2011). 'Teacher, the tape is too fast!' Extensive listening in ELT. *ELT Journal*, 65(1), 52-59. 10.1093/elt/ccq015
- Segalowitz, N. (2010). *Cognitive bases of second language fluency*. New York: Routledge.
- Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics*, 16(3), 371-391.
- Taguchi, E., & Gorsuch, G. J. (2002). Transfer effects of repeated EFL reading on reading new passages: A preliminary investigation. *Reading in a Foreign Language*, 14(1), 43 - 65.
- Taguchi, E., Takayasu-Maass, M., & Gorsuch, G. J. (2004). Developing reading fluency in EFL: How assisted repeated reading and extensive reading affect fluency development. *Reading in a Foreign Language*, 16(2), 70-96.

- Tran, Y. T. N. (2012). The effects of a speed reading course and speed transfer to other types of texts. *RELC Journal*, 43(1), 23-37. 10.1177/0033688212439996
- Wajnryb, R. (1990). *Grammar dictation*. Oxford: Oxford University Press.
- Wood, D. (2001). In search of fluency: What is it and how can we teach it? *The Canadian Modern Language Review*, 57(4), 573-589.

# TEACHING AND LEARNING ACADEMIC WRITING: NARRATIVES OF FUTURE DESTINATION

Martin Andrew and Zina Romova

Victoria University, Melbourne and Unitec Institute of Technology, Auckland

## Abstract

*This study contributes to an ongoing project on academic writing portfolios and relates their contents and forms to student destinations and imagined communities. Tertiary writing programs such as English for Special Purposes (ESP) and English for Academic Purposes (EAP) need more specificity and focus in their teaching and assessment of tasks for academic purposes in order to create disciplinary identities. Drawing on a series of 41 elicited student narratives from two cohorts over two semesters, this study considers what this 'focus' might comprise and describes how a portfolio approach to academic writing prepares students for generic writing skills and strategies while engaging with the types of texts students will read and create in future destinations. The study uses student voices to propel a narrative enquiry into what motivates them to participate in the unit 'Academic Writing' and what they realise is useful for their future disciplinary identities.*

## Background and purpose

This paper reports on a qualitative research project relating the content and form of the writing portfolios of students enrolled in the first-year tertiary unit 'Academic Writing' (AW) to three dimensions. The first of these is a focus on student 'destinations', which we relate to the concepts of 'disciplinary communities' and 'imagined communities'. Secondly, we respond to the identification in recent scholarship of the need for more specificity in creating 'disciplinary identities' (Coffin, et al., 2005) for 'professional membership' (Flowerdew, 1993). Thirdly, and more broadly, we produce findings stressing the role of the 'portfolio approach' in engaging students with discourses needed for future destinations. In short, this paper comprises an investigation into how AW programs in English as an Additional Language (EAL) engage students to invest in assessed portfolios where they analyse and create text types characteristic of their destinations.

We begin by analyzing student needs. Clearly, there is the instrumental need to achieve the outcomes of course; but there is a more integrative need for students to write for their future disciplinary communities, communities of their imaginings. Gardner (1985, p. 10) defined such motivation as "the extent to which an individual works or strives to learn the language because of a desire to do so and the satisfaction experienced in this activity." Acknowledging Gardner's (1985) view of motivation as a concept involving goal, effortful behaviour, desire to attain the goal and positive affect, we use the term 'investment' referencing the poststructuralist non-fixedness of identity and desire (Norton, 2000) to capture the centrality of students' desires to identify with academic and professional destinations as more representative of the

sociocultural reality of students in AW in 2014. AW utilises an investment-focused pedagogy that acknowledges students “for the complexity underlying their motivations, desires, and hopes for the future” (Pittaway, 2004, p. 216). At the same time it resituates the intercultural academic literacies AW students studying in a New Zealand context require for a globalised world, noting that such conceptions of literacy incorporates “competencies, attitudes and identities in addition to understandings” and crosses ‘cultural boundaries’ (Heywood, 2002, p. 10).

Since a great deal of recent research in both social identity theory (Norton, 2000) and future selves theory (Dörnyei & Ushioda, 2011) has argued that student identities are shaped by a range of motivational factors including desired destinations, we view the concepts of the *discourse community* and more latterly *disciplinary communities* as an inseparable part of the social constructivist idea of students as apprentices to a specified or privileged discourse (Woodward-Kron, 2004) desiring to belong to future communities. Borg (2003) saw as the major characteristic of discourse communities the sharing of goals and use of written communication to engage in repertoire-sharing. Swales (1988, 1990) had viewed them in terms of employing genres, each with their own fields of language and each characteristic of the particular community. To apply this notion pedagogically is to understand that the texts characteristic of discourse communities comprise ‘teachable’ texts and practices (skills, strategies, conventions, ways of structuring, cultural understandings, ways of being). With our better understanding of the role of desired destination as a key focus of investment in AW, discourse communities are, perhaps, more appropriately called ‘disciplinary communities’ (Coffin, et al., 2005).

This concept aligns with Anderson’s (1983) notion of ‘imagined communities’ since they are, as Kanno and Norton (2003) point out, desired, not current, discourse communities. In summarising literature on imagined communities, Norton and Gao (2008) make the case that “the people in whom learners have the greatest investment may be the very people who represent or provide access to the imagined community of a given learner” (p. 114). Students imagine themselves as members of future academic, local, national or professional communities using the language of these communities in specific ways characteristic of them (Kanno & Norton, 2003). In terms of engaging students of AW in producing texts encountered in their future destinations, instructors need to enquire into these communities and collect samples of their repertoire. Teaching writing moves from being an individual activity to a social one: “learning to write is part of becoming socialised to the academic community – finding out what is expected and trying to approximate it” (Silva, 1990, p. 17). Two insights from Caroline Coffin and her collaborators (2005) are helpful at this point:

As they progress through the university, students are often expected to produce texts that increasingly approximate the norms and conventions of their chosen disciplines, with this expectation peaking at the level of postgraduate study (p.2).

Students have greater control over their writing if they are helped by lecturers to develop an explicit awareness of how different disciplines employ different text types and how these text types construct and represent knowledge (both through their text structure and through their use of register) (p. 46).

These ideas led us to ask (i) how considering imagined discourse communities impacts the teaching and learning of AW and (ii) how such a pedagogy enhances students' investment by enabling their creating and recreating of discourses of these future destinations (Woodward-Kron, 2004)? We partially answered the first part of this enquiry in another article (Andrew & Romova, 2012), and in the present study we continue with the latter part of the question. This investigation also allows us to develop our earlier demonstration that the benefits of a portfolio-based pedagogy and mode of assessment provide spaces to negotiate cross-cultural and individual voices within the conventions of the discourse (Romova & Andrew, 2011). We also continue the argument that portfolios provide multiple opportunities for rehearsing a variety of text types, creating an "album of literacy performances" (Ferris & Hedgcock, 2005, p. 322). We argue that in an AW program such literacy performances need ideally to suit student future destinations, eclectic though they may be.

### **Context, participants and pedagogy**

The data was collected in the context of the unit Academic Writing, a year-one, 36-hour, 12-week EAL program aimed at both EAL majors and students from the wider university community who had achieved IELTS 6.0. Within the course, EAL learners meet the academic demands of tertiary study in their chosen fields. Specifically, the 41 students in the study aimed to develop abilities to write discourses aligned with various present and future majors: Nursing, Business Studies, ECE, Computer Science, Communication, Medicine, Statistics and Social Practice.

Participants, all aged between 18 and 39, hailed from China, Hong Kong, South Korea, Japan, Russia, Vietnam, Somalia, Ethiopia, Israel, Tonga, Nepal, and Malaysia. The gender ratio (14M-27F) is typical of EAL programs in the institution. National identity is not a variable of significance in the current study although clearly future studies will explore intercultural dimensions.

Within AW students produced portfolios comprising seven text types from critiques to expository and argumentative essays. The pedagogy working towards this final output involved learners in weekly multi-draft formative written tasks and used peer and tutor micro- and macro-level feedback techniques (Hamp-Lyons, 2006; Hyland & Hyland, 2006; Leki, 2006; Rollinson, 2005). Class tasks focused on instruction of text features so students "can better understand how to make a piece of writing more effective and appropriate to the communicative purpose" (Reppen, 2002, p. 322). While similar in genre, the text types students produced were tailored to suit the communicative purposes of their imagined and desired communities (Borg, 2003; Ramanathan & Kaplan, 2000) via the collection of a library of disciplinary-specific texts (Flowerdew,

1993; Hinkel, 2002). Hence, while producing critiques, some students might critique nursing texts, and others those from the discipline of Statistics. Swales (1988) offers a rationale for such pedagogy: he views communities as systems where the multiple beliefs and practices of text users overlap and intersect. Further, Hyland (2002) argues: “the teaching of key genres is seen as a means of helping learners gain access to ways of communicating that have accrued cultural capital in particular communities” (p. 125).

Effectively, as much as practicable, the program catered to students’ present career choices. Pedagogical interventions include teacher monitoring and conferencing, peer review and collaborative group work. To be specific, the pedagogy involves focusing on unpacking the generic features of authentic texts belonging to students’ desired disciplinary and professional futures (Flowerdew, 2000; Reppen, 2002) in terms of the *action* they aim to achieve.

A range of principles informing the teaching and learning incorporated in portfolio assessment will enable readers to better understand the situated nature of the pedagogical approach used in AW. From Granville and Dison (2005), we understand that the processes of multi-drafting that imparts critical and reflective elements to students’ work are crucial to creating increased discourse awareness within such an approach. Portfolios are sites for practicing membership of imagined communities by highlighting applications of the literacy practices of those communities (Johns, 1997, 2002; Hyland, 2000). Importantly, students’ reflections on each draft feed back into the teaching and point to a formative and developmental function (Lam & Lee, 2009). A further feature of the approach is multi-drafting, which itself comprises the sequence: collection, reflection, selection, and ongoing peer and teacher feedback (Hamp-Lyons & Condon, 2003). Johns (1995) maintained academic literacies develop through the multiple (re)production of target texts. We have itemised a range of these literacies from appreciations of paraphrasing, brainstorming and self-editing to the role of listenership during teacher monitoring and conferencing in another article (Romova & Andrew, 2011). In this paper, we wish to emphasise the importance of developing individual learners’ understandings of a range of generic text types from future ‘imagined’ discourse communities (Ramanathan & Kaplan, 2000) and to expand on the argument that writing is social practice as well as process (Coffin, et al., 2005). To do so is to pay tribute to Carolyn Miller’s (1984) ‘rhetorical’ view in relation to L1 writers writing in disciplines: genres incorporate social actions and reflect understandings about participation in communities with “cultural rationality” (p. 165).

## **Methodology**

We reconstructed narratives of the 41 participants’ learning trajectories. These narratives used a range of qualitative data. Focus groups were held in week 2 and three weeks after the completion of the 12-week course, with guided questions. Transcriptions of the focus group sessions were triangulated with students’ reflections on their learning tasks. The researchers’ goal is to construct narratives of destination as

part of an extended study, but for the current report we present analysis of narrative while our narrative enquiry percolates. The findings we report here are thematic (Sandelowski, 1995) to embody the authentic, reflective, evaluative insights of real learner experience and bring out “indigenous themes” (Patton, 1990) and avoid descriptive statistics. Effectively we present an analysis of student narratives of destination. Ethically, all students formally agreed to participate and for their words to be used. All names are pseudonyms.

The interviews in week two and what is effectively week 15 had specific foci. In week two students responded to cues about reasons for enrolling; identifiable challenges of writing academic texts; their future imagined communities; the writing skills they perceived as needed for future studies; and ways they imagined they would use writing in the future. In week 15, responses focused on the impacts of regular writing/rewriting, feedback and follow-up in the target genres; any learning about structure and discourse; the usefulness of academic styles of writing to their future needs; and, with an eye on critical pedagogies, whether the style of writing taught on the course is culturally different from students’ expectations. Students also spoke of emerging and developing literacies (brainstorming, self-editing, summarising and paraphrasing) and their connections with their future plans.

## Findings

We organised the findings into three broad themes:

1. *Target genres and future destinations*: Writing academic texts in target genres enables investment in AW due to the texts’ connections with future discourse communities and destinations;
2. *Generic features of academic texts and imagined communities*: Micro- and macro-level learning from the generic text types rehearses practices found in students’ imagined communities;
3. *The universality of planning and organising as literacy practices*: AW promotes literacy practices that are recognised as valuable for learners’ long-term futures.

### *Target genres and future destinations*

Student narratives demonstrate a future-focused orientation that is vocational (‘to practice medicine’), academic (‘to write reports for future studies’) and sociocultural (‘to interact with future co-workers’). The chance to write authentic text types appears to enhance learner investment.

To paraphrase Mabel (a future Business major), the literacy practices of AW (she calls them “skills” and “conventions”) are applicable to a business context (her imagined community). She adds her creation of a business-style paragraph made her feel she had worked “usefully”. Irma (Future destination: ECE) stated: “I think this type of writing [can help] my future study.” Asked to be more specific (and hence of relevance to theme 2 above), she states she values “restructuring the writing to suit ‘academic



writing” and “finding the right words for the topic.” Sue (also ECE) was more specific: “understanding the logic of the expected order – topic sentence and conclusion” are as useful for ECE as for any discipline. Farah (Business) appreciated the work at a metacognitive level: “developing the thinking skills needed to fit in with learning expectations of the genre.” Emma (Computer Sciences) contributed the insight that “thinking in a logical and chronological order while focusing on sentence structure too” are strategies she will require “next year.” It is, of course, impossible to cover all desired genres, and this too is reflected in the narratives with Miwa (ECE) lamenting: “for the future I want to learn not only this argumentative essay but also other different genres of writing.” William (Nursing) articulated the connection most broadly: “What I am learning to do here is related to what I want to do. I am getting ready for further study – looking into the future.”

### ***Generic features of academic texts and imagined communities***

The narratives we are unfurling tell the tale that learning from text types helps to engage learners. Yuichi (bound for the Royal New Zealand Police College) is forced to think about learning writing in a way that challenges his lexically-focused comfort zone. Academic writing genre, he says, “brings its own stress, so you can’t merely focus on vocabulary...I need more logic, as in academic reports.” This “logic” he needs is clearly related to the internal coherence of academic discourse. Kirma (ECE) specifically thought at the discourse level and nominated ‘textual organization’ as a key learning gain: “the process – pre- writing, outlining and so on – controls my ideas when I write my essay for Academic Writing and for Education”. She calls it “a skill [she] will reapply in her later career.” Helen (ECE) would agree, seeing ‘outlining’ as a crucial macro-level skill: “A good outline is guarantee of a good draft. I have learnt the writing process in AW and I will apply it in my studies in Education.” Interestingly, she explicitly looks to her future community.

For Ella (Nursing), brainstorming came closest to giving her an eureka moment about the generic features of academic texts: “Brainstorming...is the cornerstone that makes your whole essay link well. AW for IELTS and TOEFL is different from AW for nursing. The idea of logical development of text is different”. Michael (Computing) sums up the generic nature of AW as taught in this program: “In every subject, there is AW though they differ from one another.” Jad (Architecture) appreciates the formula due to his personal interest rather than a vocational one: “Writing is my interest as I was a journalist in the university newspaper in China. So I’d like to write articles. My major’s architecture. When we do models, we don’t need AW, but I like it – that’s my interest.”

Farat (Business) observed starting with the thesis and then writing topic-based sentences helps those wanting to write academically in subjects other than merely English. Emma (Computer Science) contributed that understanding the “conventions of structure [that] affect the coherence of an essay and give it its overall quality” is a crucial outcome for her as a student bound for a discipline beyond EAL. Yohana

(Health Science) related AW to a personal imagined community:

My auntie died from cancer – because there was not enough medical care. So I am planning to be a surgeon – not only to operate on people, but also to find the reason of cancer and to find treatment for it. AW will help me write reports of my discoveries + research skills.

Similarly, Dan sees big ambitions beyond the AW horizon:

AW reflects the way we think, so it helps the person to develop as a critical thinker, so it's a good tool to have in the future...You have to develop your thought, you have to deliver your thinking, to organise, to edit. So it helps you to reflect on yourself...In writing you have more time to develop what you want to say, to research and to have more thinking about your ideas than in speaking on the spot...It comes back to the role of education in the world. It is to increase awareness of people, to kill diseases and to help improve our lives.

Investment in imagined communities is tied to perceived learning gains in AW. Tala (Social Practice) reflected: "I need to improve my AW as at tertiary level you are going to do a lot of writing, and my writing is not up to standard."

Occasionally those already employed enrol in AW, as in the case of Nicky who has a job in insurance: "I have come here to study only English so that I could come up in my job. My position is supervisor and I need to write to insurance companies. Underwriter." Nicky sees a connection between the work of an AV program and potential for promotion. Similarly, Wanli (career in nursing) narrates her own reasons for investing in AW:

If you want to write a good assignment, you need to do AW. For your AW, speaking well is not enough. I have worked in a private hospital for several years. Sometimes, my letter to the pharmacy or the doctor came back as they needed to confirm what exactly I needed. So AW is vital for us, especially for this kind of professional work. I am a registered nurse and got my registration six years ago, but I have complaints from families...That's why I've come to this course. It is not only the speaking, but the AW that is going to help me as we must connect with doctors, or other nurses, or with the community, or a specialist from somewhere else and hear from them.

On the other hand, some students report that their future discipline will not specifically require AW:

I plan to go to Uni next year, and I've never been trained in writing academic things. My Chinese friend told me Chinese students have a lot of difficulty at Uni in writing academic things. I chose statistics because I need a good job in the future. Secondly, there are fewer assignments in statistics (Esson, Statistics).

Actually, I don't like AW – it's too hard for me, but I want to go onto further study. I am thinking of becoming an interpreter for courts and hospitals. And

good writing influences speaking. They go together: writing and speaking (Kenny, Interpreting/Translation).

Kenny's oral rather than written orientation and Esson's unfamiliarity with 'writing academic things' appear to reduce their investment in AW; nonetheless their comments concede to the need and usefulness of the subject.

### ***The universality of planning and organising as literacy practices***

The third group of key indigenous themes relates to specific literacies – viewed as skills, strategies, techniques and procedural 'how to's – identified in focus group interviews and reflective writings. Dan (Psychology) observes: "The structuring of an essay and organising it in an academic way. It's like a new language I need to learn to speak to meet the expectations." He says he sees AW as a foundation for academic qualifications. Emily (Business) described her perceived learning capital:

How to improve in the future: copy some good phrases while reading, rehearse and practice them...There is not a shortcut to improve my English in a sudden way, but at least I have got some strategies to make it look better.

Vinna (Nursing) emphasised her learning of the value of prewriting and outlining: These "control my ideas when I write my essay – very central". Sadya had spoken to a family member who had graduated in Nursing and was now at Middlemore Hospital. She regularly complained of her needs to summarise patients' details, and specifically of her lack of vocabulary. This story shows Sadya's investment in utilising AW as a site for consciously gaining enough vocabulary to write essays. Yoh (Business) says of writing essays in his discipline in the future: "At Uni, it'll be busy and I'll have to do a lot of research and reference the sources...(these) can be useful in other types of writing: report writing and case study." He has heard reports that 'Kiwi' students in the disciplines are often impatient with EAL students: "At Uni, we are studying with native speakers, and they expect us to be as confident/competent as native speakers." A desire to be seen as competent, a desire to keep face if you like, is central to Yoh's investment in learning AW. He says later, understanding a sociocultural motivation, "When we write, it helps us to develop relationships with people". Thorne, with the same future in mind, speaks of summarising and paraphrasing. In business studies, "I have to read and summarise a lot of documents. It may help me collect the main idea and understand the whole thing". Emma (Computer Science) reported learning gains in planning and organising: "The process of AW (pre-writing and outlining) helped me to organise ideas simply and start to write easily."

Jenny (Business), like Helen (cited above) nominated 'outlining': "Above all, outlining is the best thing for me...and I feel that if I prepare the outline well and in detail, then the time of writing an essay gets shorter." Like the other voices reported in this section, Jenny claims applying literacy practices better equip her.

## Discussion

Examining the findings, we find that the acts of learning to write point to the process of becoming socialised to future, imagined communities. We observe learners “in continual discussion, analysis and evaluation of their processes and progress as writers” (Hamp-Lyons & Condon, 2003, p. 15). We find evidence of both enhanced reflective capacity (Kathpalia & Heah, 2008) and the evolving literacy practices that are embedded in academic writing (Adamson, 1992; Johns, 1995). Specifically, students attend to such literacy practices as outlining and paraphrasing (Keck, 2006), self-editing (Ferris, 2005) and brainstorming (Rao, 2007). Interestingly, we find a particular awareness of macro-level, discourse-level thinking (Hyland, 2003, 2005). The evidence supports Silva’s observation that portfolios “[enable] learners to find out what is expected (in their future imagined communities) and then try to approximate it” (Silva, 1990, p. 17).

In the students’ reported emerging understandings, we see that students’ needs are connected to being and becoming members of future discourse and disciplinary communities. Students’ aspirations accord with understandings of imagined communities as places of the heart and mind that reference identity (Norton & Gao, 2008). As this paper works towards concluding, it is clear that an AW program needs to consider learners’ investments, that is, their next destination. To do so is to understand from the outset of the course not just their desired identities and future selves (Dörnyei & Ushioda, 2011) but also their spatial correlative: their destinations. Only from this point can we sculpture a curriculum that strives to allow spaces for individual identity even as it follows a curriculum inscribed within top-down institutional and governmental expectations.

## Conclusion

First-year degree level adult learners report English learning benefits from creating text types characteristic of their imagined communities and desired destinations. Specifically, they speak of gaining increased understanding of the discursive and generic features of academic texts necessary for participating in future discourse communities, whatever their future (or current) subject or vocation may be.

We might ask as a result of this discussion if EAP is indeed the way of the future. As Coffin and her co-authors wrote:

As the provision of writing instruction has increased, higher-level courses in academic writing have been developed. In some cases these courses link disciplinary lecturers with writing specialists to focus on disciplinary forms of writing, as in ‘learning communities’ (Coffin, et al., 2005, p. 6).

This paper, however, has offered an instance of an AW program that strives to replicate the known advantages of focused, vocational EAP curricula, but to do so in a way that allows for the individual writer as well as the writer of academic genre.

This paper has argued that a discourse/disciplinary community-based pedagogical approach can impact on learners' investments in an AW program, and that such a program highlights the literacy practices characteristic of future discourse communities. Students value these practices as capital that they can use in their future studies, workplaces and lives. This paper advocates, for teachers of tertiary academic writing, a pedagogical approach that emphasises the theoretical link between the pedagogical use of portfolios as "albums" of "individualised" genre-focused texts and the learners' future, imagined, disciplinary communities (Ferris & Hedgcock, 2005; Ramanathan & Kaplan, 2000). To teach the student in front of us now, we need to have knowledge of their desired destination.

### Limitations

Any study relying on the self-reports of participants has the clear limitation of lacking triangulation by an objective observer. However, the goal of narrative research is increasingly in line with the foregrounding of participants' voices as naturalistic and authentic data that validate identity and experience. The findings above also come from a specific context – a tertiary institute in the Western part of Auckland – and may not be directly applicable to other contexts. The overall trajectory of the narratives reported here will, we hope, be insightful and applicable to a range of contexts.

### References

- Adamson, H. (1992). *Academic competence: Theory and practice: Preparing ESL students for content courses*. New York: Longman.
- Anderson, B. (1983). *Imagined communities: Reflections on the origin and spread of nationalism*. New York: Verso.
- Andrew, M., & Romova, Z. (2012). Genre, discourse and imagined communities: The learning gains of academic writing learners. *Journal of Academic Language and Learning*, 6(1), A77-A88.
- Borg, E. (2003). Key concepts in ELT: Discourse community. *ELT Journal*, 57(4), 398-400.
- Coffin, C., Curry, M-J., Goodman, S., Hewings, A., Lillis, T. M. & J. Swann (2005). *Teaching academic writing: A toolkit for higher education*. New York: Routledge.
- Dörnyei, Z., & Ushioda, E. (2011). *Teaching and researching motivation*. Harlow: Pearson Education.
- Ferris, D. (2005). Teaching students to self-edit. In J. C. Richards & W. Renandya (Eds.), *Methodology in language teaching: An anthology of current practice* (pp.328-333). Cambridge: Cambridge University Press.
- Ferris, D., & Hedgcock, J. (2005). *Teaching ESL composition: Purpose, process and practice* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Flowerdew, J. (1993). An educational or process approach to the teaching of professional genres. *ELT Journal*, 47(4), 305-316.
- Flowerdew, J. (2000). Discourse community, legitimate peripheral participation, and the non-native-English-speaking scholar. *TESOL Quarterly*, 34(1), 127-150.

- Gardner, R. C. (1985). *Social psychology and second language learning: The role of attitudes and motivation*. London: Edward Arnold.
- Granville, S., & Dison, L. (2005). Thinking about thinking: Integrating self-reflection into an academic literacy course. *Journal of English for Academic Purposes*, 4, 99-118.
- Hamp-Lyons, L. (2006). Feedback in portfolio-based writing courses. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 140-161). New York: Cambridge University Press.
- Hamp-Lyons, L., & Condon, W. (2003). *Assessing the portfolio: Principles for practice, theory and research*. Cresskill, NJ: Hampton Press.
- Heywood, M. (2002). From international to intercultural: Redefining the international school for a globalized world. *Journal of Research in International Education*, 1(9), 8-32.
- Hinkel, E. (2002). *Second language writers' text: Linguistic and rhetorical features*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Hyland, K. (2000). *Disciplinary discourses: Social interactions in academic writing*. London: Longman.
- Hyland, K. (2002). Genre: Language, context and literacy. *Annual Review of Applied Linguistics*, 22, 113-135.
- Hyland, K. (2003). Genre-based pedagogies: A social response to process. *Journal of Second Language Writing*, 12, 17-29.
- Hyland, K. (2005). *Genre and second language writing*. Michigan: University of Michigan Press.
- Hyland, K., & Hyland, F. (Eds.), (2006). *Feedback in second language writing: Contexts and issues*. Cambridge: Cambridge University Press.
- Johns, A. (1995). Teaching classroom and authentic genres: Initiating students into Academic cultures and discourses. In D. D. Belcher & G. Braine (Eds.), *Academic writing in a second language: Essays on research and pedagogy* (pp. 277-292). Norwood, N.J.: Ablex.
- Johns, A. (1997). *Text, role and context: Developing academic literacies*. Cambridge: Cambridge University Press.
- Johns, A. (2002). *Genre in the classroom: Multiple perspectives*, Mahwah, NJ: Laurence Erlbaum.
- Kanno, Y., & Norton, B. (2003). Imagined communities and educational possibilities: Introduction. *Journal of Language, Identity and Education*, 2(4), 241-249.
- Kathpalia, S., & Heah, C. (2008). Reflective writing: Insights into what lies beneath. *RELC Journal*, 39(3), 300-317.
- Keck, C. (2006). The use of paraphrase in summary writing: A comparison of L1 and L2 writers. *Journal of Second Language Writing*, 15(4), 261-78.
- Lam, R., & Lee, I. (2009). Balancing the dual function of portfolio assessment. *ELT Journal*, 64(10), 54-64.
- Leki, I. (2006). "You cannot ignore": L2 graduate students' response to discipline-based written feedback. In K. Hyland & F. Hyland (Eds.), *Feedback in second*

- language writing: Contexts and issues* (pp. 266-287). Cambridge: Cambridge University Press.
- Miller, C. (1984). Genre as social action. *Quarterly Journal of Speech*, 70, 151-67.
- Norton, B. (2000). *Identity and language learning: Gender, ethnicity, and educational change*. London: Longman.
- Norton, B., & Gao, Y. (2008). Identity, investment and Chinese learners of English. *Journal of Asian Pacific Communication*, 18(1), 109-120.
- Patton, M. Q. (1990). *Qualitative evaluation and research methods*. Thousand Oaks, CA: Sage Publications.
- Pittaway, D. A. (2004). Investment and second language acquisition. *Critical Inquiry in Language Studies*, 1, 203-218.
- Ramanathan, V., & Kaplan, R. (2000). Genre, authors, discourse communities: Theory and application for (L1 and) L2 writing instructors. *Journal of Second Language Writing*, 9(2), 171-191.
- Rao, Z. (2007). Training in brainstorming and developing writing skills. *ELT Journal*, 61(1), 100-106.
- Reppen, R. (2002). A genre-based approach to content writing instruction. In J. C. Richards & W. Renandya (Eds.), *Methodology in language teaching: An anthology of current practice* (pp. 331-327). Cambridge: Cambridge University Press.
- Rollinson, P. (2005). Using peer feedback in an ESL writing class. *ELT Journal*, 59(1), 23-30.
- Romova, Z., & Andrew, M. (2011). Teaching and assessing academic writing via the portfolio: Benefits for learners of English as an additional language. *Assessing Writing*, 16(2), 112-122.
- Sandelowski, M. (1995). Qualitative analysis: What it is and how to begin. *Research in Nursing and Health*, 18, 371-375.
- Silva, T. (1990). Second language composition instruction: Developments, issues, and directions in ESL. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 11-23). Cambridge: Cambridge University Press.
- Swales, J. M. (1988). Discourse communities, genres and English as an international language. *World Englishes*, 7(2), 211-220.
- Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.
- Woodward-Kron, R. (2004). 'Discourse communities' and 'writing apprenticeship': An investigation of these concepts in undergraduate education students' writing. *Journal of English for Academic Purposes*, 3, 139-161.

---

## REVIEWS

---



**Barnard, R., & McLellan, J. (Eds.). (2014). *Codeswitching in university English-medium classes: Asian perspectives*. Bristol: Multilingual Matters. 978-1-415-78309-089 (pbk.), 224 pp.**

**Reviewed by Lynn Grant, AUT University**

The editors explain in the preface to this edited book that “the perceived importance of English as *the* global language of communication, technology and business has led university authorities to promote the use of English as the medium of instruction” (p.1), but note the “resurgence of publications arguing that codeswitching in English language instruction can be socially, pedagogically and educationally valuable” (p.2). The book is comprised of international case studies, with each of the eight chapters focusing on particular university classrooms in different Asian contexts.

Following Ernest Macaro’s overview of research issues in classroom codeswitching (CS) the overview shows that each of the 8 chapters ends with a relevant commentary about the issues raised in the chapter:

- Chapter 1 – Ching-Yi Tien reports an exercise in reflective practice on the author’s use of CS in an Introduction to Linguistics course in a Taiwanese university
  - The commentary discusses the findings from the perspective of Conversation Analysis and relates them to CS in Hong Kong
- Chapter 2 – Lili Tan explores the use of CS by two EFL teachers in a university in Beijing
  - The commentary compares the findings with those from elsewhere in China and Chinese teachers’ beliefs about CS
- Chapter 3 – Simon Humphries considers CS in two Japanese institutions
  - The commentary considers the extent to which the findings would apply to his own teaching context elsewhere in Japan
- Chapter 4 – Chamaipak Tayjasanant compares the contrast the use of CS in 4 different national contexts in Thailand
  - The commentary gives an overview of the language and education situation in Bhutan and makes comparisons with the Thai situation
- Chapter 5 – Le Van Canh reports on a Vietnamese university teacher’s use of the first language in four EFL classes
  - The commentary from Indonesia points out the scarcity of research into the issues of CS in both Vietnam and Indonesia, especially in university contexts
- Chapter 6 – Five lecturers look at the attitudes towards CS of three Bruneian English-Malay bilingual tutors in a university language centre
  - The commentary considers the extent to which the language-in-education situation in Malaysia is similar to that in Brunei

- Chapter 7 – Kenneth Keng Wee Ong and Lawrence Jun Zhang report on the linguistic analysis that the authors conducted on samples of CS by 22 undergraduate students in a Singapore university
  - The commentary discusses the dominance of English in Philippine society and the commonplace occurrence of CS, especially among the educated classes, and its use as a natural form of communication
- Chapter 8 – Moyra Sweetnam Evans and Ha Rim Lee, writing about Korean students at a New Zealand university, analyse written extracts from a reading comprehension task
  - The commentary begins by discussing the intense pressure in Korean education to learn English and the recent moves towards English-medium instruction in universities

The book ends with an afterword by Andy Kirkpatrick in which he comments on the issues addressed by Macaro and the contributors. The editors note that Macaro and Kirkpatrick have reflected on the significance of these Asian university case studies, and provided pointers towards a theory and rationale of classroom CS. The editors also note that “many of the contributors make a case for a more tolerant and nuanced view of the use of the L1 in L2 classrooms, thus attempting to redress the previous dominance of the zero-tolerance position” (p.8).

The book, while offering new information about CS in an Asian university context, encourages further research, especially into CS in multicultural contexts. In particular, research into the advantages of code-switching in the language classroom would be welcomed.

---

**Griffiths, C. (2013). *The strategy factor in successful language learning*. Bristol: Multilingual Matters. ISBN 978-1-84769-940-4 (pbk.) 220 pp.**

**Reviewed by Sara Amani, University of Auckland**

The role of strategies in successful language learning has deservedly received increased attention over the past forty years or so. During this period there have been numerous attempts to define and classify strategies (e.g. Oxford, 1990), to explore their theoretical underpinnings (e.g. O'Malley & Chamot, 1990), and to show the relationship between strategies on the one hand and learner and situational variables on the other (e.g. Kyungsim & Leavell, 2006). In *The strategy factor in successful language learning*, Carol Griffiths offers a comprehensive overview of research into language learning strategies from conceptual, quantitative, qualitative and pedagogical perspectives. Aware of the lack of definitional consensus among scholars over core characteristics of language learning strategies, she argues that it is essential to define the strategy construct in order to conduct meaningful research. To this end, she

follows an extensive review of previous literature and defines strategies as “activities consciously chosen by learners for the purpose of regulating their own language learning” (p. 36). She identifies the types of strategies more typical of higher- and lower-level learners, and identifies a significant correlation between strategy use and successful language learning. Following other strategy theorists and researchers, Griffiths adopts a cognitive perspective in which, she assumes, learners consciously process, manage and act upon their own learning. However, she then goes on to consider the theoretical base of strategy theory, which encompasses behaviourism, complexity/chaos theory, sociocultural theory, activity theory, and schemata theory among others.

The book comprises four main chapters. The first chapter sets up a thorough conceptual framework for understanding the strategy construct by presenting an overview of related terminology, definitions, and categorizations, proposing a definitive definition of language learning strategies, and exploring the theoretical underpinnings and inconsistencies. The strength of this overview lies in the concise language in which Griffiths presents the basic concepts, embeds them within a theoretical standpoint, and considers strategy effectiveness in relation to a broad range of learner variables. The book then goes on to report a number of experimental studies that have explored some of the issues involved in strategy research. While the second chapter seeks to explore the correlation between frequency and quantity of strategy use with successful language learning from a quantitative perspective, the third attempts to interpret individual variations from a qualitative perspective. These chapters take the reader step-by-step through methodologies and data analyses which may be adaptable for use in other teaching/learning contexts. The book concludes with a comprehensive review of the pedagogical research addressing learner, situational, target, and teacher training variables in strategy-based instruction. The final chapter considers strategies from teaching/learning perspectives, and provides a user-friendly guide for a wide audience including undergraduate students, teacher educators, teachers, teacher trainees, and researchers.

Griffiths attempts to capture the richness and complexity of strategy theory by adopting an eclectic theoretical base; however, this means she does not address the need for a sound theoretical foundation for research. And since she is no stranger to the controversy over strategy categorization, she recommends a *post hoc* rather than an *a priori* classification in which strategies are grouped on a case-by-case basis according to particular learners, targets, contexts, and goals involved. This pragmatic advice does not, however, help achieve a clear consensus among the extant set of strategy categories. Both researchers and educators still need to be able to widely apply a consistent set of theoretically sound categories. Overall, the scope of the ideas expressed in this book makes it a great resource for a wide range of readers. The author’s clear descriptions of a wealth of terminology and classifications, and explicit guidance on how to draw pedagogical implications for classroom practice and teacher education make the text accessible and manageable for researchers and teachers alike.

Although the book should not be considered as a recipe for devising strategies and techniques for teaching or teacher training, it can provide the preliminary knowledge of strategies as determining factors in successful language learning.

## References

- Kyungsim, H., & Leavell, A. (2006). Language learning strategy use of ESL students in an intensive English learning context. *System*, 34(3), 399–415.
- O'Malley, J. M., & Chamot, A. (1990). *Learning strategies in second language acquisition*. Cambridge: Cambridge University Press.
- Oxford, R. L. (1990). *Language learning strategies: What every teacher should know*. Boston: Heinle & Heinle.

**McGrath, I. (2013). *Teaching materials and the roles of EFL/ESL teachers: Practice and theory*. London: Bloomsbury. ISBN 978-1-4411-4369-3 (pbk.). 240 pp.**

**Reviewed by Patrick Coleman, Lincoln University**

English language teachers crave quality teaching materials. Given the depth and breadth of options on the market today, choosing a range of new teaching materials that meet learner needs as well as inspire teachers is a complex process. Poor selection or indeed poor delivery on the part of the teacher can lead to dull lessons or frustration from teachers trying to make a coursebook relevant. In 2002, Ian McGrath previously published *Materials evaluation and design for language teaching*, which was a practical, classroom-based guide to materials evaluation and design. This new volume builds on his previous work by comprehensively exploring the field. In the Preface to the book, McGrath describes an aim of presenting the work of scholars, researchers and practitioners side by side, so that gaps between their perspectives (and ways of bridging those gaps) become identifiable.

The book is divided into three parts. Part One focuses on external or theoretical perspectives. Chapter 2 is a reasonably detailed explanation of the process of coursebook production and the roles that publishers and coursebook writers play in this. Chapter 3 delves into areas such as coursebook selection and adaptation based on the current professional literature. McGrath makes an important argument as to why teachers need to adapt, since “coursebooks are written for everyone and no one” (p. 59). Teacher decisions around what and how to supplement are also explored, and McGrath notes that there has been little in the way of research in these areas. He ends the chapter by pointing out that, to date, learners have been neglected in research on materials evaluation and design. Chapter 4 considers the role of teacher educators, and McGrath sets up the debate between anti- and pro-coursebook views, which he uses as an opportunity to critique the anti-coursebook view. He emphasises the importance of

language learning materials, and that, therefore, their “selection, use and design cannot be consigned to the periphery of a teacher education programme” (p. 100). He highlights the usefulness of coursebooks, especially for beginning teachers as they develop their teaching skills.

Part Two deals with teacher and learner perspectives, and how their expectations are reflected in practice. Chapter 5 discusses materials (e.g. high school coursebooks) evaluation studies across the world, which reveal teachers’ needs for coursebooks that are up-to-date, engaging and relevant. McGrath notes that while the process of teacher evaluation of materials can “inform action”, there is still a lack of published research in this area (p. 126). Probably the most overlooked area of coursebook research is learner perspectives, which are explored in chapter 7. In chapters 5 to 7 McGrath makes it clear that there is a gap between the theory outlined in Part One and actual classroom practices. In Chapter 8, he explores possible reasons for the gap using terms taken from Zheng and Davison (2008) of contextual constraints like “external, internal and situational forces”. He also adds individual constraints including inexperience and time, but ultimately concludes there is no one defining explanation for the gap. The final two chapters form Part Three. Chapter 9 considers implications for stakeholder groups such as teachers, managers and coursebook writers. Chapter 10 explores implications for teacher educators, and a proposal for material evaluation and design. Overall, McGrath calls for all the parties in the process to work together not just in terms of materials development, but also in teacher education as a way of developing teacher abilities to “make a real difference” (p. 219).

The need to consider theoretical and practical aspects in the process of creating, making, evaluating and using teaching materials will always be an issue for teachers in whatever context they teach. This text is therefore recommended for graduate students and teachers involved in the design and evaluation of English language teaching materials. McGrath has succeeded in providing a well-researched and, at times, provocative volume on the slowly expanding field of materials development and design.

## References

- McGrath, I. (2002). *Materials evaluation and design for language teaching*. Edinburgh: Edinburgh University Press.
- Yuen, K. (1997). Review of Allwright’s (1981) paper “Why use textbooks?” B.Ed. assignment. University of Nottingham.
- Zheng, X. & Davison, C. (2008). *Changing pedagogy: Analysing ELT teachers in China*. London: Continuum.
- Zhu, H. (2010). EFL teacher development: A reflective model. *Modern English Teacher*, 19, 60-63.
-

**Christison, M.A., & Murray, D.E. (2014). *What English language teachers need to know Volume III: Designing curriculum*. New York: Routledge. ISBN 978-0-415-66255-0. 256 pp.**

**Reviewed by Rosemary Wette, University of Auckland**

This is the third volume in a series of handbook-type publications on ELT. The first two volumes in this series were published in 2011 and carried the sub-titles “Understanding Learning” (basic background information) and “Facilitating Learning” (planning, instructing, and assessing processes). In this volume Murray and Christison again draw on their extensive experiences as teachers, scholars and teacher educators, and turn their attention to the ELT curriculum. They state in the Preface an aim of producing a resource that will be of use to pre-service, novice and experienced teachers working with learners of all levels of proficiency and ages in all types of courses and contexts. The book is certainly wide-ranging in content, though I wonder if a single volume on the ELT curriculum could ever meet the diverse range of needs, interests, course types and levels, types and levels of learners that currently exist. While this volume gives examples from a range of contexts, it also reflects the backgrounds of its authors in that it provides more information about curriculum practices in the United States and Australia than elsewhere.

The book is divided into three main sections, and the list of content topics that follows shows its breadth. Topics covered in Part I are contexts for ELT curricula (definitions, textbooks, models of curriculum development, the hidden curriculum and curriculum change); socio-political and historical influences (levels of curriculum, progressivism and “romantic radicalism” in the USA); the world of ELT (World Englishes, native and non-native speakers of English, local and nativised varieties); and the technological context (CALL). Part II is similarly wide-ranging, and includes key processes in curriculum design (from context and needs analysis to assessment and course evaluation), using the curriculum to connect lessons courses and programs (laterally and vertically), and quality assurance issues (national accreditation systems in a number of countries are described). Part III provides information about linguistic-based curricula (structural, functional, genre-based, vocabulary and skills-based approaches); Part IV describes content-based curricula (CLIL, topics, situations); Part V outlines learner-centred curricula (negotiated, humanistic, task-based); and Part VI gives an overview of learning-centred curricula that are outcomes, competency or standards-based. While it is true that these are all potential elements of an EAL curriculum, they are not of equal importance, so the decision to give them all more or less equal prominence is a little strange. Some are commonly-used core syllabus pivots, while others (e.g. situations, notions) usually appear only as components of a multi-dimensional syllabus, and I wonder if this might be confusing or even misleading to pre-service or novice teachers.

Coherence could be an issue in a book that includes relatively brief information on such a range of topics; however, in this volume it is enhanced by the use of a short vignette at the beginning of each chapter describing a teaching, learning or teacher education situation taken from the authors' extensive experience. However, it is a little disappointing that no teacher voices, other than those of the two authors, could have been included here or elsewhere in the book. Short tasks are positioned throughout each chapter, and these provide a visual break for the reader as well as possibilities for the teacher educator. A number of tasks are based on resources of some kind, and I found these particularly useful for the graduate student course I teach. Tasks comprising sets of questions or task verbs are also helpful.

The book has a number of clear strengths, as well as a few weaknesses. One shortcoming is the unevenness of source literature. End of chapter reference lists in a book of this kind should provide readers with the most influential texts on each topic, and while this is true for some chapters (e.g. *Genre-based approaches*; *Content-based curricula*), others (e.g. *Quality Assurance*; *Functional-Notional*; *Topics/Situations*; *Using the curriculum to connect lessons, courses and programs*) cite a fairly sparse number of resources, omit key texts (e.g. the *Task-based curriculum* chapter), or are all less than recent (e.g. *Humanistic curriculum*). The book's many strengths include its general accessibility in terms of style and content for its target audience, and the experience of its two authors, which help this volume to successfully bridge the theory-practice divide. While not quite as comprehensive as its aims and claims, it is nevertheless a valuable handbook-type resource on topics that are fundamental to ELT.

**Phillips, T. & Phillips, A. (2012). *Progressive Skills in English Level 4 Course Book*. [incl. Video DVD and Audio CD1&2]. Reading: Garnet. ISBN: 978-1-85964-685-4 (PB). 237 pp.**

**Phillips, T. & Phillips, A. (2013). *Progressive Skills in English Level 4 Teacher's Book*. Reading: Garnet. ISBN: 978-1-85964-687-8 (PB). 295 pp.**

**Phillips, T. & Phillips, A. (2012). *Progressive Skills in English Level 4 Workbook*. [incl.audio CD] Reading: Garnet. ISBN: 978-1-85964-686-1 (PB). 91 pp.**

**Reviewed by Margaret Bade, UNITEC**

*Progressive Skills in English* is a multi-level series which builds the skills of students preparing for academic English. In the review pack the Level 4 of the *Progressive Skills in English* series comprises a Teacher's Book, Coursebook (includes 4 CDs and a video DVD) and Workbook (with audio CD). This level is intended for Intermediate

to Upper Intermediate (CEF B1-B2; IELTS: 4.0-6.0) students who have been studying English for a number of years. As the Introduction to the course book (p.7) states, the series “helps students in all four skills: Listening – to lectures; Speaking – in tutorials and seminars; Reading – for research; and Writing – assignments.” It is rather a shame that the Level 4 Coursebook has the CDs and DVD attached in a pouch at the front and back, since it makes the book rather bulky to pick up and use comfortably. A Contents page and two pages of Book Maps outline each aspect of the topics and skills associated with five themes. The Introduction sets out the basic format, including additional elements featured in each theme: Everyday English, a Knowledge Quiz and a Portfolio. At the back are 20 pages of resources, 45 pages of transcripts, and a nine-page word list (also in the Teacher’s Book), which presents alphabetically the words from every skill section of the five themes.

Precise details on every lesson including answers and model essays are found in the Teacher’s Book, which is excellent value. For each lesson it details the objectives and a general introduction, highlights the grammar, optional activities, transcripts and answers as well as a closure section, and provides (through methodology notes) some useful tips on giving feedback in writing. The authors’ approach to the skills methodology is explained, and possible routes through the coursebook from a 25-hour course to a 120-hour course are provided. The authors have chosen four reasonably relevant themes: geography and the modern world; communication, media and advertising; living life to the full; and the past, present and future of food - and present them in an enterprising way. They have divided the themes into discrete skills, so that, for example, the food theme has the following sub-topics: Listening: *Agriculture through history*; Speaking: *Interfering with nature?* Reading: *Should man be a herbivore?* and Writing: *GM: The future or the end?* The four main skills’ sections for the themes comprise around 30 pages and all have well-planned exercises following a set pattern. For example, in the Listening sections, the headings are: *Vocabulary for listening*, *Real-time listening*, *Learning new listening skills*, *Grammar for listening* and *Applying new listening skills*. Each *Learning new* section provides a checklist of rules and examples to suit deductive or inductive approaches. For speaking and listening there is a colour-coded box on the right of the page for a pronunciation check as well. A special feature of the coursebook is the separate vocabulary list of around 40 words for each set of five lessons. The list contains items that are linked to the theme and “most are expected to be new to the majority of students in any class” (TB p. 13). A website offers further practice.

The workbook comes with a CD and accompanying grammar and vocabulary-related exercises. Transcripts are included at the back and students can locate the answers to most exercises from these. The material in the course is appropriate for Intermediate-Upper-Intermediate level students of academic English, who will be challenged to engage in the language material critically. Grammar and vocabulary feature alongside the skills. There are plenty of practice opportunities for all skills, as well as opportunities to research on-line and in journals.



Increasingly, Level 4-type courses in New Zealand cater for students who are studying in a range of disciplinary areas, and there is much here that such students would find helpful. Even if they are not directly relevant, texts relating to themes such as reality TV, the benefits of genetically modified foods, conventions in narrative fiction, violence in stories for children, communication inventors, island tourism and water problems in a developed country will appeal to a range of students. Teachers in New Zealand will, of course, recognise the cultural flavour of this UK coursebook, but the authors explicitly encourage students to relate the theme areas to their own countries. The audio CDs and DVD lectures feature a variety of English accents. All in all, the components of *Progressive Skills in English Level 4* provide a rich variety of language learning for a skills-focused course, with detailed support from the Teacher's Book and additional practice in the Workbook. The authors have used their extensive experience in the field of academic English to produce a flexible package which IELTS teachers and teachers of academic English Level 4 will welcome.

**Tomlinson, B. (2013). *Developing materials for language teaching*. London: Continuum. ISBN 9781441186836. 576 pp.**

**Reviewed by Mhairi Mackay, Wintec**

*Developing materials for language teaching* (2<sup>nd</sup> edition) follows the first edition of 2003, and is published at a time of increasing interest in materials development and its role as a bridge between SLA theory and practice. The second edition is informed by a desire to reflect the knowledge and experience of L2 practitioners as well as of linguistic experts, which has led to a wide and interesting range of contributions that aim to inform, stimulate and provoke across a variety of materials development concerns. The contributors (twenty in all including the editor) give an international view of aspects of current practice and include Paul Nation (on vocabulary), Alan Maley (on writing from a creative perspective), Claudia Saraceni (on adapting courses), Ken Hyland (also on writing), and Duriya Aziz Singapore Walla (on systemic functional theory and course book design). Unlike other recent publications in this field (e.g. the excellent McGrath publication of 2013), Tomlinson is not solely concerned with ESL, and materials for learners of other languages are also included.

Weighing in at nearly 1kg, the book is a bit of a door stopper; however, it is also available as an eBook. The contents of the book are presented in five main sections which respectively focus on evaluation and adaptation of materials; principles and procedures of materials development; developing materials for target groups; developing specific types of materials; and materials development and teacher training. Blended Learning (BL), Content and Language Integrated Learning (CLIL) and other relatively new instructional approaches are discussed in several areas of the book (see for example page 176 for a summary of creative teaching approaches, and

pages 207-221 which focus on BL), and are seen as reasons why many teachers will return to their current materials with a desire to rethink, develop, adapt, and create anew. For those of us facing any sort of change in our L2 teaching and learning environments, this book provides thinking space as well as practical approaches to principled materials design.

Our tertiary institution, like many English language course providers throughout New Zealand, is in the process of offering new language qualifications such as the New Zealand Certificates in English Language (NZCEL), and professional discussions in recent months have tended to focus on how they will be delivered. This book provides an excellent resource for informed and robust discussion of the decisions involved in adapting, creating and preparing materials. Harwood (2012) has also provided teachers with an excellent resource that combines theoretical and practical perspectives on materials design and development, and which further includes discussion tasks to deliberately stimulate this process of informed and robust materials design. In combination, these two texts provide experienced practitioners with valuable professional development in current scholarship and practice in materials development.

Tomlinson (2013, p.x) emphasises the role of the “informed and reflective” classroom teacher as the ultimate mediator between the theory and practice of materials development and use. With powerful commercial interests publishing materials and courses, it is increasingly important that the community of teachers and learners are actively involved in the discussion of what actually works, and what is needed to support SLA in a wide range of authentic local settings. Therefore, this book also acts as an important access point for the L2 teaching practitioner seeking good materials to support useful L2 learning to the increasingly complicated world of SLA theory and academic expertise. This book is highly recommended for teachers, and would be an excellent text to consider for selection for a teacher education course.

## References

Harwood, N. (Ed.), (2012). *English language teaching materials: Theory and practice*. Cambridge: Cambridge University Press.

---

**Vicary, A. (2014). *English for Academic Study: Grammar for writing*.  
Reading: Garnet. ISBN: 978 1 78260 070 1. 240 pp.**

**Reviewed by Martin White, University of Auckland**

*Grammar for writing* is part of the English for Academic Study series published by Garnet Education in conjunction with the University of Reading. This series covers the four skills, and in addition to a book on EAS Writing there is also a book on extended

writing and research skills. Its underlying premise can be found in the introduction where the author stresses the importance of grammar by pointing out the correlation between grammatical accuracy and good academic essays, as expressed in the statement that “it is impossible to produce an academic essay of a high standard if grammatical knowledge is weak” (p.6). There are many other grammar books with exercises available in publication, but one point of difference is that this book focusses on academic writing, and not only seeks to improve and increase grammatical knowledge but also aims to teach students how to apply it to academic writing tasks. It is designed as a self-study book for students who have a level equivalent to IELTS 4.0-6.5.

*Grammar for writing* has nine units which follow the same structure in every unit. Each unit has three sections, with the first two focussing on expanding grammatical knowledge, and the third part aiming to consolidate the grammar points of the previous sections and show how they can be used in written texts and essays. Each unit starts with a list of learning objectives, and continues with self-graded study tasks which constitute the core of the book. Navigation through the book is facilitated by the use of different colours for grammatical explanations and notes. Additionally, each unit finishes with a quiz for students to check their grasp of the material featured in that unit. A set of answers is available online.

As an example of the content of the book, the subject of Unit 6 is expressing shades of meaning using modal and semi-modal verbs to show certainty and uncertainty, remote possibility and advice and practising using modal verbs in active and passive voice. While there are indeed some conventional exercises such as gap filling and changing active to passive exercises, the author has provided more context than is usual in such books, and offers exercises that relate to the type of writing expected in an EAP classroom or examination. One exercise has the students using modals in an extract to a student essay – a conclusion to an essay – where advice or recommendations must be made. Although this is a gap-filling exercise, it is not just completing isolated sentences but, properly done, the student produces with a realistic piece of academic text. Another task seeks to teach the notion of stance through the development of a point of view in an argument essay and how, through using intensifying words, shades of meaning can differ. Striving for contextualisation and relevance must have made it harder for the authors to devise exercises that can be navigated without assistance from a teacher, and means that instructions tend to be very detailed. An example, again from unit 6, is: *In each sentence below, label the subject (S) verb (V) and object (O) of the second clause. Then rewrite the sentences using the passive with a modal verb, making a tight link between the clauses and omitting the redundant words in bold.* Hopefully, students will be able to identify the second clause and also understand what is meant by making a tight link.

A common complaint about grammar exercise books is that they often feature pages of decontextualised sentences requiring mostly filling in gaps or manipulating a lexis

item such as a verb form. Thankfully, for the most part *Grammar for writing* avoids such activities and is a welcome addition to Garnet's list of published books for two principal reasons. The first is that there is considerable variety in the exercise types on offer which have direct relevance to the types of tasks a student of academic English is required to undertake whether for an exam or a course assignment. Secondly, the emphasis in this book is specifically on academic writing and the grammar needed for accurate and effective writing in academic contexts. Just as a budding guitarist finds that knowledge of a few pentatonic scales does not easily translate into Claptonesque type solos, so the New Zealand survey of EAP teachers (Barnard & Scampton, 2007) reported that the majority agreed that it was difficult for their students to transfer their grammatical knowledge into the production of authentic language. *Grammar for writing* can help to bridge this gap and prove to be of benefit both to the independent learner and for occasional use in the classroom.

## References

- Barnard, R., & Scampton, D. (2008). Teaching grammar: A survey of EAP teachers in New Zealand. *New Zealand Studies in Applied Linguistics*, 14(2).

## NOTES FOR CONTRIBUTORS

1. Contributions to *The TESOLANZ Journal* are welcomed from language educators and applied linguists within and outside Aotearoa/New Zealand, especially those working in Australia and countries in the South Pacific.
2. Contributions should in general be no longer than 5000 words.
3. Referencing conventions should follow that specified in the Publication Manual of the American Psychological Association (6<sup>th</sup> Edition). This publication is available in most university libraries. In the text, references should be cited using the author's last name and date of publication. If quotations are cited, the reference should include page numbers (Brindley, 1989, pp. 45–46). The reference list at the end of the article should be arranged in alphabetical order. The reference list should only include items specifically cited in the text.
4. As far as possible, comments and references should be incorporated into the text but, where necessary, endnotes may be placed after the main body of the article, before the list of references, under the heading Notes.
5. All graphics should be suitable for publication and need no change.
6. It is understood that manuscripts submitted have not been previously published and are not under consideration for publication elsewhere.
7. Enquiries and draft submissions should be sent by email to the Editor, Dr Angela Joe, Victoria University of Wellington, [angela.joe@vuw.ac.nz](mailto:angela.joe@vuw.ac.nz). The preferred format is WORD.
8. All submissions should be accompanied by a full mailing address, a telephone number and, if available, an email addresses and/or fax number.
9. Submissions will be considered by the Editor and members of the Editorial Board.
10. Those interested in submitting a book review should contact the Review Editor, University of Auckland, [r.wette@auckland.ac.nz](mailto:r.wette@auckland.ac.nz)
11. The closing date for the submission of manuscripts for 2015 is **Monday 24 August, 2015.**