

---

# **NEW ZEALAND STUDIES IN APPLIED LINGUISTICS**

---

---

**VOL. 14 (2) 2008**

---

**ALANZ**

**APPLIED LINGUISTICS ASSOCIATION  
OF NEW ZEALAND**

---

# NEW ZEALAND STUDIES IN APPLIED LINGUISTICS

---

## EDITORS

**John Bitchener**

**Gillian Skyrme**

## EDITORS' ASSISTANT

**Heather Meyer**

## EDITORIAL BOARD

Gary Barkhuizen (University of Auckland)

Roger Barnard (University of Waikato)

Helen Basturkmen (University of Auckland)

Allan Bell (Auckland University of Technology)

Ian Bruce (University of Waikato)

Catherine Elder (University of Melbourne)

Dana Ferris (University of California)

Janet Holmes (Victoria University of Wellington)

Marcia Johnson (University of Waikato)

Andy Kirkpatrick (Hong Kong University)

Margaret MacLagan (University of Canterbury)

Hedy McGarrell (Brock University)

Shawn Loewen (Michigan State University)

Paul Nation (Victoria University of Wellington)

Bonny Norton (University of British Columbia)

Rhonda Oliver (Edith Cowan University)

Donna Starks (University of Auckland)

Elaine Vine (Victoria University of Wellington)

Cynthia White (Massey University)

*New Zealand Studies in Applied Linguistics* is published by the Applied Linguistics Association of New Zealand (ALANZ). Contributions to the journal can be submitted in the form of articles, short reports and summaries, and book reviews. Manuscripts following the guidelines for contributors included with this volume should be sent to:

Associate Professor John Bitchener  
Co-Editor, *New Zealand Studies in Applied Linguistics*  
School of Languages, AUT University  
Private Bag 92006  
Auckland, New Zealand  
Email: [john.bitchener@aut.ac.nz](mailto:john.bitchener@aut.ac.nz)

## NOTE

Views expressed in articles and reviews published in *New Zealand Studies in Applied Linguistics* are those of the contributors and do not necessarily reflect those of ALANZ.

**Contents**

**ARTICLES**

- Diagnostic writing ability: A rating scale for accuracy, fluency and complexity. 1  
*Ute Knoch*

- The effect and the influence of the use of video and captions on second language learning. 25  
*Judy Chai & Rosemary Erlam*

- Teachers' views and practice: The place of the mother tongue in the implementation of the Kenyan language-in-education policy. 45  
*Jennifer M. Jones*

- Teaching grammar: A survey of EAP teachers in New Zealand. 59  
*Roger Barnard & Davin Scampton*

- L1 literacy amongst generation 1b: A study of an Australian Afrikaans speaking community. 83  
*Henriette Janse van Rensburg, Aniko Hatoss & Donna Starks*

**REVIEWS** 93

- Theoretical frameworks in qualitative research.  
V. Anfara & N. Mertz (Eds.)  
*Reviewed by Judy Hunter*

- Teacher cognition and language education.  
S. Borg  
*Reviewed by Rosemary Wette*



Qualitative inquiry and research design: Choosing among five approaches (2<sup>nd</sup> Edition).

Creswell, J.W.

*Reviewed by Gary Barkhuizen*

Lessons from good language learners.

C. Griffiths (Ed.)

*Reviewed by Anne Feryok*

Conversational interaction in second language acquisition:

A collection of empirical studies.

Mackey, A. (Ed.)

*Reviewed by Roger Barnard*

Exploring the second language mental lexicon.

Singleton, D.

*Reviewed by Annette Sachtleben*

## **GUIDELINES FOR CONTRIBUTORS**

107



---

## ARTICLES

---





# **DIAGNOSTIC WRITING ABILITY: A RATING SCALE FOR ACCURACY, FLUENCY AND COMPLEXITY**

**Ute Knoch**

*University of Melbourne*

## **Abstract**

*Alderson (2005) suggests that diagnostic tests should identify strengths and weaknesses in learners' use of language, focus on specific elements rather than global abilities and provide detailed feedback to stakeholders. However, rating scales used in performance assessment have been criticised for using impressionistic terminology (Fulcher, 2003; Upshur & Turner, 1999) and often resulting in holistic marking by raters (Weigle, 2002).*

*This paper reports on the development and validation of a rating scale for assessing accuracy, fluency and complexity in EAP writing. 601 writing scripts by both native and non-native speakers of English enrolled at a large university were analysed using discourse-analytic measures of accuracy, fluency and complexity. Based on the findings, a rating scale was formulated. The study investigates whether such an empirically-grounded scale can be used to assess students' writing more reliably and with greater discrimination than more traditional measures. The validation process involved ten trained raters applying both sets of descriptors to the rating of 100 writing scripts. A quantitative comparison of rater behavior was undertaken using FACETS (a multi-faceted Rasch measurement program). Interviews were also administered to elicit the raters' perceptions of the efficacy of the two scales. The results indicate that rater reliability was substantially higher and that raters were able to better distinguish between the three different aspects of writing ability when the more detailed descriptors were used. Rater feedback showed a preference for the more detailed scale. The findings are discussed in terms of their implications for rater training and rating scale development in the context of diagnostic assessment.*

## **Introduction**

Alderson (2005) identifies a number of criteria which distinguish diagnostic tests from other types of test. For example, he argues that:

1. Diagnostic tests are designed to identify strengths and weaknesses of learner's knowledge and use of language.
2. Diagnostic tests should enable a detailed analysis and report of responses to items or tasks.

3. Diagnostic tests thus give detailed feedback which can be acted upon.
4. Diagnostic tests are more likely to be focussed on specific elements than on global abilities.

Based on the criteria presented above, Alderson (2005) wrote that diagnostic tests are generally different to proficiency tests. This poses the question whether diagnostic tests of writing should be rated with a different type of rating scale than those used in proficiency testing.

Proficiency scales commonly used for writing assessment have also been criticised for a number of reasons. Fulcher (1996), for example, suggests that most rating scales are constructed a priori. This means that the descriptors of the rating scales are constructed by experts who use their own intuitive judgements concerning the nature of language development. These scales therefore lack empirical underpinnings, and are at most validated in post hoc studies. Other criticisms levelled at commonly used rating scales are that scale descriptors are often not precise enough, and therefore do not provide sufficient information for raters to base their decision on (Upshur & Turner, 1995; Watson Todd, Thienpermpool, & Keyuravong, 2004) and that the scale points are frequently described by a list of features that may not actually co-occur (Upshur & Turner, 1995) or develop in step (Young, 1995). Furthermore, Weigle (2002) was able to show that raters often rate holistically when using an analytic rating scale and adjust their scores to match the analytic criteria.

It seems therefore that rating scales commonly used for proficiency testing are not suitable in a diagnostic context. If this is the case, what features should a rating scale for diagnostic assessment display? Considering Alderson's (2005) criteria of diagnostic tests, it would mean that a rating scale for diagnostic testing should be more detailed than a scale for proficiency testing. It should also not list different aspects of writing ability in the same level descriptor, as this would lower the diagnostic power of the test and therefore lower the potential of the feedback to students.

As a result of the ideas developed above, this study set out to investigate whether an empirically-developed ratings scale with detailed level descriptors functions more reliably and validly when students' accuracy, fluency and complexity in writing are diagnosed.

### **Accuracy, fluency and complexity**

Measures of accuracy, fluency and complexity are often used in second language acquisition research as they are believed to provide a balanced picture of learners' language ability (Ellis & Barkhuizen, 2005). 'Accuracy' refers to 'freedom of error' (Foster & Skehan, 1996, p.305), 'fluency' refers to 'the processing of language in real time' (Schmidt, 1992, p.358) with a focus on 'the primacy of meaning' (Foster &

Skehan, 1996, p.304) and ‘complexity’ is ‘the extent to which learners produce elaborated language’ (Ellis & Barkhuizen, 2005, p.139)<sup>1</sup>.

Underlying these measures is a psycholinguistic view of L2 proficiency and the assumption that learners may prioritise one aspect over another in certain situations. That L2 acquisition is a differentiated rather than a unitary phenomenon was first proposed by Meisel, Clahsen and Pienemann (1981). They advanced the Multidimensional Model of L2 acquisition and showed that certain learners achieve fluency at the expense of complexity and accuracy, whilst others might emphasise accuracy and complexity over fluency.

### **Accuracy, fluency and complexity in language testing**

Studies in the area of second language acquisition have been able to show that learners prioritise these different aspects in different situations and under different task conditions (see for example Ellis & Yuan, 2004; O'Loughlin & Wigglesworth, 2003). However, these findings have not been transferred into language testing practice. Most rating scales, for example, focus on aspects of language like grammar and vocabulary and therefore the areas of accuracy, fluency and complexity are confined to the same level descriptors. Typically accuracy and complexity of grammar, sentence structure and vocabulary are rated together. However, especially in a diagnostic context, it would make more sense to have raters scrutinise these aspects separately, to provide a more complete picture of a learner's writing ability.

Alderson (2005) also suggests that diagnostic tests should focus on specific abilities rather than global abilities. Therefore, rating scales used in diagnostic assessment should not be signified by the same type of imprecise wording as rating scales used in most proficiency tests. This study, therefore, proposes the use of discourse analytic measures as the basis for the level descriptors of the scale. This idea was also proposed by Iwashita, McNamara, and Elder (2001) who criticised measures of accuracy, fluency and complexity used in research as being too complex and time-consuming to be practical under operational testing conditions. They therefore propose a rating scale based on the aspects of accuracy, fluency and complexity rather than the use of the actual measures.

### **The study**

The aim of this study is to investigate whether accuracy, fluency and complexity can successfully be operationalised into a rating scale to diagnose writing ability. The study was undertaken in two phases. During the first phase, 601 writing samples were analysed to establish the accuracy, fluency and complexity displayed by writers at five levels of writing ability. Based on these findings, an empirically-based rating scale was developed. During Phase 2, the validation phase, ten raters rated 100 writing samples. The findings were compared to previous ratings of the same 100 scripts by the same raters using a pre-existing rating scale typical of scales used in

---

<sup>1</sup> Different definitions for these measures exist. For the purpose of this study, these definitions will be used.

proficiency testing of writing. At the end of the rating round, seven raters agreed to participate in more detailed interviews.

The following research questions were addressed:

1. What are the features of accuracy, fluency and complexity displayed at different levels of expository writing?
2. How valid and reliable are measures of accuracy, fluency and complexity when used in a rating scale for diagnostic assessment?
3. What are raters' perceptions of using measures of accuracy, fluency and complexity in a rating scale?

## Method

### Context of the research

#### The assessment instrument

DELNA (Diagnostic English Language Needs Assessment) is a procedure designed to identify the English language needs of first-year undergraduate students following their admission to the University, so that the most appropriate language support can be offered. The assessment includes a screening component which is made up of a speed-reading and a vocabulary task. This is used to eliminate highly proficient users of English and exempts these from the diagnostic procedure. The diagnostic component comprises objectively scored reading and listening tasks and a subjectively scored writing task.

The writing section is an expository writing task in which students are given a table or graph of information which they are asked to describe and interpret. Candidates have 30 minutes to complete the task. The writing task is routinely double (or if necessary triple) marked analytically on nine traits (Organisation, Coherence, Style, Data description, Interpretation, Development of ideas, Sentence structure, Grammatical accuracy, Vocabulary and Spelling) on a six point scale ranging from four to nine. The rating scale has no descriptors to rate fluency. Three trait scales on the DELNA scale are designed to measure accuracy and complexity, but as is the case in many rating scales designed for writing proficiency testing, these aspects are mostly confined to the same trait scales. For example, the trait scales for 'sentence structure' and 'vocabulary and spelling' refer to both range and accuracy. The trait scale entitled 'grammatical accuracy' refers only to accuracy. The DELNA rating scale reflects common practice in rating scales developed for performance assessment in that the descriptors are graded using adjectives like '*adequate*', '*appropriate*', '*significant*', '*satisfactory*' or '*limited*'.

#### The writing samples

To identify the specific features of accuracy, fluency and complexity used by writers taking DELNA, 601 writing samples were randomly selected.

### The candidates

The L1 of the students (as reported by a self-report questionnaire) was varied. 42% (or 248 students) have an Asian first language, 36% (217) are native speakers of English, 9% (52) are speakers of a European language other than English, 5% (31) have either a Pacific Island language or Maori as first language and 4% (21) speak either an Indian or a language from Sri Lanka as first language. The remaining 4% (22) were grouped as 'Other'. Eleven students did not complete the self-report questionnaire. The scripts used in this analysis were all rated by two DELNA raters. In cases of discrepancies between the scores, the scores were averaged and rounded (in the case of a .5 result after averaging, the score was rounded down).

### The raters

The ten DELNA raters taking part in this study were drawn from a larger pool of raters based on their availability. All raters have high levels of English proficiency although not all are native speakers of English. All raters have several years of experience as DELNA raters and take part in regular training moderation sessions either in face-to-face sessions and online (Elder, Knoch, Barkhuizen, & von Randow, 2005; Knoch, Read, & von Randow, 2007).

### **Procedures – Phase 1: Analysis of writing samples**

The first step of Phase 1 was to divide the 601 writing scripts into five groups according to the average score of the two ratings awarded.

After an initial pilot study to select the measures of accuracy, fluency and complexity that most successfully distinguish between the five different levels of DELNA writing, the following discourse analytic measures were selected. For accuracy, the percentage of error free t-units proved to differentiate most clearly between the different levels of writing during the pilot study. Both t-units and errors were coded by hand by the researcher. Fluency was divided into two categories following Skehan (2003), repair fluency and temporal fluency. Repair fluency was operationalised as the number of self-corrections, whilst temporal fluency was measured as the number of words. This measure was possible, because the writing was produced under a time limit of 30 minutes. The number of self-corrections was counted by the researcher. The number of words was established with the help of a Perl program. Complexity was divided into two sub-categories, grammatical and lexical complexity. Grammatical complexity was measured as the clauses per t-unit. Both t-units and clauses were coded by the researcher. Lexical complexity was measured by the number of words from the academic word list (Coxhead, 2000). The number of AWL words was established using the program VocabProfile (Cobb, 2002). For all measures that involved manual coding, inter-coder reliability was established by having a second researcher double code a subset of 50 scripts. To establish if variables successfully differentiated between the five levels of writing, ANOVAs or Welch tests (in cases where the assumption of equal variances was violated) were calculated.

## **Procedures – Phase 2: Rating scale validation**

### Data collection

The raters rated 100 scripts using the DELNA criteria and then the same 100 using the new scale. A counter-balanced design was not possible for practicality reasons, however because the two rating rounds were completed two months apart, there was some confidence that the raters were not able to remember any of the scripts from the previous rating round. The 100 scripts were selected to represent a range of proficiency levels. Before the rating rounds, all raters participated in a rater moderation session to ensure they were thoroughly trained.

After rating each set of 100 scripts, a subset of seven raters was interviewed. All raters were invited to participate in the interviews, but only seven were available at the time. The interviews were semi-structured. The questions focussed on the raters' perceptions of the two scales and their experiences using them. The interviews were audio-recorded and later transcribed by the researcher.

### Data analysis – quantitative data:

The results of the two rating rounds were analysed using multi-faceted Rasch measurement in form of the computer program Facets (Linacre, 2006). The model states that the likelihood of a particular rating on a given rating scale from a particular rater for a particular student can be predicted mathematically from the proficiency of the student and the severity of the rater.

Before the data was analysed, several hypotheses were developed for comparing two rating scales. Each of these hypotheses relates to a group of statistics generated by the FACETS program. Each of these is discussed below.

#### *Discrimination of the rating scale:*

The first hypothesis was that a more discriminating rating scale can be seen as superior. The more levels of candidate ability a group of raters can discern with the help of a rating scale, the better the scale is functioning. The candidate separation ratio is an excellent indicator of the discrimination of the rating scale. The higher the separation ratio, the more discriminating the rating scale is.

#### *Rater separation:*

The next hypothesis was that a well functioning rating scale would result in small differences between raters in terms of their leniency and harshness as a group. The rater separation ratio provides a measure of the spread of the rater severity measures (Myford & Wolfe, 2004). The higher the rater separation ratio, the more the raters differed in terms of severity in their ratings.

### *Rater reliability:*

The third hypothesis was that a necessary condition for validity of a rating scale is rater reliability (Davies & Elder, 2005). A scale that results in higher levels of rater reliability can be seen as superior.

FACETS provides two measures of rater reliability: (a) the rater point-biserial correlation index (or single rater - rest of raters correlation), which is a measure of how similarly the raters are ranking the candidates, and (b) the percentage of exact rater agreement, which indicates the percentage of how many times raters awarded exactly the same score as another rater. Both types of rater reliability statistics were deemed necessary based on Stemler (2004), who cautioned against the use of just one statistic.

### *Variation in ratings:*

Because rating behaviour is a direct result of using a rating scale, it was further contended that a better functioning rating scale would result in fewer raters rating either inconsistently or overly consistently (by overusing the central categories of the rating scale). The idea behind this was that if a rater is unsure what level to award when using a rating scale, the rater might either rate inconsistently or resort to a play-it-safe method and overuse the inner categories of a rating scale and avoid the outside band levels.

The measure indicating variability in raters' scores is the rater infit mean square value. Rater infit means square values have an expected value of 1 and can range from 0 to infinity. The closer the calculated value is to 1, the closer the rater's ratings are to the expected ratings. Infit mean square values significantly higher than 1.3 (following McNamara, 1996 and Myford & Wolfe, 2000) denote ratings that are further away from the expected ratings than the model predicts. This is a sign that the rater in question is rating inconsistently. Similarly, values lower than .7 indicate that the observed ratings are closer to the expected ratings than the Rasch model predicts. This could indicate that a rater is rating very consistently. However, it is more likely that the rater is overusing certain categories of the rating scale, normally the inside values.

### Data analysis – interviews

The analysis of the interview data focussed on identifying recurring themes in the data. It was hoped that some of the findings of the quantitative analysis could be explained by the interview data. All data was coded by the main researcher, but double-coding of the themes was undertaken by a second researcher.

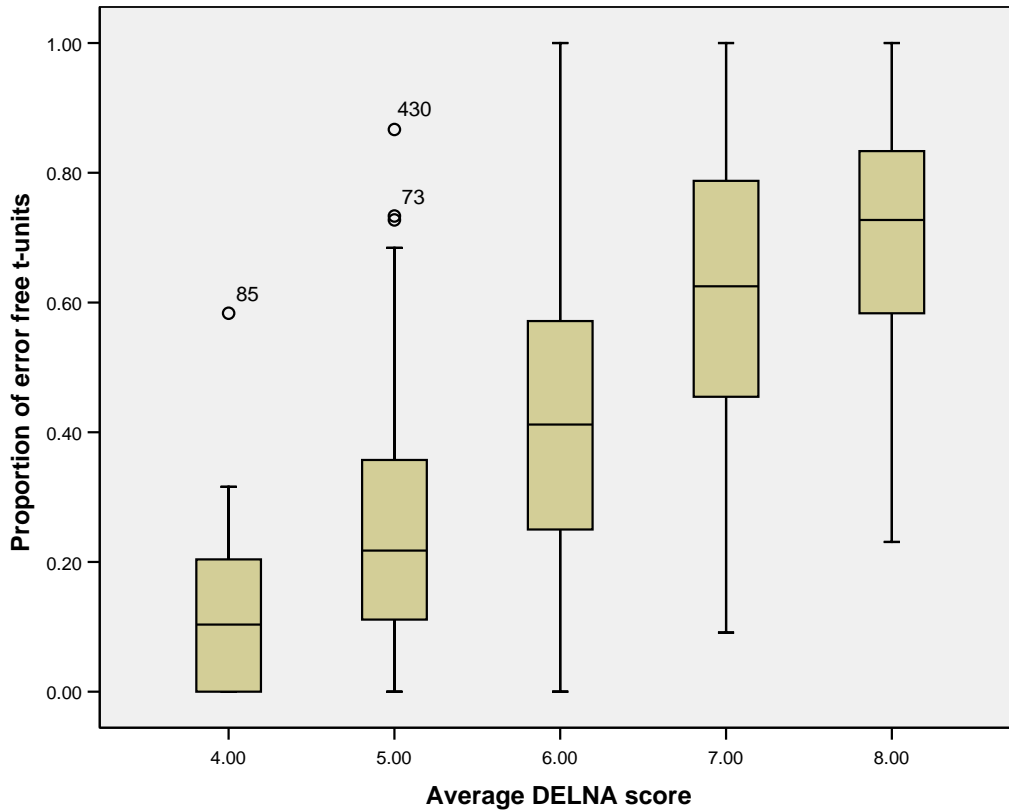
## **Results**

The results will be discussed following the three research questions.

### **1) What are the features of accuracy, fluency and complexity displayed at different levels of expository writing?**

Accuracy

Accuracy was measured as the *percentage of error-free t-units*<sup>2</sup>.



**Figure 1: Distribution of proportion of error-free t-units over overall sample and DELNA sublevels**

The side-by-side box plots in Figure 1 depict the distribution of the proportion of error-free t-units (vertical axis) over the different DELNA bands (horizontal axis). The variable successfully distinguished between the different levels, with some overlap.

**Table 1: Descriptive statistics - Proportion of error-free t-units**

DELNA level	M	SD	Minimum	Maximum
<b>4</b>	<b>.15</b>	<b>.18</b>	<b>0.00</b>	<b>0.58</b>
<b>5</b>	<b>.25</b>	<b>.18</b>	<b>0.00</b>	<b>0.87</b>
<b>6</b>	<b>.42</b>	<b>.24</b>	<b>0.00</b>	<b>1.00</b>
<b>7</b>	<b>.61</b>	<b>.22</b>	<b>0.09</b>	<b>1.00</b>
<b>8</b>	<b>.71</b>	<b>.21</b>	<b>0.23</b>	<b>1.00</b>

Table 1 above shows the descriptive statistics for each of the five proficiency levels. An ANOVA revealed significant differences between the different band levels,  $F(4, 576) =$

<sup>2</sup> Percentages are represented as proportions of 1 in the data below.



60.28,  $p < .001$ . The Games-Howell post hoc procedure showed statistically significant differences between two adjacent pairs of levels, levels 5 and 6 and levels 6 and 7.

### Trait scale: accuracy

The rating scale for accuracy (Table 2 below) was designed, so that the raters did not have to count each error-free t-unit. It was rather the idea that they would estimate the proportion of error-free t-units when reading a script. It was further decided that raters did not need to be trained to identify t-units in these data, because a brief analysis of t-unit borders showed that these coincided in over 90% of the cases with sentence breaks.

**Table 2: Rating scale – accuracy**

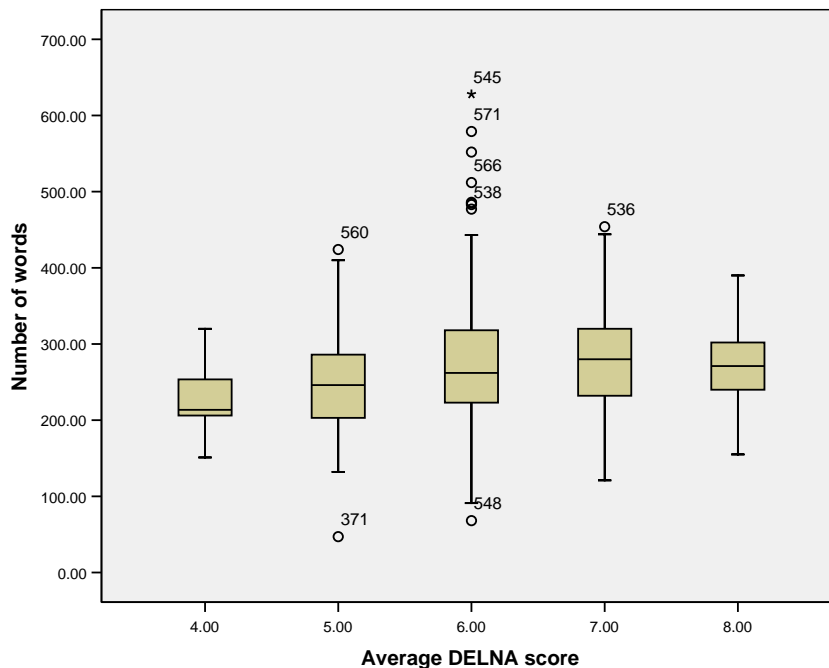
9	8	7	6	5	4
All sentences error-free	Nearly all or all sentences error-free	About $\frac{3}{4}$ of sentences error free	About half of sentences error-free	About $\frac{1}{4}$ of sentences error-free	Nearly no or no error-free sentences

Although the analysis of the scripts only showed five distinct levels of accuracy (because no scripts at level 9 were included in the analysis), a sixth level was added to the trait scale of accuracy to acknowledge completely error-free scripts.

### Fluency

#### Temporal Fluency

The variable chosen for temporal fluency was the *average number of words* per script.



**Figure 2: Distribution of number of words per essay over overall sample and DELNA sublevels**

**Table 3: Descriptive statistics – Average number of words per script**

DELNA level	M	SD	Minimum	Maximum
<b>4</b>	<b>226.67</b>	<b>42.86</b>	<b>151</b>	<b>320</b>
<b>5</b>	<b>244.98</b>	<b>58.63</b>	<b>47</b>	<b>424</b>
<b>6</b>	<b>273.66</b>	<b>79.18</b>	<b>68</b>	<b>628</b>
<b>7</b>	<b>281.00</b>	<b>67.69</b>	<b>121</b>	<b>454</b>
<b>8</b>	<b>273.32</b>	<b>54.33</b>	<b>155</b>	<b>390</b>

The box plots in Figure 2 and the descriptive statistics in Table 3 both indicate that although the average number of words generally increased as the writing level rose, there was much overlap. There also seemed to be a ceiling effect to the variable, indicating that writers at levels 6, 7 and 8 seemed to produce a very similar number of words on average. Whilst there was therefore a clear difference between the number of words produced on average between levels 4 to 6, for levels 6 and above the variable did not discriminate successfully between the writers.

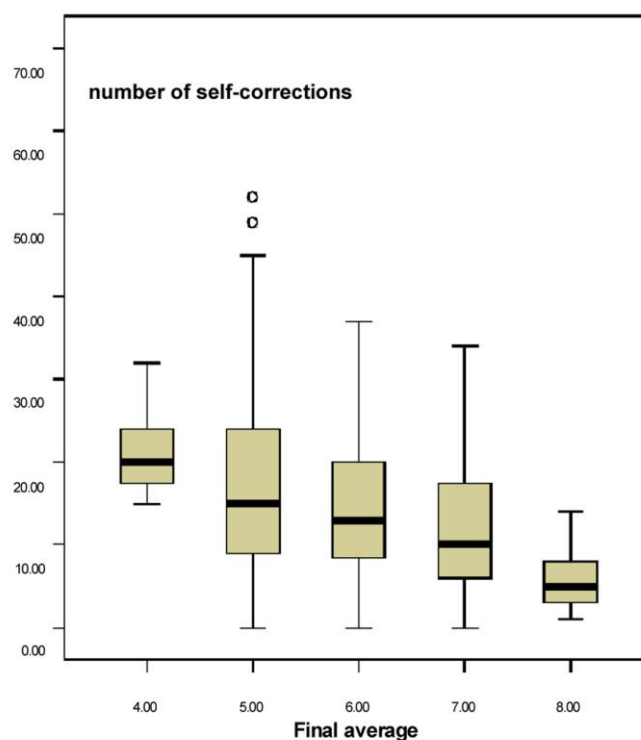
An ANOVA revealed a statistically significant difference between the five band levels,  $F(4, 577) = 5.82$ ,  $p < .001$ . The Games-Howell procedure revealed that the only adjacent levels that were significantly different were levels 5 and 6.

### **Trait scale: temporal fluency**

It was decided not to include temporal fluency into the rating scale because there was little evidence from the analysis of the scripts that there are differences between the different levels of writing in terms of the number of words that writers produce. The mean did not increase monotonically; there was a ceiling effect after level 6.

### *Repair Fluency*

The variable chosen for repair fluency was the *number of self-corrections*.



**Figure 3: Distribution of number of self-corrections over overall sample and DELNA sublevels**

Although there was considerable overlap, the measure discriminated between the different DELNA bands (see Table 4 below), showing that the lower the level of the writer, the more self-corrections were made<sup>3</sup>.

**Table 4: Descriptive statistics - Number of self-corrections**

DELNA level	M	SD	Minimum	Maximum
<b>4</b>	<b>21.33</b>	<b>5.19</b>	<b>0</b>	<b>32</b>
<b>5</b>	<b>17.21</b>	<b>11.41</b>	<b>0</b>	<b>52</b>
<b>6</b>	<b>15.00</b>	<b>9.58</b>	<b>0</b>	<b>64</b>
<b>7</b>	<b>12.38</b>	<b>9.57</b>	<b>0</b>	<b>57</b>
<b>8</b>	<b>6.96</b>	<b>5.84</b>	<b>0</b>	<b>37</b>

Because the assumption of equality of variances did not hold in this case, a Welch test was performed which revealed statistically significant differences between the different groups,  $F(4, 60.7) = 4.14, p = .005$ . However, the Games-Howell procedure revealed that no immediately adjacent levels were significantly different.

### **Trait scale: repair fluency**

On the basis of these findings, the rating scale for fluency (Table 5) was only based on the variable ‘number of self-corrections’. The scale largely followed the findings from the analysis. Levels were slightly adjusted to allow for better distinctions between bands. For example, band level 8 was designed to include no more than five self-corrections although the analysis of band 8 resulted in a mean of just under 7 and so on. As with accuracy, a sixth level (level 9) was added to the scale to acknowledge scripts with no self-corrections.

**Table 5: Rating scale - repair fluency**

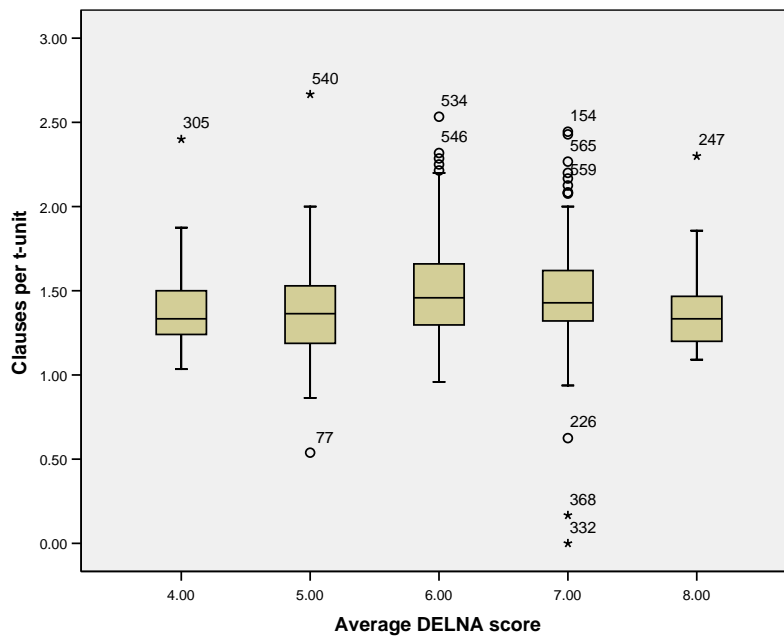
<b>9</b>	<b>8</b>	<b>7</b>	<b>6</b>	<b>5</b>	<b>4</b>
<b>No self-corrections</b>	<b>No more than 5 self-corrections</b>	<b>6-10 self-corrections</b>	<b>11 – 15 self-corrections</b>	<b>16-20 self-corrections</b>	<b>More than 20 self-corrections</b>

### Grammatical complexity

#### *Grammatical complexity*

The variable chosen to analyse grammatical complexity was *clauses per t-units*.

<sup>3</sup> A further analysis showed that if the variable is controlled for the number of words per script, the variable is even more discriminating between levels. For reasons of space it was not reproduced.



**Figure 4: Distribution of clauses per t-units over overall sample and DELNA sublevels**

The box plots in Figure 4 and the descriptive statistics in Table 6 below show that the variable failed to differentiate between scripts at different ability levels. This means that, in contrast to what was expected, higher level writers did not use more complex sentences (more subordination). Overall, very little subordination was used in the scripts as is indicated by the mean of 1.46 for all scripts included. That is, fewer than every second t-unit included subordination.

**Table 6: Descriptive statistics - Clauses per t-unit**

DELNA level	Mean	SD	Minimum	Maximum
<b>4</b>	<b>1.45</b>	<b>.39</b>	<b>1.03</b>	<b>2.40</b>
<b>5</b>	<b>1.39</b>	<b>.27</b>	<b>.54</b>	<b>2.67</b>
<b>6</b>	<b>1.50</b>	<b>.29</b>	<b>.96</b>	<b>2.53</b>
<b>7</b>	<b>1.48</b>	<b>.32</b>	<b>.00</b>	<b>2.44</b>
<b>8</b>	<b>1.42</b>	<b>.30</b>	<b>1.09</b>	<b>2.30</b>

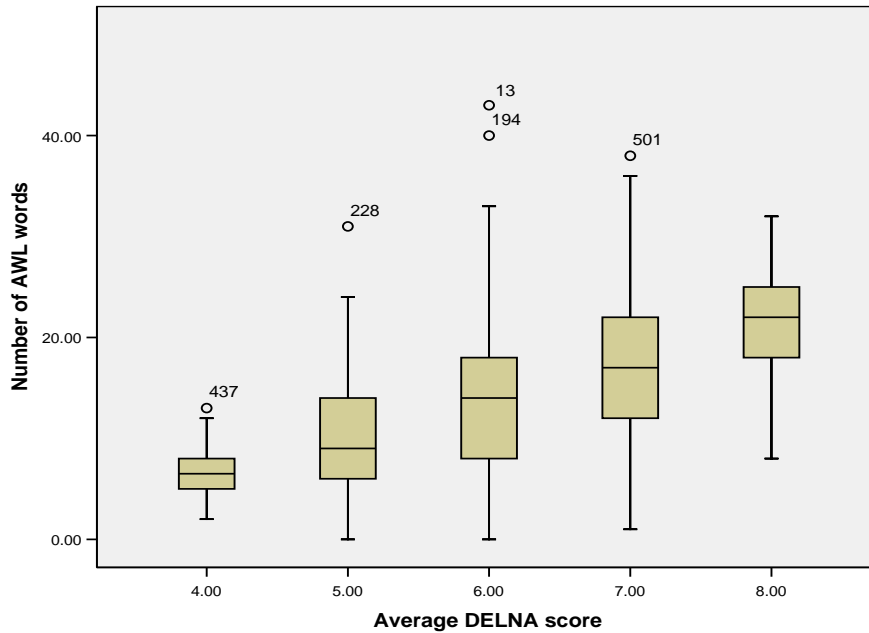
An ANOVA was performed which returned a statistically significant result,  $F(4, 575) = 3.08$ ,  $p = .016$ . The Games-Howell procedure showed that the only adjacent band level pair that was significantly different was level 5 and 6.

### Trait scale: grammatical complexity

Based on these findings, the decision was made not to include this variable into the rating scale as the changes of the variable as the levels increase is not linear.

### Lexical complexity

*The number of words in the Academic Word List (AWL)* was recorded as part of the analysis of VocabProfile. As the boxplots in Figure 5 and Table 7 indicate, this variable differentiates well between the different levels of writing<sup>4</sup>.



**Figure 5: Distribution of number of AWL words over overall sample and DELNA sublevels**

**Table 7: Descriptive statistics - Number of words in AWL**

DELNA level	M	SD	Minimum	Maximum
<b>4</b>	<b>6.91</b>	<b>3.09</b>	<b>2</b>	<b>13</b>
<b>5</b>	<b>10.25</b>	<b>5.97</b>	<b>0</b>	<b>31</b>
<b>6</b>	<b>13.99</b>	<b>7.69</b>	<b>0</b>	<b>43</b>
<b>7</b>	<b>17.11</b>	<b>7.71</b>	<b>1</b>	<b>38</b>
<b>8</b>	<b>21.24</b>	<b>6.09</b>	<b>8</b>	<b>32</b>

Because the assumption of equal variances was not satisfied in this case, a Welch procedure was performed, which revealed statistically significant differences between the groups,  $F(4, 66.22) = 39.99, p < .001$ . The Games-Howell procedure showed that all adjacent pairs of band levels differed significantly statistically.

<sup>4</sup> See Note 2

**Trait scale: lexical complexity**

It was decided that six levels of AWL words would be difficult to distinguish for raters. Therefore, only four levels were created, joining levels 4 with 5 and 8 with 9 (Table 8).

**Table 8: Rating scale – lexical complexity**

9	8	7	6	5	4
<b>Large number of words from academic wordlist (more than 20)</b>		<b>Between 12 and 20 AWL words</b>	<b>5 - 12 words from AWL</b>	<b>Less than 5 words from AWL</b>	

- 2) Do the ratings using the two different sets of trait scales relating to accuracy, fluency and complexity differ in terms of (a) the discrimination between candidates, (b) rater spread and agreement, (c) variability in the ratings, and (d) what the different traits measure?

The results for this research question will be discussed in two sections. The first relates to the analysis of individual trait scales (considering points a to c of the research question), whilst the second relates to the comparison of the rating scales as a whole (considering point d of the research question).

The first section below focuses on a comparison of individual trait scales. In this case, direct comparisons between trait scales on the DELNA scale and on the new scale are made where possible. It was, for example, possible to compare the trait scale for accuracy on the new scale with the trait scale of ‘grammatical accuracy’ on the DELNA scale directly. Where no direct comparisons were possible, the FACETS results for the individual trait scales will be presented alone.

*Accuracy scales*

The first two trait scales are those relating to accuracy (see Table 9 below).

It can be seen that the candidate separation ratio for the new scale was higher than that for the DELNA scale, which suggests that the new scale was more discriminating. The statistics indicating rater separation and reliability show that the raters rated more similarly in terms of leniency and harshness when using the new scale (indicated by the lower rater separation ratio) and ranked the candidates more similarly (rater point biserial) and also chose the same band level of the rating scale more often (percentage exact agreement).

**Table 9: Rating scale statistics for accuracy**

DELNA scale - Grammatical accuracy	New scale - Accuracy
<b>Candidate discrimination:</b> <b>Candidate separation ratio: 4.68</b>	<b>Candidate discrimination:</b> <b>Candidate separation ratio: 5.07</b>
<b>Rater separation and reliability:</b> <b>Rater separation ratio: 5.85</b> <b>Rater point biserial: .80</b> <b>% Exact agreement: 37.8%</b>	<b>Rater separation and reliability:</b> <b>Rater separation ratio: 3.71</b> <b>Rater point biserial: .91</b> <b>% Exact agreement: 46.8%</b>
<b>Variation in ratings:</b> <b>% Raters infit high: 20%</b> <b>% Raters infit low: 10%</b>	<b>Variation in ratings:</b> <b>% Raters infit high: 0%</b> <b>% Raters infit low: 0%</b>

Table 9 above also presents the percentage of unusually high or low infit mean square values exhibited by the raters. If, for example, two of the ten raters displayed very high infit mean square values, then the table indicates that twenty percent of raters showed this tendency. Whilst three raters displayed either unacceptably high or low infit mean square values when using the DELNA scale, no raters rated with too little or too much variation when applying the new scale for accuracy.

In summary, it can be argued that when the two accuracy scales were compared, all indicators point to the fact that the new scale functioned better.

#### *Vocabulary and spelling/lexical complexity scales*

Table 10 below shows a comparison of the four groups of statistics for the two rating scales focusing on lexis.

In this case, the discrimination of the new scale was only slightly greater than that of the existing scale. That the candidate separation of the new scale was higher than that of the existing rating scale is surprising given that the new scale had two fewer band levels. Because the discrimination of the two scales cannot be easily compared in this way, Linacre (personal communication, July 2006) offered the formula in Equation 1 below (an application of the Spearman-Brown Prophecy formula to this situation) to equate the separation ratios of two scales with differing band levels.

**Table 10: Rating scale statistics for vocabulary/spelling and lexical complexity**

DELNA scale – Vocabulary and spelling	New scale – Lexical complexity
<b>Candidate discrimination:</b> <b>Candidate separation ratio: 4.38</b>	<b>Candidate discrimination:</b> <b>Candidate separation ratio: 4.54</b>
<b>Rater separation and reliability:</b> <b>Rater separation ratio: 3.48</b> <b>Rater point biserial: .78</b> <b>% Exact agreement: 40.6%</b>	<b>Rater separation and reliability:</b> <b>Rater separation ratio: 6.65</b> <b>Rater point biserial: .85</b> <b>% Exact agreement: 49.7%</b>
<b>Variation in ratings:</b> <b>% Raters infit high: 20%</b> <b>% Raters infit low: 20%</b>	<b>Variation in ratings:</b> <b>% Raters infit high: 10%</b> <b>% Raters infit low: 10%</b>

Candidate separation new scale = square root (no. of levels in new scale – 1/ no. of levels in old scale – 1)* candidate separation old scale
---

**Equation 1: Equation used to predict the candidate separation of two scales with differing levels**

If the empirical candidate separation ratio (new scale) found in the FACETS analysis exceeds what the formula predicts, then the new scale is more discriminating. The result of the formula predicts that if converted to only four levels, the existing scale would only have a candidate separation ratio of 3.39. Therefore, the new scale was clearly more discriminating, even though it had fewer band levels.

As was found with the accuracy trait scales, both the rater point biserial correlation coefficient and the exact agreement were higher for the new scale. However, interestingly, the rater separation ratio indicates that the raters were more spread out in terms of severity when using the new scale. So, although they seemed to be ranking the candidates more similarly when using the new scale, the raters as a group were more varied in terms of severity.

When using the new scale, fewer raters rated either inconsistently or overly consistently (only 20% of raters compared to 40% of raters when applying the existing descriptors).

*Sentence structure scale*

The existing DELNA scale has a trait scale for sentence structure. The descriptors in this scale refer both to accuracy and complexity of sentences. Accuracy of sentences



is covered in the new scale in the trait scale for accuracy, whilst the analysis of the writing scripts showed no differences in grammatical complexity between the different levels of writing. Therefore, for completeness of the results section, the scale statistics for the DELNA sentence structure trait scale are presented here.

**Table 11: Rating scale statistics for sentence structure**

DELNA scale – Sentence structure
<b>Candidate discrimination:</b> <b>Candidate separation ratio: 4.16</b>
<b>Rater separation and reliability:</b> <b>Rater separation ratio: 3.90</b> <b>Rater point biserial: .77</b> <b>% Exact agreement: 40.2 %</b>
<b>Variation in ratings:</b> <b>% Raters infit high: 20%</b> <b>% Raters infit low: 20%</b>

Table 11 above shows that the trait scale was discriminating. The rater separation ratio indicates that the raters were not all rating exactly alike in terms of severity. The single rater - rest of raters correlation (point biserial) was high, as was the exact agreement. However, of the ten raters, four rated with either too much variation (inconsistently) or with too little variation (underusing the extreme levels of the scale).

### *Repair fluency scale*

One trait scale not found in the existing rating scale, but included in the new scale, was the trait scale for repair fluency. Table 12 below displays the rating scale statistics for this new scale.

The candidate separation ratio indicates that the discrimination of this new scale was high. The inter-rater reliability, as indicated by the point biserial correlation, was also high (.93), more so than for all other trait scales examined so far. The same can be said for the exact agreement (61.9%). The rater separation ratio, however, indicates large differences in severity between the most severe and the most lenient rater in the group.

Nearly half of the raters were identified as rating either inconsistently or too consistently.

**Table 12: Rating scale statistics for repair fluency**

New scale – Repair fluency
<b>Candidate discrimination:</b> <b>Candidate separation ratio: 5.82</b>
<b>Rater separation and reliability:</b> <b>Rater separation ratio: 5.34</b> <b>Rater point biserial: .93</b> <b>% Exact agreement: 61.9%</b>
<b>Variation in ratings:</b> <b>% Raters infit high: 20%</b> <b>% Raters infit low: 20%</b>

**Comparison of whole rating scales:**

It was further of interest how the rating scales as a whole compared. The FACETS analysis gave some indication that the trait scale on the existing DELNA scale were highly correlated and this was not the case with the trait scales on the new scale. To investigate this in more detail, and to establish which trait scales correlated with which (information which FACETS does not provide), a Spearman correlation was performed for each scale. The differences between the two rating scales were striking. Whilst, as the FACETS analysis suggested, the trait scales on the existing scale resulted in correlations ranging from .825 to .851 (Table 13), the correlations of the scores on the new scale were significantly lower, only ranging from .202 to .567 (Table 14). This suggests that the traits on the new scale measure different aspects of writing, whilst this is not the case for the traits on the existing DELNA scale. A discussion of this finding will follow later.

**Table 13: Correlations for existing DELNA scale**

	<b>Sentence Structure</b>	<b>Grammatical accuracy</b>	<b>Vocabulary and spelling</b>
<b>Sentence structure</b>	<b>1.00</b>		
<b>Grammatical accuracy</b>	<b>.833**</b>	<b>1.00</b>	
<b>Vocabulary and spelling</b>	<b>.825**</b>	<b>.851**</b>	<b>1.00</b>

\*\* = Statistically significant ( $p = .000$ )

**Table 14: Correlations for new scale**

	<b>Accuracy</b>	<b>Repair fluency</b>	<b>Lexical complexity</b>
<b>Accuracy</b>	<b>1.00</b>		
<b>Repair fluency</b>	<b>.253**</b>	<b>1.00</b>	
<b>Lexical complexity</b>	<b>.567**</b>	<b>.202**</b>	<b>1.00</b>

\*\* = Statistically significant (p = .000)

### 3) What are raters' perceptions of using measures of accuracy, fluency and complexity in a rating scale as compared to more conventional rating scales?

The third and final research question set out to establish the raters' perceptions of the two different rating scales. Seven of the ten raters were interviewed after both rating rounds were completed.

One of the most commonly emerging themes from the interviews was that the raters thought the descriptors in the DELNA scale were often too vague for them to arrive easily at a score. In the extract below, Rater 3 describes his problems when choosing the appropriate band level<sup>5</sup>:

Rater 3: [...] sometimes I look at [the descriptor] I'm going 'what do you mean by that?' [...] You just kind of have to find a way around it cause it's not really descriptive enough, yeah

Almost all raters reported similar sentiments. The most common way of dealing with the vagueness of the DELNA scale was that raters started rating holistically, that is, instead of awarding different scores for each category on the trait scale, they let their overall impression of the whole script guide their scoring on individual trait categories. This can, for example, be seen in the extract below:

Rater 10: Well there's always a bit of a problem with [the category of] vocabulary anyway in deciding you know the difference between *extensive*, *appropriate*, *adequate*, *limited* and *inadequate*. [...] so I think I just go with gut instinct on that one.

What was interesting, however, was that the raters did not resort to this strategy when rating with the new rating scale. In the extract below, for example, Rater 5 describes how the ratings on different categories on the rating scale were more differentiated when using the new scale.

Interviewer: So you said you sometimes form a global impression with the DELNA scale, did you ever do that with this scale?

<sup>5</sup> All extracts from interviews presented here were chosen to be representative of the findings from the majority of the transcripts.

Rater 5: no, I don't think I could have, I don't think I did, really, no it was all over the place. I mean with one script it tend to come in lots of things, so I couldn't really say this was a six

When the raters were using the new scale, they did not know how an overall score would be computed from the sub-scores on the rating scale (because this was not a focus of the study). This seemed to help the raters to rate analytically rather than holistically. The extract below is an example:

Interviewer: Did it bother you that you didn't know what the final score was going to be with the new scale

Rater 10: actually, no, it didn't bother me. I guess I just gave up that idea of this script is really a six. Maybe I still had it in my head, so I'd say, oh well, this is really a seven, but they got a four for, you know, paragraphing. Four for whatever it was that they scored low on.

## **Discussion**

DELNA is a diagnostic assessment system. To establish construct validity for a rating scale used for diagnostic assessment, we need to turn to the limited literature on diagnostic assessment. Alderson (2005), as reported in the introductory section, compiled a list of features which distinguish diagnostic tests from other types of tests, four of which are central to this study. Each of these four statements will now be discussed in turn.

Alderson's first statement calls for diagnostic assessments to identify strengths and weaknesses in a learner's knowledge and use of language. Both rating scales compared in this study were analytic scales and were designed to identify strengths and weaknesses in different aspects of the learners' writing ability. However, the correlational analysis showed that the new scale was more successful in distinguishing between different aspects of writing than the existing DELNA scale. Therefore, it could be argued that the new scale was more successful in identifying different strengths and weaknesses.

Although developed as an analytic scale, the existing DELNA scale seemed to lend itself to a more holistic approach to rating. It is possible, as hypothesised in this study, that the rating scale descriptors do not offer raters sufficient information on which to base their decisions and so raters resort to a global impression when awarding scores. This would explain why, when using the empirically developed new scale with its more detailed descriptors, the raters were able to discern distinct aspects of a candidate's writing ability.

Some studies have in fact found that raters display halo effects only when encountering problems in the rating process (e.g. Lumley, 2002; Vaughan, 1991). Lumley, for example, found that when raters could not identify certain features in the descriptors, they would resort to more global, impressionistic type rating. This study

suggests that the halo effect and impressionistic type marking might be more widespread than has so far been reported. The halo effect is usually seen in the literature as being a rater effect that needs to be reduced or even eliminated by rater training. However, Cascio (1982, cited in Myford and Wolfe, 2003, p. 396) notes that the halo effect is ‘doggedly resistant to extinction’. This study has shown that simply providing raters with more explicit scoring criteria can significantly reduce this effect. It could therefore be argued that the halo effect is not necessarily only a rater effect, but also a rating scale effect.

However, it was not established in this study whether the raters rated analytically because they were unfamiliar with the new scale. It is possible that extended use of the new scale might also result in more holistic rating behavior. This point will be taken up in the suggestions for further research.

Alderson’s (2005) second and third statements assert that diagnostic assessments should enable a detailed analysis and report of responses to tasks and that this feedback should be in a form that can be acted upon. Both rating scales lend themselves to a detailed report of a candidate’s performance. However, as evident in the quantitative analysis, if the raters at times resort to a holistic impression to guide their marking when using the DELNA scale, this will reduce the amount of detail that can be provided to students. If most scores are, for example, centred around the middle of the scale range, then this information is less useful to students than if they are presented with a more jagged profile of some higher and some lower scores which therefore affords a clear indication of which aspects of their writing they need to focus on.

Alderson’s fourth statement states that diagnostic tests are more likely to be focussed on specific elements rather than on global abilities. If a diagnostic test of writing is aimed at focussing on specific elements, this needs to be reflected in the rating scale. Therefore, the descriptors need to lend themselves to isolating more detailed aspects of a writing performance. The descriptors of the new scale were more focussed on specific elements of writing because they were based on discourse analytic measures. The band descriptors on the DELNA scale generally reflect more global abilities, with vaguer, more general band descriptors.

The way the scores are reported is also important. It is not effective to use an analytic scale and then average the scores when reporting back to stakeholders, because this will result in a more global impression of the performance and important information is therefore lost. Currently the writing scores are reported to test takers as one averaged score with brief accompanying descriptions about their performance in fluency, content and form. Academic departments only receive one averaged score. None of this advice, however, focuses on details of their writing performance. In this way, the current practice is more representative of proficiency tests or placement tests.

Finally, it was also important to establish the stakeholders’ perceptions of the efficacy of the two scales for diagnostic assessment. Only the raters’ opinions were determined.

Most raters reported encountering problems when using the DELNA descriptors. Almost all of these comments were related to the descriptors being too vague or non-specific for raters to be able to easily decide on a score. One reason mentioned in this respect is the use of adjectives like ‘*extensive*’, ‘*appropriate*’ or ‘*adequate*’. Raters were very aware that these could mean different things to different raters. This problem has also been pointed out by a number of researchers (for example Brindley, 1998; Mickan, 2003; Upshur & Turner, 1995; Watson Todd, Thienpermpool, & Keyuravong, 2004). Furthermore, recent evidence from think-aloud protocols of the rating process lends support to the fact that raters struggle with vague descriptors. Smith (2000), for example, found that raters had ‘difficulty interpreting and applying some of the relativistic terminology used to describe performances’ (p. 186). Shaw (2002) noted that about a third of the raters he interviewed reported problems when using the criteria but he did not specify what specific problems they encountered. Similarly, Claire (cited in Mickan, 2003) reported that raters regularly debate the rating scale descriptors in rater training sessions and describe problems in applying descriptors with terms like ‘*appropriately*’. It can therefore be said that there is a growing body of research available that supports the results obtained for the interviews. Raters do often seem to find the descriptors vague and consider this to be problem.

The rater comments about the new scale as a whole were generally positive. The raters liked the fact that the level descriptors were more explicit and objective and provided more guidance than the descriptors raters were used to. They reported that it was much easier for them to ‘let go’ of impressionistic marking.

## Conclusion

This study set out to establish whether a rating scale based on empirical investigation of actual writing samples and with detailed level descriptors might be more suitable for the diagnostic assessment of writing than a pre-existing measure.

The results showed that the empirically-developed rating scale resulted in a more differentiated score profile which is therefore more useful in a diagnostic context. The ratings on the individual trait scales were also found to be more reliable and discriminating than those using the existing rating scale.

The study has some clear limitations. There were only ten raters in this study, which is a relatively small sample size. A larger group of raters would have added to the generalisability of the study. Also, it would have been preferable to have a fully counter-balanced design for this study. Some raters should have rated half of the scripts using the new scale first and then the DELNA scale, and the other half should have started using the DELNA scale. As mentioned earlier, this was not possible for practicability reasons, however the raters did mention that they were not able to remember any of the 100 writing scripts from the first rating round.

There are three implications for the current study. The first relates to rating scale development. Rating scale developers need to keep the context of the assessment in

mind. When the purpose of the writing assessment is to establish the overall proficiency of a student, then a rating scale with vaguer level descriptors is sufficient. However, if we want to provide students with detailed score profiles, as is the case in a diagnostic context, then a rating scale with more detailed level descriptors seems to be more successful because, as this study was able to establish, raters use the scale more analytically.

The second implication relates to score reporting. In a diagnostic context, the scores on an analytic rating scale need to be reported to stakeholders separately, so that as much information as possible is related. There is, therefore, no room for averaged scores.

The final implication is concerned with rater training. Raters need to be made aware of the importance of rating analytically. This is especially important when the assessment context is diagnostic. So whilst the descriptors were shown in this study to influence the rating behaviour and processes, rater training needs to also play an important role in raising the awareness of raters.

This study has provided an initial piece of evidence that rating scales used in diagnostic assessment should be designed in a different manner than those used in proficiency testing. More research is necessary, however. For example, it is not clear if the raters in this study would have used the new scale in a different way if they would have been used to it. It is possible that raters, after having used a rating scale for a while, resort back to holistic rating. This needs to be established in further research. It would also be useful to canvass the ideas of stakeholder groups other than the raters. For example, what type of feedback would students find helpful and what type of feedback is effective? How detailed should the feedback be and what categories on the rating scale are less useful for the different stakeholder groups involved. Further research is clearly necessary.

## References

- Alderson, C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment*. London: Continuum.
- Brindley, G. (1998). Describing language development? Rating scales and SLA. In L. F. Bachman & A. D. Cohen (Eds.), *Interfaces between second language acquisition and language testing research*. Cambridge: Cambridge University Press.
- Cascio, W. F. (1982). *Applied psychology in personal management*. Reston, VA: Reston Publishing Company.
- Cobb, T. (2002). Web Vocabprofile. Retrieved 12 December 2005, from <http://www.lex tutor.ca/vp/>
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213-238.
- Davies, A., & Elder, C. (2005). Validity and validation in language testing. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning*. Mahwah, NJ: Lawrence Erlbaum.
- Elder, C., Knoch, U., Barkhuizen, G., & von Randow, J. (2005). Individual feedback to enhance rater training: Does it work? *Language Assessment Quarterly*, 2(3), 175-196.
- Ellis, R., & Barkhuizen, G. (2005). *Analyzing learner language*. Oxford: Oxford University Press.
- Ellis, R., & Yuan, F. (2004). The effects of planning on fluency, complexity, and accuracy in second language narrative writing. *Studies in Second Language Acquisition*, 26, 59-84.

- Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition*, 18, 299-323.
- Fulcher, G. (1996). Does thick description lead to smart tests? A data-based approach to rating scale construction. *Language Testing*, 13(2), 208-238.
- Fulcher, G. (2003). *Testing second language speaking*. London: Pearson Longman.
- Iwashita, N., McNamara, T., & Elder, C. (2001). Can we predict task difficulty in an Oral Proficiency Test? Exploring the potential of an information-processing approach to task design. *Language Learning*, 51(3), 401-436.
- Knoch, U., Read, J., & von Randow, J. (2007). Re-training writing raters online: How does it compare with face-to-face training? *Assessing Writing*, 12, 26-43.
- Linacre, J. M. (2006). Facets Rasch measurement computer program. Chicago: Winsteps.
- Lumley, T. (2002). Assessment criteria in a large-scale writing test: What do they really mean to the raters? *Language Testing*, 19(3), 246-276.
- McNamara, T. (1996). *Measuring second language performance*. Harlow, Essex: Pearson Education.
- Meisel, J., Clahsen, H., & Pienemann, M. (1981). On determining developmental stages in natural second language acquisition. *Studies in Second Language Acquisition*, 3, 109-135.
- Mickan, P. (2003). 'What's your score?' An investigation into language descriptors for rating written performance. Canberra: IELTS Australia.
- Myford, C. M., & Wolfe, E. W. (2000). *Monitoring sources of variability within the Test of Spoken English assessment system*. Princeton, NJ: Educational Testing Service.
- Myford, C. M., & Wolfe, E. W. (2004). Detecting and measuring rater effects using many-facet Rasch measurement: Part II. *Journal of Applied Measurement*, 5(2), 189-227.
- O'Loughlin, K., & Wigglesworth, G. (2003). Task design in IELTS academic writing Task1: The effect of quantity and manner of presentation of information on candidate writing. *IELTS Research Reports*, 4, 89-129.
- Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition*, 14, 357-385.
- Shaw, S. D. (2002). IELTS writing: Revising assessment criteria and scales (Phase 1). *Cambridge Research Notes*, 9, 16-18.
- Skehan, P. (2003). Task-based instruction. *Language Teaching*, 36, 1-14.
- Smith, D. (2000). Rater judgments in the direct assessment of competency-based second language writing ability. In G. Brindley (Ed.), *Studies in immigrant English language assessment*. Sydney: National Centre for English Language Teaching and Research.
- Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research and Evaluation*, 9(4), 1-20.
- Upshur, J. A., & Turner, C. E. (1995). Constructing rating scales for second language tests. *ELT Journal*, 49(1), 3-12.
- Upshur, J. A., & Turner, C. E. (1999). Systematic effects in the rating of second-language speaking ability: Test method and learner discourse. *Language Testing*, 16(1), 82-111.
- Vaughan, C. (1991). Holistic assessment: What goes on in the rater's mind? In L. Hamp-Lyons (Ed.), *Assessing second language writing in academic contexts*. Norwood, NJ: Ablex Publishing Corporation.
- Watson Todd, R., Thienpermpool, P., & Keyuravong, S. (2004). Measuring the coherence of writing using topic-based analysis. *Assessing Writing*, 9, 85-104.
- Weigle, S. C. (2002). *Assessing writing*. Cambridge: Cambridge University Press.
- Young, R. (1995). Discontinuous interlanguage development and its implications for oral proficiency rating scales. *Applied Language Learning*, 6, 13-26.



# **THE EFFECT AND THE INFLUENCE OF THE USE OF VIDEO AND CAPTIONS ON SECOND LANGUAGE LEARNING**

**Judy Chai & Rosemary Erlam**

*The University of Auckland*

## **Abstract**

*The benefits of using video and captions for improving general L2 reading and listening comprehension have been well documented, however what is lacking is research that explores what contribution they may make to learning beyond just comprehension. This paper presents a study that aimed to address this gap by investigating how video plus captions impacted on the learning of second language words and phrases.*

*Twenty Chinese learners of English participated in the study. Participants were divided into two groups; one group watched a short video sequence with captions and the other the same sequence with no captions. All were pretested prior to the treatment and completed a posttest and delayed posttest. Participants also completed an interview. Results supported a positive correlation between the presence of captions and learning of unknown words and phrases, suggesting that the use of captions does enhance micro-level learning. In the interview, participants were positive in their responses to the use of captioned video in language learning contexts. The majority of those in the captions group reported that they prioritized the reading of captions in watching the video, some reported that they were unable to pay attention to both sound and pictures at the same time.*

## **Introduction**

With the evolution of information technology and its impact on society and education, exploiting and combining audio and visual aids to stimulate and enhance learning has become a trend in second/foreign language instruction over the years. Using video (with or without captions) is regarded as a major way to supplement target language input to learners, especially for those in an EFL setting.

## **Information processing theories**

Research pertaining to the effect of video and captioned materials on second language learning is based on theories of how information is processed by humans. Proponents of the *Single Channel Hypothesis* (Broadbent, 1958; Travers, 1967; Simatos & Spencer, 1992) claim that a human's perceptual system generally functions as a single channel system with information from only one channel gaining access to it at any given time. On the other hand, the *Dual Coding Theory* postulates the existence of two different storage systems for information in human cognition,

one nonverbal and the other verbal, which are specialized for imagery and language processes respectively (Chun & Plass, 1997). Information is much easier to retain and retrieve when dual-coded because of the availability of two mental representations instead of one (D'Ydewalle & Van Rensbergen, 1989; Paivio, 1986). This theory fits well with an understanding of working memory as a system with a limited storage capacity (Baddeley, 1992). Presenting material bimodally (i.e., combining visual and verbal modes) may enlarge working memory capacity, leading to improved learning outcomes. Research on L1 learning has confirmed the advantage of dual mode presentation and demonstrated that the use of multimedia and captions has a positive influence on language learning (Goldman, 1996; Goldman & Goldman, 1988). It has also shown redundant and meaningful visual images to be beneficial for enhancing learning as they can clarify or serve as reference points for speech or written text. As the maxim says, "A picture is worth a thousand words". The benefits of using captions and video in a L2 context has also been demonstrated. Research has shown an impact on reading and listening comprehension, listening and word recognition, as well as incidental vocabulary learning (Garza, 1991; Danan, 1992; Neuman & Koskinen, 1992; Borrás & Lafayette, 1994; Guillory, 1998; Chung, 1996; Baltova, 1999; Markham, 1999).

### **From comprehension to acquisition**

While past research has shown that ESL/EFL learners using video and captioned materials as 'comprehensible input' show significant improvement in listening and reading comprehension and incidental vocabulary acquisition, studies have not shown to what extent this input can help learners turn unknown target language forms into known and acquired forms. As many SLA researchers have observed, in contexts where learners' attention is focused primarily on comprehending the meaning of a passage or a conversation learners tend to act on a principle of 'least-effort' and fail to notice features of the target language which differ from their level of current knowledge and the use of that language (Schmidt, 1990; Ellis, 1997, 2003; Lightbown et al., 2002) since "the kind of processing necessary for comprehension is different from the kind of processing required for acquisition" (Lightbown et al., 2002: 459). The study presented in this paper departs from the research reported above in that it seeks to address whether video and captions can lead to learning of new linguistic knowledge, for example, words and phrases (micro-level learning) beyond just comprehension of content (macro-level learning).

For learning to occur it is generally acknowledged that learners need to notice target forms in language input (Gass, 1988). According to Schmidt (1990) there are a number of ways in which learners may be helped to do this; for example, they may be given a specific goal by a teacher, cuing them to attend to certain language features. The interest in attention to input as a key to development of a learner's interlanguage system has led a number of researchers to explore strategies that may induce learners to heed new language. One of these is Batstone's (2002a) 'given-to-new' principle where learners are facilitated to make sense of new language as a result of using contextual or cotextual cues and prior knowledge. Batstone (2002a) claims that

learners can use these cues to “infer the meaning of various lexical items” (p.7) or even the function of new grammar. For vocabulary learning this principle would seem to be particularly effective as there is general consensus that the use of contextual cues is a good strategy for learning and retaining new vocabulary (Husltijn, 1992; Ying, 2001) and research evidence to show that learners can derive the meanings of new words with the help of surrounding context (Nation, 1990; Ruhe, 1996). However, as Batstone (2002a, 2002b) points out, for learning to occur, learners must not merely be asked to get the gist of a given text or sentence by exploiting cues (so that they avoid the need to engage any further with the new language), they must interact with surrounding cues and pay attention to a novel linguistic item and connect the unknown with what is already known. Through this process they will make form/function mappings – connections between language forms and the meanings they convey.

### **Potential advantages and disadvantages of using video and captions**

A number of researchers have documented the strong motivational potential of the use of video in a language learning context. Vanderplank (1988), Secules et al. (1992), Stempleski (1992) and Gruba (2005) note that, given the choice between audio and video, language learners prefer learning via watching video. Furthermore, research suggests that, once people have adjusted to captions, processing them appears to be an effortless activity (D’Ydewalle et al, 1991; D’Ydewalle & Gielen, 1992).

On the other hand, there are also a number of potential disadvantages. Firstly, while in today’s world people are increasingly exposed to captions on TV/at cinemas and are thus used to processing them, the experience of reading captions in a L2 is another matter. Captions are presented at a rate of approximately 120 words a minute, providing a challenge to even the most accomplished developing readers (Spache, 1981). There is also the potential that reading captions may detract attention from other visual and/or auditory information, meaning that viewers may miss out on valuable clues. While, as discussed above, there is evidence that presenting information bimodally (i.e., visually and verbally) may enhance learning outcomes, some have questioned whether learners are able to cope with reading, viewing and listening all at the same time (Williams & Snipper, 1990). One may then wonder to what extent reading captions may mean that students do not have the chance to develop or practice other language skills, for example, listening comprehension. Vanderplank (1990) calls for research that investigates how text, sound and vision are used by learner viewers.

### **The present study**

The present study investigates whether input in the form of video and captions can help learners turn unknown target language forms into known and acquired forms. It looks at whether learners can infer the meaning of new words and whether learners

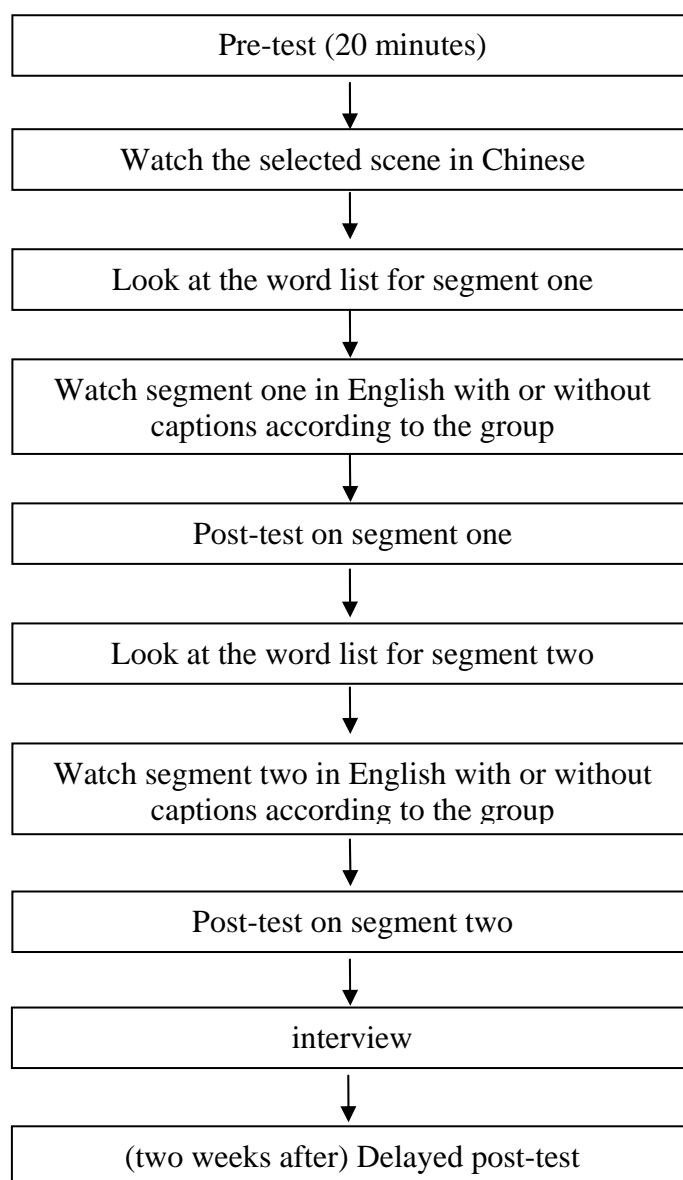
can be motivated to engage with this input and make it become ‘intake’ instead of simply comprehending it.

The research questions are:

1. Does video and captions lead to better learning and retention of unknown target words/phrases than video alone?
2. Among the three linguistic features targeted (i.e., single-words, phrasal verbs and phrases), which are the most effectively learnt by the use of video plus captions?
3. How do learners feel about learning English through video and captions?
4. What information do students pay most attention to in watching video with captions: captions, sound, pictures?

### Procedure

A total of 20 Chinese native speakers whose English proficiency was upper-intermediate, took part in the study. They were randomly divided into two groups: a *no captions group* and a *captions group*. Each participant filled out a background questionnaire prior to viewing the video, and then completed a pretest which examined their prior knowledge of targeted words and phrases. Participants then watched the movie excerpt on a computer with a pair of headphones in their native language (Chinese). The provision of L1 audio was intended as preliminary ‘scaffolding’ to assist learners to comprehend the story line since the present study focuses on words and phrases (micro-level) rather than on content comprehension (macro-level). The researcher made sure each participant understood the story line content completely after the viewing by asking some comprehension questions in Chinese. Participants were then shown a list of words and phrases (Appendix 1) and told that these would appear in the scene they were going to watch and that they needed to try and listen for these words and the sentences in which they occurred. They had two minutes to look at the list before watching the same excerpt again in English, with or without captions. This time the scene was divided into two segments so that the participants did not have the burden of having to remember all the target words at once. In other words, participants watched each segment in English with or without captions and then did a posttest on that segment straightaway. Participants then participated in a short interview. A final posttest was administered after two weeks. The overall procedure of the study is illustrated in Figure 1.



**Figure 1: Procedure of the present study**

### **Participants**

The participants were 20 ESL students in two language schools in Auckland, designated as upper-intermediate by their teachers. Their band scores on the IELTS ranged from 5 to 6. It was hypothesized that students of this level were appropriate for this study because they would have less difficulty comprehending the meaning of words/sentences containing complex language and ascertaining the meaning of target words with textual and visual support than students of lower proficiency.

There were 8 females and 12 males. All participants had Chinese nationality and their average age was 22. The mean age at which they started learning English was 11 and the average number of years they had spent living in an English speaking country was 2 years 8 months. All participants reported that they had not seen the movie chosen for this study.

### **The video**

A scene of about 10 minutes in length was selected from the DVD movie *Howl's Moving Castle*. See Appendix 2 for the script of the film. Mandarin, Cantonese and English dubbed versions were used since there were 4 participants whose native language was Cantonese. Several criteria suggested by Arcario (1992) for selecting a scene were taken into consideration, as listed below:

- (a) degree of visual support – the verbal messages need to be supported and contextualized by the visual elements
- (b) clarity of picture and sound – both need to be of good quality
- (c) speech delivery – speech needs to be clear and the accent should not depart significantly from the standard accent of television broadcasters
- (d) language content and level – the content of a scene must not be offensive and the complexity of linguistic items needs to be considered
- (e) length of sequence – the recommendation of optimal segment length is from 30 seconds to 12 minutes long (Willis, 1983, cited in Arcario, 1992).

The scene selected for this study was 10 minutes long and was kept entirely intact as it appeared in the full-length movie.

### **The target words/phrases**

The test was aimed to assess participants' knowledge of words, phrasal verbs and phrases. Phrasal verbs and phrases were included as target items because they are extremely common in English. Furthermore, they are difficult for ESL learners because they are rarely transparent in meaning. All target items in this study were chosen from the dialogue of the selected scene following advice from three experienced ESL teachers. These teachers were asked to identify any vocabulary, syntax and phrases which L2 learners of English at upper proficiency level would be unlikely to know. Only those items with which two out of the three teachers were in agreement were included as target words/phrases. Two other key considerations were taken into account in the selection of each test item. Firstly, there needed to be a clear link between verbal (audio and/or captions) and non-verbal (images) so that the item would likely be encoded dually during viewing (Garza, 1991; Danan, 1992) and secondly, only those items that were spoken clearly by the characters were selected. There were a total of 21 target words and phrases (8 nouns and 2 adjectives, 6 phrasal verbs, 5 phrases).

### **The test**

Two tests were designed to be used as equivalent versions – one was used as the pretest and delayed posttest, the other was used as the posttest. There were seven sections to the test and a total of 45 items (15 items testing words, 19 testing phrasal verbs and 11 testing phrases). The test was designed to measure receptive rather than productive knowledge of the target items. Individual sections tested the ability to match words with their definitions, the ability to choose words from a list to complete sentences, the ability to choose the preposition that collocated correctly with a phrasal

verb and the ability to give definitions (either in English or the L1) of idioms. See Appendix 3 for examples of test items. Reliability of the tests was calculated using Cronbach's alpha over all testing episodes. Results were as follows: pretest,  $\alpha = .183$ ; posttest,  $\alpha = .779$ ; delayed posttest,  $\alpha = .754$ . It was hypothesized that the reason for the low reliability estimate of the pretest is because of a low SD (3.07); given that participants did not know the target words/phrases, there was not much variation on the pretest.

### The interview

All the participants were asked to participate in a short semi-structured interview to discuss how they had proceeded with the task of working out the meaning of the target words/phrases and what they thought about learning English through video and captions. (see Appendix 4 for interview questions). Participants could choose to answer the questions in English or in their L1. Participants' responses were audio-taped, translated (if necessary), and transcribed afterwards by the researcher.

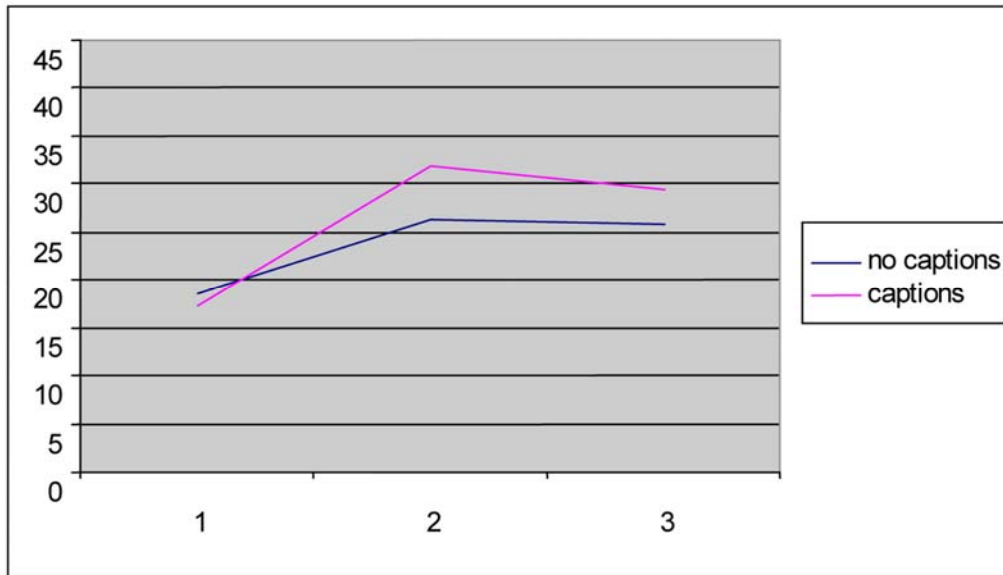
## Results

Descriptive statistics were calculated for each testing episode. A series of 2 x 3 mixed factorial design split plot ANOVA was performed with type of treatment (between subjects) and time of test (within subjects) as independent variables and test scores as the dependent variable. In addition, in order to compare the *no-captions group's* gain with the *captions group's* gain from pretest to posttest the pretest score was subtracted from the posttest score for each participant. Then an independent-samples *t* test was performed to test whether the means of the two treatment groups' change scores differed. Table 1 presents descriptive statistics for both the *no-captions group* and the *captions group*.

**Table 1: Descriptive statistics of overall test scores**

		no- captions			captions	
		n = 10			n = 10	
	Mean	%	SD	Mean	%	SD
pre-test	18.50	41.11	2.99	17.20	38.22	3.16
post-test	26.20	58.22	3.46	31.80	70.67	6.43
delayed	25.80	57.33	4.18	29.40	65.33	5.85

Figure 2 shows the gains that both groups made over the three testing episodes. Although both groups made gains from pretest to posttests, the *captions group's* gain was larger than that of the *no captions group*.



**Figure 2: Gains made by both groups over three testing episodes**

Table 2 presents the results of the split plot ANOVA. As can be seen from the table, the interaction between time and treatment was significant ( $p = .000$ ) as was the main effect for time ( $p = .000$ ). On the other hand, the main effect for treatment did not reach significance. This tells us that if we ignore all other variables, the *captions group's* overall performance was not significantly different from that of the *no captions group*.

**Table 2: Results of split plot ANOVA for the overall test scores**

	df	F	Sig.	Partial Eta Squared
time	2	146.21	.000*	.890
treatment	1	2.01	.174	.100
time * treatment	2	12.49	.000*	.410

Table 3 shows post hoc pairwise comparisons for the within-subjects contrast. For both treatment groups, significant differences were found ( $p < .0005$ ) between pretest and posttest as well as between pretest and delayed posttest. In other words, both groups improved from pretest to posttest and the learners in both groups had maintained some knowledge of their newly learnt words/phrases two weeks after treatment.

**Table 3: Post hoc pairwise comparison for within-subjects contrast on overall test score**

Treatment				SE	Sig.
no-captions	pre-test	vs	post-test	1.175	.000*
	pre-test	vs	delayed	1.064	.000*
captions	pre-test	vs	post-test	1.175	.000*
	pre-test	vs	delayed	1.064	.000*



In order to answer research question 2 (which of the three targeted linguistic features was most effectively learnt by the use of video plus captions) the test was divided into three sections and a series of 2 x 3 mixed factorial design split plot ANOVA were performed with type of treatment (between-subjects) and type of question (within-subjects) as independent variables and test scores as the dependent variable.

There was a significant interaction between time and treatment, a significant main effect for time and a significant main effect for treatment for one section of the test only, that testing for the learning of phrases. This tells us that the *captions group* performed significantly better than the *no captions group* in this section. Results of the split plot ANOVA are given in Table 4 and descriptive statistics for this section of the test are given in Table 5 (there were 11 items in this part of the test)

**Table 4: Results of split spot ANOVA for the scores of the phrase section**

	df	F	Sig.	Partial Eta Squared
time	2	23.76	.000*	.569
treatment	1	5.52	.030*	.235
time * treatment	2	4.74	.026*	.208

**Table 5: Descriptive statistics of scores of the phrase section**

		no-captions			captions	
		n = 10			n = 10	
	Mean	%	SD	Mean	%	SD
pre-test	5.40	49.09	1.43	5.10	46.36	1.10
post-test	6.50	59.09	1.43	8.10	73.64	1.37
delayed	6.70	60.91	1.34	8.40	76.36	1.35

These results demonstrate that the *captions group* outperformed the *no-captions group* on the phrase section, suggesting that the use of video plus captions enhanced learners' learning of phrases. Differences between the two groups were not statistically significant for any other section of the test.

All participants, in both groups, reported in the interview, following the treatments, that they thought that they had learnt something. Words like plucky, tacky and phrases such as suit yourself and you are a natural were the target language most commonly mentioned by participants. Nine of the ten participants in the *no-captions group* said that they would have preferred to watch the video with captions, one participant liked it the way it was. This participant commented:

as for this movie, I don't think I would have needed English captions as I understood the meanings of most sentences and I don't know if I could handle three inputs simultaneously.

All twenty participants stated that they enjoyed/preferred learning English via watching video (with or without captions) and fourteen mentioned that it increased their interest in and motivation for learning. Five participants pointed out that using

videos in the classroom would complement the existing language teaching curriculum but that it would not be successful if students just simply watched a video once or did not do any exercises to help them learn.

All participants in the *no-captions group* reported paying attention to both sounds and pictures. Since captions were not available, they stated that the only way to catch the target words and phrases was to listen to the dialogue very carefully. They also stated that visual images such as body movements, gestures, facial expressions and surroundings were important and helpful for stimulating imagination and guessing the meanings of unknown words/phrases.

Table 6 shows how each participant in the *captions group* allocated their attention while watching the video. The numbers indicate the ranking given to each information source according to how participants reported prioritizing their attention (1 being their first priority). In those instances where the number one is repeated, it means that the participant gave equal attention to two information sources. Results show that eight out of ten participants reported that they paid most of their attention to the captions. Reasons given were: a/ reading is easier than listening, b/ reading is more helpful for learning, c/ reading captions is a habit and d/ captions are eye-catching. Of these eight participants, three said that listening to speech was their lowest priority since they felt that only captions and pictures could help them learn as many target words as possible. Three participants said that they could not attend to three sources of information at the one time, thus they decided to prioritise the most helpful. Participant 1 commented:

I paid most of my attention to reading captions, pictures came second, and the sound came last, because it's impossible to do three things at once. So I decided to prioritise reading captions, mainly because reading is easier than listening and in terms of helpfulness, captions were ranked the first.

The two participants who did not prioritise captions stated that they looked only at the captions if they did not understand the dialogue.

First I focused on the sound and pictures, and if I heard some words that I couldn't understand, I looked at the captions to search for these words. So yeah, I read captions only if I needed to.

## Discussion

Research question one asked whether video plus captions leads to better learning of unknown words/phrases than video alone. Since no statistically significant effect was found between the two treatment groups, we cannot conclude that the *captions group*

<sup>1</sup> The *no captions group's* 95% confidence interval lower bound = 22.77 and upper bound = 29.63, whereas the *captions group's* lower bound = 28.37 and upper bound = 35.23 (based on the means of the posttest).

was clearly superior to its counterpart. However, given the fact that the *captions group* started with slightly lower mean scores on pretests and that there were slightly overlapping confidence intervals for the means of the two groups<sup>1</sup>, we can conclude that the results may have been different if the two groups had been equal on the pretest. In addition, the fact that the interaction effect turned out to be statistically significant (i.e., in comparison with the *no captions group*, the *captions group* had higher gains from pretest to posttest) means that video plus captions led to better learning of unknown words/phrases than video alone. This result shows that captioned video material has the potential to impact on micro-level learning (i.e., intentional learning of new linguistic knowledge) as well as on macro-level learning (i.e., video content comprehension). Contrary to the concern that presenting duplicate information (e.g. spoken words plus verbatim written words) may overwhelm learners' attentional capacity and impede learning (Sweller, 1988), the results clearly indicate that learners learnt more words from captioned video than learners in the *no captions group*.

**Table 6: Summary of each participant's attention allocation (the *captions group*) ranking from 1 to 3 (1 being most attention)**

participants	sound	pictures	captions
1	3	2	1
2	3	2	1
3	2	1	1
4	2	1	1
5	2	1	1
6	3	2	1
7	1	2	1
8	1	1	2
9	3	2	1
10	1	2	3

Research question two asked which of the three linguistic features targeted (words, phrasal verbs and phrases) were most effectively learnt through video plus captions. The answer to this question is phrases; the *captions group* achieved a statistically higher score on this part of the test than did the *no captions group*. Participants in the *captions group* learnt the meanings of the phrases: suit yourself, for once, you are a natural and don't give me that. One can surmise that participants were successful in learning these phrases because they consisted of words that were familiar to them and because the surrounding contextual and/or cotextual cues were clear. This would provide evidence of Batstone's 'given to new' principle (2002a). In contrast, they did not guess the underlying meaning of someone is on the prowl, maybe because they did not understand the word 'prowl' and because the context did not assist in guessing. Significant gains for the *captions group* on the sections of the test assessing learning of phrases, suggests

that the use of video plus captions can help students learn colloquial language and how and when native speakers use it.

Research question 3 asked how learners felt about learning English through video and captions. All participants who were exposed to captions reacted very positively and believed that the captions helped them to learn more words. All 20 participants agreed that watching video with captions was a good way to learn new L2 material. Some participants commented that they would continue to make use of this method in learning English (either through self study or suggesting that their teachers use it in the classroom).

Research question 4 asked what information students paid most attention to in watching video with captions, that is, captions, sound or pictures. There was a definite tendency for participants in the *captions group* to prioritise reading captions when captions were present. This confirms the theory that captions are processed mandatorily (D'Ydewalle et al., 1991; Van Lommel et al., 2006). The result was that, for some participants, captions competed with sound. They reported times when they were unable to read and listen at the same time. Captions, also, at times competed with pictures. Whilst most participants reported being able to switch flexibly from captions to visual images and back again, this was not always the case. There were cases when participants reported that they had not had enough time to look at pictures after they had finished reading the captions and trying to work out what they meant. These results perhaps endorse the conclusion made by Rubin (1995), that teachers may be best not to use captions if they want to use video to contribute to L2 learners' listening skills. Viewing captioned video may lead learners to prioritise reading over listening and prevent them from making full use of auditory and other visual clues. In the light of these results, teachers should maybe change the combination of input (captioned or uncaptioned) when using video, according to their objectives and learners' language proficiency.

A small study of this nature has a number of limitations. The small sample size means that results need to be taken with caution. In addition, the interview could have gone further in trying to determine exactly how participants used captions, pictures and sound in order to ascertain the meaning of unknown language. This would be an obvious area for future research, using perhaps stimulated recall as a data elicitation strategy.

## Conclusion

This study demonstrates that captioned video material does not only enhance comprehension of content but also helps learners learn new L2 material. Results show that, of the three linguistic features tested, it was the learning of phrases that was most enhanced by the use of video plus captions.

Exposing L2 learners to video plus captions, therefore, seems a pedagogically healthy approach that may be incorporated into a classroom setting. Results from this study

suggest that learning may be enhanced if students have some relevant background knowledge (participants in the study watched the video first in their L1), if learners are introduced to new words before viewing as a prompt for noticing and if the context and cotext surrounding the language are clear enough to enable guessing. However, more research is warranted in order to determine which of these factors contribute most to effective learning and to what extent.

Participants were positive about the opportunity to learn through the use of video with captions. The majority of them prioritized reading captions over attending to auditory information and a number of them felt that they were unable to attend to three information sources at the same time. In this respect, in a teaching situation, it may be wise to vary the use of video with captions and without captions so that learners have the opportunity to attend to auditory information as well.

## References

- Arcario, P. (1992). Criteria for selecting video materials. In S. Stempleski & P. Arcario (Eds.), *Video in second language teaching: Using, selecting, and producing video for the classroom*. U.S.: Teachers of English to Speakers of Other Languages.
- Baddeley, A. D. (1992). Working memory. *Science*, 255, 556-559.
- Baltova, I. (1999). Multisensory language teaching in a multidimensional curriculum: The use of authentic bimodal video in core French. *The Canadian Modern Language Review*, 56(1), 31-48.
- Batstone, R. (2002a). Making sense of new language: A discourse perspective. *Language Awareness*, 11(1), 14-29.
- Batstone, R. (2002b). Contexts of engagement: A discourse perspective on 'intake' and 'pushed output'. *System*, 30, 1-14.
- Broadbent, D. E. (1958). *Perception and communication*. New York: Pergamon Press.
- Borras, I., & Lafayette, R. C. (1994) Effects of multimedia courseware subtitling on the speaking performance of college students of French. *The Modern Language Journal*, 78, 61-75.
- Chun, D. M., & Plass, J. L. (1997). Research on text comprehension in multimedia environments. *Language Learning and Technology*, 1(1), 60-81.
- Chung, J. M. (1996). The effects of using advance organizers and captions to introduce video in the foreign language classroom. *TESL Canada Journal*, 14(1), 61-65.
- Danan, M. (1992). Reversed subtitling and dual coding theory: New directions for foreign language instruction. *Language Learning*, 42(4), 497-527.
- D'Ydewalle, G., & Gielen, I. (1992). Attention allocation with overlapping sound, image and text. In K. Rayner (Ed.), *Eye movements and visual cognition: Space perception and reading*. New York: Springer-Verlag.
- D'Ydewalle, G., Praet, C., Verfaillie, K., & Van Rensbergen, J. (1991). Watching subtitles television: Automatic reading behaviour. *Communication Research*, 18(5), 650-666.
- D'Ydewalle, G., & Van Rensbergen, J. (1989). Developmental studies of text-picture interactions in the perception of animated cartoons with text. In H. Mandl & J. R. Levin (Eds.), *Knowledge acquisition from text and pictures*. Elsevier, Amsterdam, Holland.
- Ellis, R. (1997). *SLA research and language teaching*. Oxford: Oxford University Press.
- Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.
- Garza, T. J. (1991). Evaluating the use of captioned video materials in advanced foreign language learning. *Foreign Language Annals*, 21(2), 225-241.
- Gass, S. (1988). Integrating research areas: A framework for second language studies. *Applied Linguistics*, 9, 198-217.

- Goldman, M., & Goldman, S. (1988). Reading with closed-captioned TV. *Journal of Reading*, 31(5), 458-461.
- Goldman, M. (1996). If you can read this, thank TV. *TESOL Journal*, 6(2), 15-18.
- Gruba, P. (2005). *Developing media literacy in the second language classroom*. Sydney: Macquarie University.
- Guillory, H. G. (1998). The effects of keyword captions to authentic French video on learner comprehension. *CALICO*, 15(1-3), 89-108.
- Hulstijn, J. H. (1992). Retention of inferred and given word meanings: Experiments in incidental vocabulary learning. In P. Arnaud & H. Bejoint (Eds.), *Vocabulary and applied linguistics*. London: Macmillan.
- Lightbown, P. M., Halter, R., White, J. L., & Horst, M. (2002). Comprehension-based learning: The limits of 'do it yourself'. *Canadian Modern Language Review*, 58(3), 427-464.
- Markham, P. (1999). Captioned videotapes and second-language listening word recognition. *Foreign Language Annals*, 32(3), 321-328.
- Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Rowley, MA: Newbury House.
- Neuman, S. B., & Koskinen, P. (1992). Captioned television as comprehensible input: Effects of incidental word learning from context for language minority students. *Reading Research Quarterly*, 27(1), 95-106.
- Paivio, A. (1986). *Mental representation: A dual-coding theory approach*. New York, New York: Oxford University Press.
- Ruhe, V. (1996). Graphics and listening comprehension. *TESL Canada Journal*, 14(1), 45-60.
- Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11, 129-158.
- Secules, T., Herron, C., & Tomasello, M. (1992). The effect of video context on foreign language learning. *The Modern Language Journal*, 76, 480-490.
- Simatos, A., & Spencer, K. (1992). *Children and media: Learning from television*. Manutius Press: Liverpool.
- Spache, G. (1981). *Diagnosing and correcting reading disabilities*. Boston, MA: Allyn & Bacon.
- Stempleski, S. (1992). Teaching communication skills with authentic video. In S. Stempleski & P. Arcario (Eds.), *Video in second language teaching: Using, selecting, and producing video for the classroom*. U.S.: Teachers of English to Speakers of Other Languages.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12, 257-285.
- Travers, R. M. W. (1967). *Research and theory related to audiovisual information transmission (Rev. ed.)*. Kalamazoo, Mich.: Western Michigan University.
- Vanderplank, R. (1988). The value of teletext sub-titles in language learning. *ELT Journal*, 42, 272-281.
- Vanderplank, R. (1990). Paying attention to the words: Practical and theoretical problems in watching television programmes with uni-lingual (CEEFAQ) subtitles. *System*, 18(2), 221-134.
- Van Lommel, S., Laenen, A., & d'Ydewalle, G. (2006). Foreign-grammar acquisition while watching subtitled television programmes. *British Journal of Educational Psychology*, 76, 243-258.
- Williams, J., & Snipper, G.C. (1990). *Literacy and bilingualism*. New York: Longman.
- Ying, Y. S. (2001). Acquiring vocabulary through a context-based approach. *English Teaching Forum*, 39(1), 18-21.

## Appendix 1 : The word list

### Segment One

Suit yourself	castle	hide ..
Fog	tore (past tense of 'tear)	prey ...
Moustache	escort	you are a natural
Draw		

### Segment Two

float	balcony	wizard	don't give me that
on the prowl	for once	tacky	stand(ing)
plucky	spell	regards	

## Appendix 2: The script of the selected scene

(target words/phrases are bolded)

### Segment One

[in the hat shop]

<woman> Sophie, we just closed the shop. You've done enough work.  
Why don't you come out with us this time?  
<Sophie> No, I better finish this. You go and have fun.  
<woman> All right, **suit yourself**.  
Let's go, girls.  
<girl A> Wait for me!  
<girl B> I'm ready.  
<girl C> Do I look OK?  
<girl D> Oh! It's Howl's **castle**!  
<girl B> Where?  
<girl A> Howl?  
<girl A> Do you think Howl will go into town?  
<girl A> He's gone.  
<girl B> No, he's just **hiding** in the **fog from** those planes.  
<girl C> Did you hear what happened to Martha? They say Howl **tore**  
her heart **out**.  
<girl A> I'm scared to go out!  
<girl B> Don't worry, he only **preys on** pretty girls.  
[all laughing]  
<woman> All right, let's go.  
<girl A> Your girls are mean.  
  
<soldier A> Hey! Looks like a little mouse lost its way.  
<Sophie> Oh, no, I'm not lost.  
<soldier A> This little mouse looks thirsty. We should take her for a cup of tea.

<Sophie> No thanks, my sister's expecting me.  
<soldier B> She's pretty cute for a mouse.  
<soldier A> How old are you anyway? You live around here?  
<Sophie> Leave me alone.  
<soldier A to  
soldier B> You see? Your **moustache** scares all the girls.  
<soldier B> So? I think she's even cuter when she's scared.  
<man to Sophie> There you are, sweetheart. Sorry I'm late. I was looking  
everywhere for you.  
<soldier A> Hey! Hey! We're busy here.  
<man> Are you really? It looked to me like the two of you were just  
leaving.  
[the man made the two soldiers go away involuntarily]  
<man to Sophie> Don't hold it against them. They're actually not all that bad.  
<man to Sophie> Where to? I'll be your escort this evening.  
<Sophie> Oh, I'm, um, just going to the bakery.  
<man> Don't get alarmed, but I'm being followed.  
Act normal.  
<man> Sorry. Looks like you're involved.  
This way.  
Hold on.  
[they fled to the air] [in the air]  
<man> Now, straighten your legs and start walking.  
<man> See? Not so hard, is it?  
<man> **You are natural.**  
[landed]  
<man> I'll make sure to **draw** them **off**. But wait a bit before you head  
back outside.  
<Sophie> OK.  
<man> That's my girl.

## Segment Two

<Littie> Sophie!  
<Sophie> Lettie.  
<Lettie> What's going on ?  
Someone just told me you **floated down** onto our **balcony**.  
<Sophie> So that did happen. It wasn't a dream.  
<man2> Lettie, would you like to use my office?  
<Lettie> I should really get back to work. Thank you though.  
<Lettie> Wow. He must have been a **wizard**, then.  
<Sophie> But he was so kind to me. He rescued me, Lettie.  
<Lettie> Of course he did. He was trying to steal you heart! You are so  
lucky, Sophie. If that wizard were Howl, he would have eaten  
it.  
<Sophie> No, he wouldn't. Howl only does that to beautiful girls.



<Lettie> **Don't give me that.** You need to be more careful. It's dangerous out there. Even the Witch of the Waste is back **on the prowl.**

<Lettie> Are you listening?

<Sophie> Hun?

<Lettie> Ugh

<cake man> Lettie, the chocolate éclairs are done.

<Lettie> OK! I'll be right there.

<cake man> Thanks.

<Sophie> All right, I better get going then. I just wanted to make sure that you were doing ok.

[outside bakery]

<Lettie> Now, Sophie, do you really want to spend the rest of your life in that hat shop?

<Sophie> The shop was just so important to Father. And I'm the eldest, I don't mind.

<Lettie> I'm not asking what Father would have wanted. I want to know what you want.

<Sophie> Well  
I better be going.

<Lettie> It's your life, Sophie. Do something for yourself **for once**, will you?

<Sophie> Bye, Lettie.

[Sophie's back in the hat shop.]

<Sophie> I'm sorry, but the shop's closed now, ma'am.  
I could have sworn I locked that door. Must have forgotten.

<Witch> What a **tacky** shop!  
I've never seen such tacky little hats.  
Yet you're by far the tackiest thing here.

<Sophie> I'm afraid you will have to leave now.  
The door is over here, ma'am. We're closed.

<Witch> **Standing up** to the Witch of the Waste, that's **plucky**.

<Sophie> The Witch of the Waste?

<Witch> The best part of that **spell** is you can't tell anyone about it.  
My **regards** to Howl.

### Appendix 3: Sample items from the test (pre-test & delayed post-test)

A. Next to each word, write the letter of its meaning.

e.g. teacher	_____g_____	a. exciting event
1. balcony	_____	b. a large entrance or reception room
2. wizard	_____	c. a man who has magical power
3. plucky	_____	d. a weather condition in which water falls from the sky in small drops
4. escort	_____	e. a weather condition which causes difficulty to see an object clearly
5. fog	_____	f. a message that says you hope someone is well
etc.	_____	g. someone whose job is to teach in a school
		h. actors in a film
		etc.

Turn Over (please do not turn back)

B. Find one word from the list which fits **best** in each sentence and write it in the blank. Each word can be used **only once**.

sneak	escort	fog	regards	well
etc.				

1. I wonder how your parents have been doing lately. Please send my \_\_\_\_\_ to them.

2. Where are you heading? Let me be your \_\_\_\_\_ tonight.

etc.

C. Decide which preposition(s) collocate(s) **best** with the verb.

1. hide \_\_\_\_\_

(a) over (b) on (c) from (d) around

2. prey \_\_\_\_\_

(a) for (b) at (c) on (d) with

etc.

D. Find **one verb** and **one or two prepositions** from the lists which fit into each sentence. Each word can be used **only once**.

VERBS					PREPOSITIONS				
defend	make	give	hide	leave	at	on	up	for	off
check	draw	float	stand	prey	out	into	to	with	down
etc.									

1. Jack was so brave that he wasn't afraid to \_\_\_\_\_ people who bullied him. (find 1 verb and 2 prepositions)

etc.

E. The following questions are aimed at testing your knowledge of English phrases. Answer each question and write your answer on the line provided. You can write the answer in your **native language** or in **English**.

Example:

\* When someone says "**it's raining cats and dogs**", what do you think it means?

Correct Answer: It's raining very heavily. OR 正下着倾盆大雨。

Incorrect Answer: "Cats and dogs are falling like rain". OR 雨下得像猫狗。

1. When somebody says "**suit yourself**", what do you think it means?

etc.

F. Decide which is the best response to each sentence.

1. A: I don't think I can come to the party tonight.

B: \_\_\_\_\_.

- (a) You are a natural
- (b) For once I understand
- (c) Suit yourself
- (d) I wonder too

etc.

G. Read the statement and write T (true) or F (false) on the line.

1. \_\_\_\_\_ "Do something for yourself for once" means you always think of yourself.

etc.

## **Appendix 4: Interview Guidelines**

1. Before today, have you ever thought of using video (plus captions) to improve/learn English?
2. How do you feel about the sequence in which you watched the video (Chinese first then English)? Do you feel you've learned something from this research? If so, what have you learned?
3. (For the no captions group) When you were watching the video clip, apart from paying attention to the target words/phrases, what did you pay most of your attention to? Sounds or pictures? Why? Did you have difficulty understanding the English dialogue? Would you have liked to have watched the video with captions? Why?
4. (For the captioned group) When you were watching the video, apart from paying attention to the target words/phrases, what did you pay most of your attention to? The sound, pictures or captions? Why? Did you have difficulty understanding the English dialogue? Do you think captions helped or hindered you? Why?

# **TEACHERS' VIEWS AND PRACTICE: THE PLACE OF THE MOTHER TONGUE IN THE IMPLEMENTATION OF THE KENYAN LANGUAGE-IN-EDUCATION POLICY**

**Jennifer M. Jones**

*University of Auckland, New Zealand*

## **Abstract**

*This paper reports on an ethnographic study which investigated the varied responses of teachers to both the Kenyan language-in-education policy (Mother Tongue as subject and as medium of instruction) and the process by which it is being implemented in Sabaot, one of the fifty indigenous languages in Kenya. The Sabaot people who live on Mt Elgon are in a similar predicament to minority groups elsewhere as their language suffers the effect of colonisation and faces the pressure of globalisation. In an effort to “reverse language shift” (Fishman, 1991) and maintain the language (Pauwels, 2004), while at the same time raise education standards (Cummins, 2000), the policy was recently implemented in a number of Sabaot schools. The study was conducted in one of these schools during a time of significant unrest on Mt Elgon. Data was collected over a seven month period and consisted of analysing documents, observing and interviewing in one school. The findings show that the policy is only partially implemented as the mother tongue is only used to a small extent before students are transitioned to Kiswahili, the national language, and English, the official language. Constraints affecting policy implementation are examined as are the underlying reasons teachers are motivated to implement the policy. Implications for the Kenyan context, which may also be relevant to other multilingual contexts, are discussed.*

## **Introduction**

This paper, reporting on a study in progress, seeks to investigate the place of the mother tongue (MT) in the implementation of the Kenyan language-in-education policy. While language-in-education in Kenya is the focus of this paper, the place of the MT in education is a strongly contested issue in multilingual contexts today throughout the world. Much has been written, for example, about New Zealand Māori (Fishman, 2006; Peddie, 2003) and New Zealand immigrant languages (Starks, 2005). Not only is it argued that the language used as the medium of instruction influences the educational attainment of students (Hovens, 2002), the learning environment (Alidou, 2004), students' identity (Corson, 1999; Norton, 1997), the participation of girls and parental involvement in school (Benson, 2004), it is also argued that it may influence the very survival of indigenous languages (Nettle & Romaine, 2000). It is estimated that in the next 100 years, 90% of minority languages will die (Wolff, 2005) because of the worldwide process of globalisation (Fishman,

2001). Kenya, the focus of my paper, has 50 indigenous languages (Gordon, 2005), 12 of which are endangered (Batibo, 2005).

Kenya's language-in-education policy is supportive of the use of indigenous languages for instructional purposes. Like many other Anglophone countries, such as South Africa, Fiji and Uganda, Kenya's language-in-education policy supports MT instruction during the early years of education (Ministry of Education, 1997; Ministry of Education, Science and Technology, 2002). Kenya's policy is represented by an early-exit transitional model of bilingual education (Heugh, 2005): MT is taught as a subject during daily half hour lessons and used as the medium of instruction until the fourth year of school. At this time, an abrupt transition occurs as MT is replaced by English. English is also taught as a subject, along with Kiswahili, throughout primary and secondary school. Although the language-in-education policy is favourable towards MT, there seems to have been little official effort to implement it. While textbooks for the subject MT have been produced in 22 of the 50 indigenous languages (Mbaabu, 1996), all textbooks for content subjects are in English. In addition, the syllabus for teacher training shows that little emphasis is given to MT in pre-service training courses for teachers (Ministry of Education, Science and Technology, 2004).

In order to implement the language-in-education policy, language-in-education planning has taken place in a number of Kenyan indigenous language groups. This kind of planning not only involves setting goals and objectives for the language education system, it also addresses practical concerns such as teacher supply and materials (Ingram, 1989). One of the languages targeted by recent language-in-education planning efforts is Sabaot. Sabaot is spoken by approximately 300,000 people living on Mt Elgon on both sides of the Kenya-Uganda border, about half of whom live in Kenya (Taylor, 2002). The Sabaot have been marginalised for many years, having only gained Kenyan parliamentary representation in 1993. Until quite recently, Sabaot children were punished if they spoke their language at school. They were forced instead to speak Kiswahili or the neighbouring Kibukusu language, both unrelated to Sabaot.

Several language planning activities have been carried out as part of the process of policy implementation. These include status planning (van Els, 2005), corpus planning (Liddicoat, 2005), usage planning (Lo Bianco, 2004) and prestige and image planning (Ager, 2005). As a result of these activities, the Sabaot writing system was developed, as were teaching/learning materials for the subject MT. Moreover, approximately 300 teachers from quite a number of the 100 primary schools on Mt Elgon which are dominated by Sabaot, have participated in some form of in-service training. During this training, teachers are introduced to the Sabaot orthography, the teaching/learning materials and the rationale of using the MT for instructional purposes.

Six of the teachers who have attended training come from Sumanet School (a pseudonym, as are other names in this paper), which draws its students from the surrounding area, from about a two kilometre radius. In April 2007 the school had 16 classrooms, 26 teachers and 966 students. Classes, although large, were taught different subjects by different teachers. Approximately 95% of students in the school are Sabaot, and the remainder are mostly Bukusu or Teso. Most come from homes in the rural area where Sabaot and/or Kiswahili are spoken. Some also come from the nearby market where Kiswahili is the lingua franca. Another group who have come more recently to the school are from Kopsiro Division, which is 20 kilometres away, where the predominant language is Sabaot. This group began settling in Sumanet Area, and other parts of Mt Elgon, from October 2006 when conflict broke out between two Sabaot clans in their home area due to land disputes. By March 2007 approximately 34,200 people had been displaced because of the unrest (*Kenya: Mt Elgon preliminary appeal*, 2007) following the deaths of 89 Sabaot people and closures of 30 primary schools (Wabala & Macharia, 2007).

Before the recent conflict on Mt Elgon, I was involved in the implementation of the policy for seven years in the Sabaot language group, and that is how I, a New Zealand researcher, became interested in undertaking this study. During my time in Kenya I observed that some teachers responded very positively to the implementation of the policy, while others were less enthusiastic. I wanted to understand what lay beneath the teachers' responses. Although a number of studies have investigated the responses of teachers and other stakeholders to the implementation of the language-in-education policy in Kenya (Abagi & Cleghorn, 1990; Bunyi, 1996; Cleghorn, Merritt, & Abagi, 1989; Muthwii, 2002), these have primarily been carried out in major language groups such as Kikuyu and Luo where some provision has been made for MT to be taught, prior to and since independence. Little attention has focused, however, on more recent implementation efforts in small minority language groups apart from Pokomo (Graham, 2007) and Tharaka (Nyaga, 2005). The purpose of the study is to investigate the different responses of teachers from one minority group, the Sabaot, to the language-in-education policy and the process by which it is being implemented. The research questions addressed in this study are:

1. How are Sabaot teachers implementing the language-in-education policy (MT as subject and medium of instruction) in their school?
2. What are the Sabaot teachers' perceptions of the language-in-education policy (MT as subject and medium of instruction) and of the process by which it is being implemented?
3. What are the major forces affecting the language-in-education policy implementation process (e.g. historical, global and local influences and stakeholder attitudes to language use in education)? How do they affect the process?

## Methodology

In an effort to address these questions, an ethnographic approach was employed to guide the inquiry. An ethnographic approach places importance on gaining a holistic view (Wolcott, 2002) and an understanding of the situation from perspective of the participants (Watson-Gegeo, 1988), the insiders. An insider's perspective was needed to gain an understanding of the response of Sabaot teachers to the implementation of the policy. As ethnography is concerned with providing an in-depth understanding of the situation under investigation, the ethnography for this inquiry took place within one school.

Having obtained ethics approval from the University of Auckland Human Participant Ethics Committee and a research permit from the Office of the President, Republic of Kenya, I gained consent from participants at Sumanet School. This school was chosen as the research site because the majority of residents are Sabaot and it is located in an area in which I have long-standing connections and was still peaceful when I began data collection in March 2007. The primary participants were five female Sabaot teachers who had attended at least one five day in-service MT course at some time since 2000. One taught the Early Childhood Development class, three taught Standard One, and the other taught Standard Two. Table 1 below provides more information about the teachers and learners involved in the study.

**Table 1      Responsibilities of teachers**

<b>Names and ages of teachers</b>	<b>Class Level</b>	<b>Equivalent class level in New Zealand</b>	<b>Average age of learners</b>	<b>Class size: March 2007</b>	<b>Ethnicity of learners</b>
Mary (35)	Early Childhood Development	Pre-school Kindergarten	4-6	90	85 Sabaot 4 Bukusu 1 Teso
Anne (42) Nelly (52) Helen (44)	Standard 1	Year 1	6-8	125	118 Sabaot 2 Bukusu 4 Teso 1 Luo
Susan (32)	Standard 2	Year 2	8-9	95	91 Sabaot 1 Bukusu 3 Teso

Secondary participants were 42 stakeholders from the school and community who were concerned with the implementation of the policy at Sumanet School: parents, education officers, the head teacher, other teachers as well as some from the non-governmental organisation assisting with policy implementation. This article is based on data from the primary participants.



Four main methods were used to gather data: observation, semi-structured interviews, focus group interviews and document analysis. Field notes were also taken to record my ongoing observations and reflections. Two types of observation were used: classroom observation and informal observation. During classroom observations I observed a total of 62 lessons across a range of subjects taught by primary participants. In addition, I observed a further 20 lessons taught by secondary participants. The focus of these observations was on the teacher. Initially these observations were of very broad scope but over time as I formed tentative hypotheses of what lay beneath the teachers' practice, my focus narrowed. These lessons were recorded for later analysis and 37 were transcribed. In order to gain a holistic view of the sociolinguistic situation, I also carried out informal observation of language use in the school and community.

Following observation of teachers in the classroom, I sought to gain an understanding of their perspective of their own behaviour through two or three semi-structured interviews with each teacher of approximately one hour duration. To help me gain a holistic view of the perceptions of members of the school community concerning the implementation of the language-in-education policy, an additional 31 semi-structured or focus group interviews were held with secondary participants. These interviews were recorded for later analysis. Documents pertaining to the study were also analysed. These included policy documents, school records and other materials that related to the implementation of the language-in-education policy such as lesson plans and schemes of work. Throughout the study I worked alongside a Sabaot person, someone literate in Sabaot, English and Kiswahili who acted as an interpreter with non-English speakers and assisted with transcription.

Data collection was significantly affected by the situation of unrest on Mt Elgon. Not only was there an influx of students and teachers in Sumanet School as I collected data from March to September 2007, in early April the conflict spread to Sumanet area. Six people were massacred one night in a nearby town (Kwalia & Ng'etich, 2007). The conflict continued to escalate and by September, approximately 200 people had been killed (Siringi, 2007) including relatives of some of the participants in my study. The ethnographic approach I employed allowed me the flexibility to carry out research in this somewhat volatile context. Since it was impossible to anticipate what would happen next, in consultation with participants, I decided on a daily basis which class I would observe and/or who I would interview. I also adopted a number of precautions to reduce the risk to myself and others in this situation. For example, I developed a 'plan of escape' in the event of an attack (Sluka, 1995, p. 282) and each time I returned to the area, I checked beforehand that it was safe to do so.

Data analysis occurred simultaneously with data collection and continued after I left the research site. The sequence of analysis employed in the study broadly followed Ellis and Barkhuizen (2005): *"coding for themes - looking for patterns - making interpretations - building theory"* (italics in original, p. 259). This theory, grounded in the data, is defined by Glaser and Strauss (1967) as "grounded theory", that is,

“theory that is derived from data and then illustrated by characteristic examples of data” (p.5). The following section provides a report of my findings.

## Findings

In this section, I report on and discuss my findings with respect to the implementation of the language-in-education policy. Firstly, I provide a summary of my findings in terms of MT as subject and MT as medium of instruction. After this, I focus more specifically on two areas affecting policy implementation: the constraints teachers face and the underlying reasons for their motivation to implement the policy.

### MT as subject

As mentioned above, the policy stipulates that MT should be taught as a subject on a daily basis in Standards One, Two and Three. In Sumaneet School, Sabaot was only taught as a subject in two classes, that is, in Standards One and Two. It seems that MT was not taught as a subject in Standard Three because teachers at this level had not attended in-service training and therefore did not have the confidence to teach Sabaot.

Lesson observation data revealed that in Standards One and Two MT lessons, teachers focussed on developing students’ Sabaot language skills: listening, speaking, reading and writing. The teachers’ main focus seemed to be to develop students’ reading skills and this was facilitated by utilising the Sabaot teaching/learning materials (e.g., Boiyo et al., 2004). Teachers also sought to develop speaking fluency through singing traditional songs and reciting prayers. Writing skills were taught using the same approach that was used in other subjects, that is, by instructing students to copy new lesson content off the blackboard.

The teachers made a deliberate effort to promote the Sabaot language during MT lessons. All those who taught MT as a subject avoided mixing languages and endeavoured to use “pure” Sabaot. Despite their efforts, they unconsciously used a number of Kiswahili words such as “kweli” (really) and “sasa” (now). Another strategy some teachers used to promote the Sabaot language was to refuse to accept students’ responses unless they spoke Sabaot. For example, Helen, a Standard One teacher responded in the following way to a student who used Kiswahili during a MT lesson:

*Āmāchē chiito nyoo ng’āloolē kuutiitaab moomo. Māāmāchē Kiswahili.*  
*I want a person who speaks the mother tongue. I don’t want Kiswahili.*<sup>1</sup>

### MT as medium of instruction

The other strand of the language-in-education policy examined in the study is the place of the MT, Sabaot, as the medium of instruction. The study revealed that although the policy stipulates that the MT is used as the medium of instruction, there seems to be an unwritten policy operating in Sumaneet school, that is, to transition

<sup>1</sup>

Different fonts represent different languages: Italics are used for Sabaot, bold for Kiswahili and normal for English.

pupils as quickly and as early as possible, firstly to Kiswahili and then to English. Consequently, Sabaot is only used partially as the medium of instruction in the first two class levels (Early Childhood Development and Standard One). Kiswahili is the primary language of instruction in these levels and English is also used, to a small extent, as teachers constantly seek to integrate English vocabulary into content lessons. As revealed in the following lesson extract, Sabaot is mostly used in combination with Kiswahili and/or English to clarify the meaning of new English vocabulary:

Nelly: **Nimesema** eye *ku koong'ta*, nose *ku sēruut*, tongue *ku kiy nee?*

**I have said** eye *is an eye*, nose *is a nose*, tongue *what is a tongue?*

Students: *Mēēsiit*

*A tongue*

From Standard Two onwards, any use of the MT for instructional purposes in content subjects is avoided and Kiswahili is used as the medium of instruction with increasing use of English. As Anne explained:

I will try to avoid using MT in Class 2 and 3. I'll not use most of the time MT, but I will use Kiswahili and English.

By Standard Three, teachers use both Kiswahili and English approximately to the same degree. No use of MT is made.

### **Language choice for literacy**

Another part of policy implementation (MT as medium of instruction) involves the language used for literacy by both teachers and students. While teachers may use Sabaot, Kiswahili and/or English for orally instructing students, English is the only language used for literacy from Standard One onwards. In the Early Childhood Development class, the use of language for literacy purposes is slightly different, however. At this level, the teacher seems to alternate between using Sabaot, Kiswahili and English for literacy. Teachers seem to only give MT a limited place as they implement the policy because although they might recognise the benefits of using the MT for instructional purposes, they encounter a significant number of constraints. I will discuss these in the following section.

### **Constraints affecting the implementation of the policy**

All the teachers reported that they experienced difficulty implementing the policy due to a number of constraints. These include: the language of textbooks and examinations, parental pressure, multilingual classes, dialectical differences, inadequate in-service training and the complexity of the Sabaot orthography. I will look at the impact of each of these in turn on the process of implementation. First, since the only books for content subjects at Sumaneet School are the nationally approved textbooks which are written in English, it is somewhat difficult for teachers to use only MT for instructional purposes. As Susan observed:

In fact all of the books are written in English. Eh. So you have no choice. Sometimes you teach what is inside the book.

Not only are the textbooks in English, all the commercially produced tests for content subjects that are used on a regular basis in the school are as well, as are the final national examinations set at the end of Standard Eight. As other researchers in Kenya also found (Abagi & Cleghorn, 1990; Bunyi, 1996), the practice of testing in English has a backwash effect in the classroom because teachers tend to focus on the English words and phrases they know students will be tested on. Teachers experienced another difficulty because exams are set in English. They reported that parents put considerable pressure on them to prepare their children to pass these examinations. Mary, the Early Childhood Development teacher, explained that some parents see any use of Sabaot as “a waste of time” and if they see too much Sabaot in their children’s books, they complain. She explained the problem in the following way:

But where I will get problem is from parents. ... Because maybe a father might come in and say, “Oh no. What I’m seeing in the book of my child or my daughter is, they are always writing Sabaot words. What about other words. What about English words?”

Another constraint affecting the process of implementation is that students do not all share a common language. Mary reported that parents of the few children from other tribal groups may accuse her of discrimination if she uses too much Sabaot:

Because maybe we have these parents who are non-Sabaot. Now if they will see that all the time their children are talking, you are teaching them in Sabaot, they will complain and say, “Oh! Now maybe these people are discriminating my child. Maybe they don’t want my child so I will better take my child and take him or her elsewhere.”

The majority of students in the school, however, are Sabaot. In addition, Sabaot is reported by teachers and parents to be the first language of the recent arrivals from Kopsiro (who constitute one-third of the school population). It is somewhat questionable, however, which language actually is the MT for many of the students originally from Sumaneet community. This is because it seems that language shift is taking place. All the teachers commented that increasing use is made of Kiswahili in Sabaot homes.

The shift to Kiswahili appears to be occurring for a number of reasons. Mary observed that some Sabaot parents want to give their children a “head-start” to their education by using Kiswahili at home. She noted that some parents are “ashamed of their language” and want their children to be “civilised” and speak Kiswahili. The parents may experience a sense of shame as a result of the marginalisation that the Sabaot have experienced since it was only recently that they gained parliamentary representation. Unquestionably, the occurrence of language shift puts teachers in a difficult position because whichever language they use as the medium of instruction there will always be some students who do not understand. Mary described their dilemma well:

It is two-way traffic. If I use the MT, those ones who do not understand will always keep quiet and those ones who understand will reply. ... It happens the

same to those ones who understand Kiswahili. They will also become surprised and wonder what you're talking. So it is two-way traffic.

Yet another constraint teachers face as they implement the language-in-education policy relates to their own proficiency in speaking the Sabaot language. Because teachers have also been affected by the process of language shift, some consider themselves more proficient speaking Kiswahili than Sabaot. Consequently, these teachers have more difficulty using Sabaot as the medium of instruction, than Kiswahili, and as a result, tend to use Kiswahili. The difficulty that some teachers experience in speaking Sabaot is compounded because they do not speak Koony, the standard Sabaot dialect, which is predominantly spoken in Sumaneet area. As a result they are concerned that students may laugh at their pronunciation. Anne, who speaks the Book dialect explains the problem:

I experience difficulties especially when I am pronouncing some words. Sometimes these children laugh at me when I pronounce some words in Book.

Not only do some teachers struggle in terms of their proficiency in speaking Sabaot, all teachers find reading and writing Sabaot more difficult than Kiswahili and/or English. It seems teachers experience this difficulty for several reasons. Firstly, unlike the extensive formal instruction they have had in English and Kiswahili, the exposure they have had to the Sabaot writing system is of very short duration: a few days or weeks during in-service training. Another reason that teachers experience difficulty reading and writing Sabaot relates to the complexity of the language which is reflected in its orthography. Language features such as advanced tongue root, vowel harmony and grammatical tone that do not exist in Kiswahili or English are marked in the Sabaot orthography (Larsen, Surai, Kigai, & Mang'esoy, 1990). One teacher, Susan, found the differences between Sabaot, Kiswahili and English so significant that she described Sabaot as a "new language". The final reason teachers may have difficulty reading and writing Sabaot, is that they use it in very few domains. Apart from at school when teaching the subject MT and for religious purposes, they have little opportunity to read or write Sabaot. Sabaot teachers are not alone in experiencing difficulty writing their language. Earlier research showed that Kikuyu and Luo teachers also felt less proficient writing their MT than writing Kiswahili and English (Abagi & Cleghorn, 1990).

The final constraint affecting policy implementation is that many teachers have a negative attitude towards attending MT in-service courses for the first time. The teachers' negative attitude stems from the above constraints as well as their perception that the government provides little incentive for them to teach MT or to attend such training. Susan described her attitude towards the course prior to attending it in the following way:

In fact, when you are told to go [to the course] for the first time, nobody's willing. ... It's like we're wasting time. ... After all, you know, I can teach English, I can teach Kiswahili, why do you want me to teach MT?

Although it appears that there are an overwhelming number of constraints affecting the implementation of the policy, some teachers are endeavouring to implement it to a certain degree. In the following section I will examine some of the reasons why some teachers from Sumanet School have been motivated to take a number of positive steps towards language-in-education policy implementation despite the constraints they face.

### **Reasons for implementing the language-in-education policy**

Many of the teachers appear to be motivated to give the MT a place in the implementation of the language-in-education policy (MT as subject and as medium of instruction), even if only partially, because of their positive attitude towards the use of MT for instructional purposes. The attitude of many teachers towards the place of MT in education changed during in-service training. Susan described her change of attitude by saying:

But when you go there [to the course] and discover it's [MT instruction] a very important exercise, you come when you have changed your mind.

As a consequence of the teachers' change in attitude towards the MT, there was also a change in their language behaviour. Instead of teaching Kiswahili (or another subject) during the subject MT, teachers began teaching Sabaot. They also began using MT to a limited degree as the medium of instruction because they perceived that there were benefits in using MT for instructional purposes. Teachers can see two main benefits of MT instruction: Firstly, they are aware of the pedagogical advantages in using the MT for instructional purposes. Secondly, they desire to see the Sabaot language and culture maintained, and as a result of that maintenance, they are assured that students are secure in their identity as Sabaot. I will address each perceived benefit in turn.

The primary reason that teachers seem to be motivated to implement the language-in-education policy appears to be because they perceive that there are pedagogical advantages of MT instruction. The teachers perceive that the MT is the language which acts as the link between home and school. In their view, MT is the foundation on which other learning can take place and the language that can be used to ensure all students understand the lesson content. In order to ensure pupils understand, teachers may, for example, use Sabaot to provide an explanation of the meaning of a new grammatical item as illustrated by an excerpt from a Standard One English lesson:

*A a, yoo akeenke kēēmwooyē, "This is." Sasa, kuuyu choong'ēch kēēmwooyē kule, "These are books."*

*No, if it is one we say, "This is." Now, because there are many we say, "These are books."*

MT is also perceived to facilitate language transfer to Kiswahili and English. As Anne explained:

After you have taught them in MT, they can be able to write other languages, especially English and Kiswahili. They are able.

Teachers are anxious that their students transition well to Kiswahili and English to increase their chance of good performance in national examinations.

The other main reason that teachers are motivated to implement the policy seems to be that they want to see the Sabaot language and culture maintained. Some teachers are concerned that the future of the Sabaot language and culture is threatened because it appears to be disintegrating. These teachers want to see the continuation and not the “collapse” of the Sabaot language. Moreover, they want students to take pride in who they are, in the fact that they belong to the Sabaot tribal group and, therefore, to be strong in their identity as Sabaot. Helen explained her perspective in the following way:

But [MT] also promotes that belonging. That belonging to the Sabaot. We feel that we belong to a Sabaot community. That matter of identification. It promotes identification. It is very nice.

The motivation of Sabaot teachers to implement the policy and to give the MT a place seems to stem, therefore, from their attitude towards the language itself and their perception of the beneficial effects MT can have in education and in ensuring the future of the Sabaot language.

## **Conclusion**

The study reported in this paper has revealed that the successful implementation of the Kenyan language-in-education policy depends largely on attitudes of local stakeholders (particularly teachers and parents) towards the MT and its use in school. Despite significant opposition and tremendous constraints, some teachers were willing to attempt to implement the language-in-education policy and give the MT a place if they themselves had a positive attitude towards it. Because teachers' attitudes are often changed through in-service training, the study has also shown that effective orientation of teachers to the rationale behind MT instruction is an essential component of successful policy implementation. Moreover, the study revealed that if parents have a negative attitude towards the use of MT in school, they can have a significant negative influence on policy implementation.

The study has highlighted, therefore, a number of factors that need to be addressed when language-in-education policies that support MT (like that of Kenya) are implemented. Firstly, greater attention needs to be given to teacher education during pre- or in-service training. Hand-in-hand with teacher education, instructional materials for content subjects need to be developed together with tests and examinations in these languages. Of equal importance, language-in-education planning needs to target parents and other stakeholders so that they support, rather than inhibit, teachers in their efforts to implement the policy. According to Breen (2002), effective implementation needs commitment from the whole school and community. All stakeholders, therefore, need to be made aware of the importance of the MT, both at home and in school, if they want to see their language flourish and their children succeed

in school, while at the same time they remain secure in their identity. As efforts are made to raise stakeholders' awareness, it may be helpful to expose stakeholders to positive examples of language revitalisation such as New Zealand Māori which have included the introduction of the language into the school system.

The study has revealed that there is a need for further research into language-in-education planning and implementation in minority language groups. For example, there is a need to conduct further research to discover tangible ways that indigenous languages, the students' MTs, can be promoted and maintained through education within a multilingual context. At the same time, ongoing research is needed to determine how teachers implement language-in-education policies that support MT as the medium of instruction in both linguistically homogenous and heterogeneous settings in postcolonial contexts such as Kenya.

## References

- Abagi, J. O., & Cleghorn, A. (1990). Teacher attitudes towards the use of English: Kiswahili and Mother Tongue in Kenyan primary classrooms. *Canadian and International Education*, 19, 61-71.
- Ager, D. E. (2005). Prestige and image planning. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 1035-1054). Mahwah, NJ: Lawrence Erlbaum.
- Alidou, H. (2004). Medium of instruction in post-colonial Africa. In J. W. Tollefson & A. B. M. Tsui (Eds.), *Medium of instruction policies. Which agenda? Whose agenda?* (pp. 195-214). Mahwah, NJ: Lawrence Erlbaum.
- Batibo, H. M. (2005). *Language decline and death in Africa: Causes, consequences and challenges*. Clevedon, England: Multilingual Matters.
- Benson, C. (2004). Bilingual schooling in Mozambique and Bolivia: From experimentation to implementation. *Language Policy*, 3, 47-66.
- Boiyo, F., Nakitare, Z., Okarau, R., Pkomu, J., Simotwo, J., & Taboi, J. (2004). *Keesuman ankeesir am kuutiinyoo kootaab 1* [We read and write our language in Standard 1]. Nairobi, Kenya: Bible Translation and Literacy E.A.
- Breen, M. (2002). From a language policy to classroom practice: The intervention of identity and relationships. *Language and Education*, 16(4), 260-283.
- Bunyi, G. W. (1996). *Language, education and social selection in Kenya: An ethnographic study of two schools*. Unpublished PhD, University of Toronto, Toronto, Canada.
- Cleghorn, A., Merritt, M., & Abagi, J. O. (1989). Language policy and science instruction in Kenyan primary schools. *Comparative Education Review*, 33(1), 21-39.
- Corson, D. (1999). *Language policy in schools: A resource for teachers and administrators*. Mahwah, NJ: Lawrence Erlbaum.
- Cummins, J. (2000). *Language, power and pedagogy: Bilingual children in the crossfire*. Clevedon, England: Multilingual Matters.
- Ellis, R., & Barkhuizen, G. (2005). *Analysing learner language*. Oxford: Oxford University Press.
- Fishman, J. A. (1991). *Reversing language shift: Theoretical and empirical foundations of assistance to threatened languages*. Clevedon, England: Multilingual Matters.
- Fishman, J. A. (2001). Why is it so hard to save a threatened language? (A perspective on the cases that follow). In J. A. Fishman (Ed.), *Can threatened languages be saved?* (pp. 1-22). Clevedon, England: Multilingual Matters.
- Fishman, J. A. (2006). *Do not leave your language alone: The hidden status agendas within corpus planning in language policy*. Mahwah, NJ: Lawrence Erlbaum.



- Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. New York: Aldine.
- Gordon, R. G. (Ed.). (2005). *Ethnologue: Languages of the world* (15th ed.). Dallas, TX: SIL International.
- Graham, B. E. (2007, 6-10 August). *Questions that demand answers*. Paper presented at the 5th Pan-African Reading for All Conference, University of Ghana.
- Heugh, K. (2005, 3-5 August). *Introduction II - Theory and practice - Language education models in Africa: Research, design, decision-making, outcomes and costs*. Paper presented at the Conference on Bilingual Education and the Use of Local Languages, Windhoek, Namibia.
- Hovens, M. (2002). Bilingual education in West Africa: Does it work? *International Journal of Bilingual Education and Bilingualism*, 5(5), 249-266.
- Ingram, D. E. (1989). Language-in-education planning. *Annual Review of Applied Linguistics*, 10, 53-78.
- Kenya: *Mt Elgon preliminary appeal*. (2007). Nairobi, Kenya: Kenya Red Cross Society.
- Kwalia, B., & Ng'etich, P. (2007, Tuesday, April 3). 6 killed as gun raid rocks district hqs. *Daily Nation*, pp. 1,4,
- Larsen, I., Surai, F., Kigai, K., & Mang'esoy, P. (1990). *Korooryo ku taay: Reading and writing Sabaot*. Nairobi, Kenya: Bible Translation and Literacy (E.A.).
- Liddicoat, A. J. (2005). Corpus planning: Syllabus and materials development. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 993-1011). Mahwah, NJ: Lawrence Erlbaum.
- Lo Bianco, J. (2004). Language planning as applied linguistics. In A. Davies & C. Elder (Eds.), *The handbook of applied linguistics* (pp. 738-762). Oxford: Blackwell.
- Mbaabu, I. (1996). *Language policy in East Africa: A dependency theory perspective*. Nairobi, Kenya: Educational Research and Publications.
- Ministry of Education. (1997). *Guidelines for Early Childhood Development in Kenya*. Nairobi, Kenya: National Centre for Early Childhood Education (NACECE) Kenya Institute of Education.
- Ministry of Education, Science and Technology. (2002). *Primary Education Syllabus* (Vol. 1). Nairobi, Kenya: Kenya Institute of Education.
- Ministry of Education, Science and Technology. (2004). *Primary Teacher Education Syllabus* (Vol. 1). Nairobi, Kenya: Kenya Institute of Education.
- Muthwii, M. J. (2002). *Language policy and practices in education in Kenya and Uganda: Perceptions of parents, pupils and teachers on the use of Mother Tongue, Kiswahili and English in primary schools*. Nairobi, Kenya: Phoenix.
- Nettle, D., & Romaine, S. (2000). *Vanishing voices: The extinction of the world's languages*. New York: Oxford University Press.
- Norton, B. (1997). Language, identity and the ownership of English. *TESOL Quarterly*, 31(3), 409-429.
- Nyaga, S. K. (2005). *An evaluation of the perceived long-term effects of teaching/ learning Kimenti for the subject Mother Tongue in a Tharaka school on children's language development and identity*. Unpublished Masters, University of Huddersfield, Huddersfield.
- Pauwels, A. (2004). Language maintenance. In A. Davies & C. Elder (Eds.), *The handbook of applied linguistics* (pp. 719-737). Oxford: Blackwell.
- Peddie, R. (2003). Languages in New Zealand: Population, politics and policy. In R. Barnard & T. Glynn (Eds.), *Bilingual children's language and literacy development* (pp. 8-35). Clevedon, England: Multilingual Matters.
- Siringi, S. (2007, Monday August 13). New squad formed to crush militia. *Daily Nation*, p. 8,
- Sluka, J. A. (1995). Reflections on managing danger in fieldwork: Dangerous anthropology in Belfast. In C. Nordstrom & A. C. G. M. Robben (Eds.), *Fieldwork under fire: Contemporary studies of violence and survival* (pp. 276-294). California, London: University of California Press.

- Starks, D. (2005). Other languages of New Zealand: Introduction. In A. Bell, R. Harlow, & D. Starks (Eds.), *Languages of New Zealand* (pp. 241-247). Wellington: Victoria University Press.
- Taylor, J. E. (2002). *The music of the Sabaot: Bridging traditional and Christian contexts*. Unpublished PhD, University of Edinburgh, Edinburgh.
- van Els, T. (2005). Status planning for learning and teaching. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 971-991). Mahwah, NJ: Lawrence Erlbaum
- Wabala, D., & Macharia, D. (2007, Monday March 5th). Two summoned over clashes. *Daily Nation*, p. 9,
- Watson-Gegeo, K. A. (1988). Ethnography in ESL: Defining the essentials. *TESOL Quarterly*, 22(4), 575-592.
- Wolcott, H. F. (2002). Ethnography? Or educational travel writing? In Y. Zou & E. T. Trueba (Eds.), *Ethnography and schools: Qualitative approaches to the study of education* (pp. 27-48). Oxford: Rowman & Littlefield.
- Wolff, H. E. (2005, 3-5 August). *Introduction - Background and history - Language and language-in-education policies in Africa: Problems and prospects*. Paper presented at the Conference on Bilingual Education and the Use of Local Languages, Windhoek, Namibia.

## **TEACHING GRAMMAR: A SURVEY OF EAP TEACHERS IN NEW ZEALAND**

**Roger Barnard & Davin Scampton**

*University of Waikato*

### **Abstract**

*This paper reports on a survey of New Zealand teachers' attitudes towards grammar and grammar teaching in their own particular teaching contexts. It uses a questionnaire adapted from that used in a survey of teachers of English for Academic Purposes (EAP) in British universities (Burgess & Etherington, 2002), followed by a series of email interviews with volunteer respondents. The findings of the present study indicate that, like the teachers reported in the 2002 study, EAP teachers in New Zealand appreciate the centrality of grammar in their language teaching and have a critical awareness of many of the problems and issues involved. There is also evidence to suggest that the teachers favour the treatment of grammar through its emergence in whole texts, rather than its presentation in decontextualised sentences and structures. In this regard, there is support for an approach tending towards Focus on Form (Long, 1991; Long & Robinson, 1998). However, the teachers' comments on the importance of systematic practice of grammatical features and detailed error correction suggests that there is a preference for more extensive treatment of grammatical issues than is usually suggested by proponents of a strictly incidental Focus on Form approach.*

### **Introduction**

The teaching of grammar continues to be a matter of controversy in the field of Applied Linguistics and Second Language Acquisition (SLA.) It is generally agreed that some attention to grammatical form is useful, perhaps necessary, but many issues related to the teaching of grammar still need further research, especially into the key social factors that are an inescapable element of classroom learning. Prominent among these social factors are the personal identities of the teachers and learners – and their individual and collective constructions of classroom reality. Increasingly, the beliefs and attitudes of practising teachers are being sought to shed light on theoretical concerns in the teaching of grammar, such as the nature of implicit/explicit learning, the way that grammar is best presented, the need for various types of evaluative feedback, the role of practice, etc. The study reported here took as its starting point a questionnaire slightly modified from that published a few years ago (Burgess & Etherington, 2002) which explored the attitudes of a group of British teachers of English for Academic Purposes (EAP) regarding the role of grammar teaching in their courses. The present study surveyed a similar group of EAP teachers in New Zealand, and followed this with a series of email interviews

with a self-selected sample of the respondents. For the purposes of this report, the following broad definition of EAP is used: “those educational activities in higher education, the purpose of which is the teaching and learning of the English language required by undergraduates and/or staff” (Kennedy, 2001, p.25).

## Literature review

### The teaching of grammar

The consensus among applied linguists involved in SLA is that language learning should have a primary focus on meaning within an overall communicative framework (Ellis, 2006). However, “the inability of communicative ESL teaching alone to promote high levels of accuracy in learners is now clear” (Fotos, 1998, p. 301). Hence, over the past decade there has been a re-focus on research into, and the practice of, grammar teaching (see Nassaji & Fotos, 2004). There is current theoretical debate –even sharp controversy (for example, Sheen 2003) – between applied linguists who argue for a focus on forms (FonFS), and those who propose a focus on form (FonF).

In the former, the teacher plans a series of lessons around specific grammar points in order to: promote an explicit understanding of grammar by a variety of means; provide written and oral exercises to practise the target form; and allow frequent opportunities for the (communicative) use of the target items. FonFS is most obviously exemplified by the Presentation, Practice, Production (PPP) model, although grammar can also be explicitly and inductively taught through activities such as ‘dictogloss’ (Wajnryb, 1990), or consciousness-raising tasks based on input texts (Ellis, 1992; Ellis & Gaies, 1999).

Focus on form, on the other hand, assumes an indirect, context-based focus on grammar, rather than overt, teacher-led instruction (Doughty & Williams, 1998). Ellis (2001a) has distinguished two types within this category: incidental and planned. Proponents of *incidental* FonF assume that classroom activity is based on communicative tasks and that attention to form should be paid only when grammatical difficulties arise which lead to (or are anticipated to lead to) a communicative breakdown. At this point, remedial treatment is effected by transitory corrective feedback, and when more extended grammar treatment is needed, this should be based on grammar problem-solving tasks, rather than forms-focussed instruction. *Planned* FonF, on the other hand, involves the treatment of pre-determined grammatical features but differs from FonFs because it occurs when the learners’ attention is primarily engaged in processing meaning. It may thus be seen as an intervening point in a continuum between the other two approaches. Various empirical studies investigating FonF in some English as a Second Language contexts have been recently conducted, notably by Ellis and his associates (Ellis, 2001a and b; Ellis 2002, Ellis, Basturkmen & Loewen, 2001; Ellis, Basturkmen & Loewen, 2002). FonF appears to be consistent with current SLA theories; however, the following statement made some years ago about KAL (Knowledge About Language) still, to a

large extent, applies today about the relationship between theories of grammar instruction and teachers' beliefs and practices:

Theoretical debates and popular discussions of rationales and models for KAL in the classroom have been informed by little empirical evidence regarding teachers' current beliefs and classroom practices in the area (Brumfit, Mitchell & Hooper, 1996, p.70).

It is to the issue of teachers' beliefs that attention is now turned.

### **Teacher cognition**

Research on teachers' beliefs and the relationship of those beliefs with pedagogical practice, originated in America in the early 1970s. Following the publication of *Life in Classrooms* (Jackson, 1968), the National Institute of Education of America published a report which enunciated the need for research on teachers' thought processes (instructional beliefs) more deeply. The report stated that:

It is obvious that what teachers do is directed in no small measure by what they think. Moreover, it will be necessary for any innovations in the context, practices, and technology of teaching to be mediated through the minds and motives of teachers. (National Institute of Education of America, 1975, p.1)

Since then, the relationship between teachers' thinking and the impact of their knowledge and beliefs on instructional practices has increasingly attracted educational researchers' attention, first in America (for example, Clark & Yinger, 1977), then elsewhere. In general, research on teachers' thought processes is based on the three major assumptions: (i) teaching is largely influenced by teacher cognition, (ii) teaching is guided by teachers' thoughts and judgments, and (iii) teaching constitutes a high-level decision-making process (Isenberg, 1990). As Clark and Peterson (1986) have claimed, teaching is "substantially influenced and even determined by teachers' underlying thinking" (p. 255). However, it has long been recognised that individuals' thinking processes and belief systems cannot merely be observed or measured, but instead must be inferred from what individuals say – and this has conventionally been investigated by attitude measurement techniques, often via questionnaires pioneered in the 1920s and 1930s by Thurstone, Likert and Gutmann.

Since that time, a great deal of research into teachers' cognition has been carried out, and a wide array of constructs has been employed – not only various categorisations of thinking, knowledge and beliefs, but also diverse operational definitions of conceptions, assumptions, values, principles, decision-making, attitudes and so on – and Borg (2006, p. 272) has argued the need for a shared terminological framework to be used. In the absence of such, the view will be taken in this paper that *attitudes* are the surface expression of underlying values, beliefs and knowledge, and may not fully represent those deeper constructs for various reasons such as: an individual's

lack of explicit awareness of those underlying constructs; an internal contradiction between and within these categories; and/or a simple inability, or unwillingness, to convey these to another person. Thus, at the best, soliciting teachers' attitudes is barely scratching the surface of much deeper cognitive processes, but one which – it may be argued – is a necessary first step towards more fully exploring cognitive processing. It is also evident that what a teacher believes or knows may not always correlate highly with his or her professional practice. Thus, investigations into teachers' beliefs should be balanced by consideration of their actual behavior in planning and executing classroom activities, and the extent of the convergence or divergence between beliefs and practice explored by both researchers and participant teachers. It is important to acknowledge this as a further limitation to the issues raised by the study reported in this paper.

### **Second language teacher cognition**

The field of Second Language Teacher Education has tended to lag behind mainstream educational research in its attempt to understand the cognitive dimensions of second language teaching (Johnson, 1992a). In the 1970s, interest in classroom-based research on second language teaching and teacher education focussed on three issues: effective teaching behaviours, positive learner outcomes, and teacher-student interactions (Freeman, 2002). Only in the past two decades have second language teacher education researchers (Freeman, 1989; Richards & Nunan, 1990; Johnson, 1992a, 1992b; Woods, 1996; Borg, 1998a and b, 1999, 2003a; 2003b; 2006; Borg & Burns, 2008) drawn attention to ways in which second language teachers' cognitive processes influence their classroom instruction.

A key work in this area was Woods' (1996) longitudinal study of planning and decision-making by eight teachers in Canadian ESL classrooms, which he carried out using multiple sources of data documentary analysis: teachers' logs, interviews, observations, and stimulated recall sessions. The book provides a wealth of detailed insights into both the decision making processes and the factors which shaped these, which Woods divided into internal and external groups:

*External* factors are situational factors which teachers take into account in making decisions (or to be accurate, what teachers know, assume and believe about these factors). *Internal* factors are ones internal to the decision-making process itself, i.e., the internal structuring of decisions and the relationships of decisions to each other. (Woods, 1996, p. 128).

One important contribution made by Woods to investigating language teachers' planning processes was his notion of BAK (Beliefs, Assumptions and Knowledge); rather than assume that these are distinct categories he considered them as separate points on a semantic continuum. In this respect, he echoed somewhat earlier thinking on the knowledge base of general teachers (Grossman, Wilson & Shulman, 1989).

### **Teachers' beliefs about grammar teaching**

Borg (2003a and b; 2006) reviewed 64 studies of language teacher cognition, of which all but twelve appeared after 1994, including Woods (1996) – the only case study at book length reported in the review. Borg divided his reviews into five parts, according to whether these studies related to: prior language learning experience, teacher education, classroom practice, teaching grammar, or literacy instruction. The 38 studies which focussed on grammar teaching were considered by Borg (2006) in terms of three distinctive sub-topics: (i) teachers' knowledge of grammar; (ii) surveys of teachers' beliefs about grammar teaching; and (iii) the relationships between teachers' beliefs about grammar teaching and their classroom practices. The findings from all of these studies suggest that teachers do indeed have a set of complex belief systems about the teaching of grammar, and that these are sometimes not reflected in their classroom practices for various complicated reasons. Borg concludes his consideration of the somewhat heterogenous collection of research reports by saying "there is no suggestion in any of these studies that formal instruction is becoming less prevalent in language classrooms" (Borg, 2003b, p.100).

Of particular relevance to the present study are two surveys. The first was carried out by Burgess and Etherington (2002) to identify the attitudes of 48 British teachers of EAP in UK universities. The findings from this survey clearly indicated that the responding teachers "appear to see grammar as important for their students and to have a sophisticated understanding of the problems and issues involved in its teaching" (Burgess & Etherington, 2002, p. 450). The teachers favoured discourse-based approaches, rather than decontextualised presentation of grammar items, with an inclination towards the use of authentic, full texts and real-life tasks for practice.

The second is a more recent survey (Borg & Burns, 2008) of beliefs about the integration of grammar and skills teaching. A questionnaire comprising both open- and closed-ended items was completed by 231 teachers of English from South America (2.5%), Asia (23.2%), Europe (25.7%) and Australia and New Zealand (46%). In this survey, the number of respondents from South America was very small (n=6) and the authors (p. 461) decided to exclude their responses from their analysis of the findings. Although 76% the remaining respondents worked in universities and other Adult Education Centres, there is no indication that EAP was in focus either in the questionnaire items or in the responses. Much of the authors' discussion of their findings (pp. 476-480) was taken with considering how their respondents perceived the relationship between grammar and the teaching of language skills in apparently General English (rather than specifically EAP) classes, but they also discussed their teachers' beliefs about the centrality of grammar to language instruction, and some of the points made will be considered in the final section of the present paper.

### **The present study**

The international survey reported by Borg and Burns (2008) came to the attention of the present researchers too late to influence their own survey, and other than this no

other research into New Zealand teachers' attitudes towards grammar and its integration into language teaching has been identified. It is not clear from the international survey how many of the respondents were from New Zealand, nor how many, if any, were working in EAP contexts, rather than in more general situations. Moreover, while there were many extracts of comments from respondents from other countries, none were reported as having been expressed by New Zealand teachers. Thus a survey of the attitudes of a sample of EAP teachers in New Zealand may be seen to complement previous studies and also, to a limited extent, to add to a general academic and professional understanding about what certain teachers believe about key aspects of grammar teaching. However, it is realized that such a survey, using questionnaires and interviews as the sole sources of information – a mere 'snapshot' of opinions of a small group at a specific moment in time – can at best be only a first step of what should be a more detailed and intensive investigation into the cognitive processes underlying teachers' classroom activity.

### **Research questions**

It was decided that the same research questions used by Burgess and Etherington (2002, p.437) would guide the present study:

1. Which beliefs about grammar and grammar teaching are most widely held by EAP teachers (in New Zealand)?
2. Is there a bias towards decontextualised presentation of grammar and away from discourse-based, unified approaches?

### **Survey instrument**

With the permission of Sian Etherington, a slightly modified version of the 2002 questionnaire was used (see Appendix 1): unlike the 2002 questionnaire, which had a five-point response scale, the version used in New Zealand solicited only four responses: strongly disagree, disagree, agree, and strongly agree. The reason for this was that a four-point scale facilitates a clear analysis of positive and negative responses; there is a tendency for many respondents to regress to the central point (#3 – no opinion, or neutral) in a five-point scale – a matter which led to some uncertainty in interpreting Burgess and Etherington's discussions of their findings. The questionnaire in the present study was written in HTML, XSSI and CSS to include university branding. It included radio buttons for respondents to indicate their preferences, and one open-ended question for additional comments, as well as five general background questions. The respondents were able to review their responses, and make any changes, before submitting the completed questionnaires, which then came directly to the researchers' email addresses when respondents clicked the 'submit' box. The researchers received no indication of the identity of the respondents.

### **Survey respondents**

An online search of all the New Zealand university websites produced a total of almost 200 teachers who were initially assumed to be teaching EAP. These were all contacted by email and advised of the scope and purpose of the study and their willingness and formal consent to participate was solicited, following the university's



standard ethical procedures for conducting human research. Many of those contacted responded that they did not currently teach EAP classes and therefore declined to participate; a few others simply did not respond. However, questionnaires were completed and returned by 32 teachers working at six of the eight universities in New Zealand, and at Unitec. All of these respondents taught EAP courses to undergraduate or pre-undergraduate students as all or part of their teaching duties. In terms of the types of classes and students they taught, and the length of EAP teaching experience, the New Zealand respondents were largely comparable with the British teachers reported by Burgess and Etherington. One relatively minor difference is that all of the British teachers taught on pre-sessional EAP classes (Burgess & Etherington, 2002, p. 437), whereas some of those in New Zealand were teaching students already enrolled in undergraduate EAP courses. As did Burgess and Etherington (pp. 437), the present authors readily acknowledge the possibility of ‘volunteer bias’ in the sample of those who took the time to complete and return the questionnaire. These 32 respondents were then asked if they would be willing to follow up the questionnaire with a series of email interviews.

### **Email interviews**

Although they obtained, and discussed, some qualitative comments from their respondents, the British authors acknowledged that “[t]he lack of follow-up interviews is a major limitation to the study” (Burgess & Etherington, 2002, 449). Therefore, in the New Zealand study, all the respondents were invited to participate in a series of follow-up interviews to flesh out the bare descriptive statistics of the questionnaires. Because of the geographical spread of the participants, it was decided that these discussions should be conducted with each participant through email messages. So, over a period of two months, three sets of questions were posed to each interviewee; initially, the same questions were asked of each participant, but in the second and third sets the general questions were preceded by some tailored queries in order to follow up previous comments by the particular interviewee. (An example is provided in Appendix 3.) The information thus collected was collated and subjected to a process of grounded analysis, whereby all the data were interrogated to identify common and contrasting opinions on two or three themes.

### **Interview participants**

Eleven of the survey respondents volunteered to take part in the email interviews, all of whom had more than five years of EAP teaching experience, and all but one held an MA and/or doctorate in applied linguistics. All were teaching undergraduate (support) or pre-undergraduate (access) programmes in the same six universities and Unitec as represented in the survey, and the students they taught were within the IELTS range 4.5 - 6.5; the median being 5.5. The following data were taken from nine interviewees (three female and six male), as one withdrew after the first set of questions due to pressure of work, and another was working overseas for his university at the time, and it was felt that his responses, though valuable, would not assist an understanding of EAP teaching within a New Zealand context.

The questionnaire and – especially – the interviews provided a wealth of interesting information about the respondents' attitudes towards grammar teaching. However, Sections 4 and 5 below report only those data relating specifically to the two research questions stated above: the centrality of grammar to the respondents' EAP classes, and how grammar might be introduced and practiced.

## **Quantitative Findings: Data emerging from the preliminary survey**

The following section presents the findings (percentages rounded to the nearest tenth of a percentile) emerging from the New Zealand questionnaire, and the sequence of the discussion follows that in Burgess and Etherington (2002, pp. 440-447). For full details, please refer to Appendices 1 and 2, which provide tabulated data.

### **The role of grammar in language**

Most of the New Zealand teachers surveyed considered that grammar plays a central role in language. Two thirds (68.8%) agreed that the role of grammar is *a framework for the rest of the language* (statement 1.1) and the same number rejected the idea that grammar is merely...*a refinement for a more basic linguistic knowledge* (statement 1.3). Over half (56.3%) agreed that grammar *can be seen as blocks of language combined to create the whole* (statement 1.2), and 71.9% concurred that *grammar can be seen as an equal backbone for sufficient linguistic competence* (statement 1.4).

### **Explicit grammar teaching**

The majority of the respondents (71.9%) agreed or strongly agreed with statement 23, *My students expect teachers to present grammar points explicitly* – a point reinforced by the positive responses of just over two thirds to statement 33, *A lack of explicit grammar teaching leaves my students feeling insecure*. This possibility is strengthened by the 80.7% concurrence with statement 20, *Explicit discussion of grammar rules is helpful for students*. These findings strongly support the view that a direct approach to grammar instruction seems to be preferred by the New Zealand students – although it is important to note, as did Burgess and Etherington (2002, p. 441), that the above attitudes – while ostensibly focussing on the learners' needs and wishes – may actually reflect the pedagogical preferences of the teachers.

### **Instruction vs. exposure**

The notion that grammar can be learned through exposure to language in natural use (statement 2) was agreed by 62.6% of the respondents, although 80.6% of them also agreed or strongly agreed with statement 3, *Formal instruction helps learners to produce grammatically correct language*. These findings are not necessarily contradictory, as they are linked to the following issues.

### **Declarative and procedural knowledge**

Almost two-thirds (65.7%) of the respondents either agreed or strongly agreed with statement 21, *My students find it difficult to transfer their grammatical knowledge*

into communicative language use. The point was specifically taken up in the email interviews and some of the teachers' comments are reported in Section 5 below.

### **The importance of conscious knowledge**

Three statements sought teachers' attitudes towards the role of students' conscious knowledge. Statement 4, *Student use of language does not involve conscious knowledge of the grammatical system and how it works*, received positive responses from only 28.1% of the teachers, a point reinforced by almost two thirds (62.5%) agreeing or strongly agreeing with statement 6, *Students need a conscious knowledge of grammar in order to improve their language*. On the other hand, a very similar proposition in statement 9, *Students need to be consciously aware of a structure's form and its function before they can use it proficiently*, did not produce conclusive results, as only 46.9% of the New Zealand teachers agreed with this point.

### **Comparison and contrast of structures**

Two thirds (67.7%) of the respondents agreed or strongly agreed with statement 17, *Comparison and contrast of individual structures is helpful for students learning grammar*; of the dissenting third, only one teacher strongly disagreed.

### **The use of grammatical terminology**

78.2% of the teachers agreed with statement 34, *My students find grammatical terminology useful*. Burgess and Etherington (2002, p. 444) suggested that the British teachers' similar preference for explicit grammar teaching may be linked to their students' previous experience of grammar-based language learning, and this is possibly also the case in New Zealand. In contrast to the positive responses above, statement 39, *My students find it difficult to use grammatical terminology*, did not produce conclusive results, as there was an exact 50:50 split between those who agreed and those who disagreed, although more (12.5%) strongly agreed than strongly disagreed (3.1%). Here, the distinction may be made between reporting the students' preference for the teacher's use of grammatical terminology in contrast to the students' own metalinguistic competence.

### **Problem solving**

Statement 22, *My students are motivated by problem-solving techniques for learning grammar*, was agreed to by 67.7% of the respondents. Similarly, 61.3% of them disagreed with the contrasting statement 40, *My students are frustrated by problem-solving techniques for learning grammar*, thus confirming the learners' apparent preference for this approach to grammar learning, or at least the teachers' assumptions that this is the case.

### **Error correction**

83.9% of the teachers agreed that form-focused correction helps students to improve their grammatical performance (statement 18). Two thirds (68.8%) rejected the proposition that *Teachers find it difficult to correct student errors of grammar within a written communicative context* (statement 35). However, they were almost equally

divided in their responses to statement 16, *Teachers should only correct student errors of form which interfere with communication*, with 43.8% agreeing and the majority disagreeing.

### **Presentation in authentic, complete texts**

75% of the teachers showed agreement or strong agreement with statement 15, *Students learn grammar more successfully if it is presented within a complete text*. It was clear that the vocabulary used in authentic texts presented more difficulties than grammatical features; 78.2% supported statement 29, *My students find authentic texts difficult because of the vocabulary used*, but it was less clear whether *My students find it difficult to handle grammar presented within authentic texts* (statement 26), as there was a 50:50 split in the responses, although there was more strong agreement (15.6%) than strong disagreement (6.3%) on this issue. Just over two thirds of the teachers (68.8 %) disagreed that the use of authentic material was too time-consuming (statement 31). On the whole, the responses of the New Zealand teachers, like those of their British counterparts, indicate “no general feeling that authentic texts take too much time in the classroom or in preparation” (Burgess & Etherington, 2002, p. 446).

### **The role of practice**

Very positive attitudes towards the role of practice were evident. 80.7% of the respondents concurred with statement 5, *Students can improve their grammatical accuracy through frequent practice of structures*. Likewise, 84.4% agreed with statement 12, *Productive practice of structures is a necessary part of the learning process*. In both cases, there was a marked degree of strong agreement (19.4% for statement 5 and 40.6% for statement 12). Over two thirds (68.8%) also supported the view that *Participating in real-life tasks with language is the best way for students to develop their grammatical knowledge* (statement 14).

The distinction between practice (of structures) and ‘real life’ tasks was not clear in the questionnaire, and this point was taken up in the subsequent interviews.

## **Qualitative Findings: data emerging from the email interviews**

The first research question of this study sought to elicit EAP teachers’ beliefs about grammar and its role in their language teaching. Six of the nine interviewed teachers reinforced the point made by the questionnaire respondents that grammar had a central place in their classes. One of the points in the first interview asked the interviewees to “...explain the extent to which you focus on grammatical issues in the above class.” The following comments (using pseudonyms) were made:

There would be some focus on grammar in all of my EAP lessons (Colin).

Grammar is an important component in the paper (Charles).

A great deal of focus, because the students are very inaccurate with grammar (Brian).

It occurs in most lessons (Kenneth).

Form is the focus of the whole course (Michael).

Although I include in my weekly planning some sessions of explicit grammar teaching, a lot of incidental explanation occurs almost every day ... Students prefer explicitness when they are taught grammar and I feel they believe grammar is everything, more important than writing and reading (Emily).

There were similar comments which clearly indicated that most of the interviewed teachers believed that an *explicit* focus on grammar was important, although the following comments were also made:

I really don't overtly teach it all (Carol).

The overall approach to the course places very little focus on grammar (Simon).

We do not focus on grammar unless it arises in the course of the lesson (Penny).

The teachers took into account their learners' background when deciding to adopt an explicit focus on grammar, for example:

I would be fairly confident that the past English language learning experience of the majority of learners attending our programmes would be heavily based around *well structured courses based around the grammatical functions of English* (Simon –emphasis in original).

And some compared students from different language backgrounds:

From experience, it has always been the non Asian students that do not know the grammar rules (the Pacific Island students or the Middle Eastern students) whereas the Chinese, Malaysians, Japanese, and Koreans know them and are able to answer the questions. So I think there is some difference in knowledge levels depending on how they are taught English (Carol).

The interview data from these teachers indicated that, like their British counterparts, they had a thoughtful and “sophisticated understanding” (Burgess & Etherington, 2002, p.450), of the issues involved. These will be revealed in the following discussion, which focuses on the interview responses addressing the second of the research questions posed above:

Is there a bias towards decontextualised presentation of grammar and away from discourse-based, unified approaches?

The question was soundly rejected; the majority of the interviewees supported an approach based upon a discourse-based treatment of grammatical issues within whole texts, rather than through isolated, decontextualised presentation.

Because the actual questionnaire responses of the interviewees were unknown, they were asked to comment on the following item: #15 *Students learn grammar more successfully if it is presented within a complete text*. Seven out of nine agreed; one strongly agreed; one disagreed. The disagreement came from Penny, who explained that: “if grammar points do arise, I deal with them but not in great detail. ... Because they are having a grammar focus in their Text Based Skills class, this leaves me free to focus on EAP alone.” There was otherwise a general feeling that “isolated grammar has no contextual relevance to many students” (Brian) and “if ... we look at the grammar in the context of a written piece of text then I think that this makes more sense to them rather than simply teaching grammar in a piecemeal way” (Carol). Another teacher put it this way: “I believe that if students are exposed to grammar in the context within which it is used, they are more likely to remember and use such grammar when they find themselves (using language) in a similar context” (Simon).

A wide range of texts were used by the teachers to draw attention to grammatical issues, ranging from journal articles and abstracts to the local and international newspapers to:

an assignment that I had cobbled together from theirs – I would have picked out particularly bad examples of grammar that were particularly recurring, would have highlighted the issues and would lead a general discussion on the whys and wherefores of the mistakes (Carol).

They were asked what difficulties their students might have when dealing with grammar within complete texts. Perhaps not surprisingly, “it varies according to level – naturally – also is somewhat dependent upon their first language and previous academic achievement level” (Brian). More specifically linguistic issues were mentioned, such as

difficulties in nuance especially of time and aspect ... modality can cause lots of problems as it is quite often badly taught in course books and teachers tend to be scared of it as it's hard to get an overall handle on (Emily).

When asked if the teachers themselves found difficulties in this approach, there was general feeling that the only problem was that of time – for example:

I have no difficulties in presenting grammar in this way; on the contrary, students seem to enjoy this discourse / grammar approach. Our real problem is time because a change in how they look at a text requires time and practice, more than the four weeks between two assessment tests (Emily).

Another comment indicated the general freedom of choice enjoyed by university teachers:

I don't have any difficulty with presenting grammar problems in the way that I have outlined as it is not imposed on me by anyone else. I would have a problem if someone told me I had to do it in a certain way, but as it is my way ... I am relatively confident of and with, the students in front of me, then I don't have any problems (Carol).

The issues of practice and correction stimulated a great deal of comment. As one interviewee said with regard to the difficulty students have in transferring grammatical knowledge into communicative language use: "the majority of students have a good grasp of grammar, but the problem is getting them to apply their knowledge and write meaningful sentences" (Charles). Another comment was: "I find that many students lack awareness of grammar and this hinders effective academic writing" (Brian). There was firm agreement or strong agreement among the interviewees on the need for practice (in response to items #5, 12 and 17 on the questionnaire; please refer to Appendix 1), although notions of what constitutes 'practice' varied. For some, this was a matter of skills and strategy training, rather than focus on grammatical features. For example:

I train them and hope that they learn skills, strategies and that critical thinking and other faculties are honed. This involves practice (Brian).

Another interviewee put it like this:

Tasks that require students to involve certain language structures to successfully complete a task (or to complete it to a higher standard) are also given as additional 'homework' tasks (Charles).

Some interviewees espoused an incidental approach to the practice of forms:

Practice is a key part of classroom sessions. ... I would prefer to think of learning as more of a practice of language using target structures rather than as practice of the target structures themselves (Simon).

Another made a similar remark, but implied that she might have liked to have more forms-focussed practice:

Practice is a very important part of my classroom practice ... but the practice tasks I am referring to are to do with assignment/ report construction, not grammar. In an ideal world ... I would have grammar tasks all the time to reinforce the structures taught and let them see the constructions in a real written form (Carol).

One very experienced EAP teacher clearly indicated a reactive focus-on-form approach:

I recently responded to a student enquiry in class about ‘reduced’ relative or subordinate clauses. I gave the question some whiteboard focus, using examples from a text we had been studying ... I spent between 5 and 10 minutes on it and then moved on (Colin).

Here, however, the focus on form reported by the teacher went beyond the beyond “overtly draw(ing) students’ attention to linguistic elements as they arise *incidentally* in lessons whose overriding focus is on meaning or communication” (Long, 1991, pp. 45-46 – emphasis added), which is a central feature of a ‘strong’ form of *incidental* Focus on Form. Others reported an approach similar to planned FonF (Ellis 2001a) regarding grammar points which might give rise to particular problems:

It depends on whether you mean formal practice or productive use ... If by practice you mean solely for pedagogic purposes, I would say, only for “intricate structures”, which require special attention to their formal features (e.g. conditionals), and which for this reason are difficult to ‘pick up’ accurately (Kenneth).

Another interviewee was explicit about her proactive strategy to raise her students’ consciousness of how forms are contextualized:

I prepare a significant number of worksheets designed for individual, pair and group work. As the modules that we run are driven by topic of subject, I try to use a variety of texts, which contain the same grammar structure, which students have to identify, and then compare to the way it was used in previous texts (Emily).

Overall, the interviewees’ comments on issues relating to practice indicated that the attention to grammar was secondary to other aspects of language, such as the appropriate organisation of written texts. When they did deal with grammatical issues, however, there was clear evidence that this was more than an “*occasional* shift of attention to linguistic code features” (Long and Robinson, 1998, p. 23 – emphasis added). As one interviewee said: “Students ultimately need to – and in most cases want to – minimise errors” (Kenneth). This attention to form is also reflected in their comments about correction: reinforcing the general agreement of the questionnaire respondents to item #18, all of the interviewees agreed that form-focused correction helps students to improve their grammatical performance, and most did not find any particular difficulties in correcting grammar errors in either written or oral communication, which they did in various ways – such as:

When I mark an assignment, I will circle or underline mistakes, but I don’t re-write things for them, sometimes I will add ‘s’ if there is a plural



mistake and sometimes I will cross out words or phrases that are repetitious or redundant – and I will write ‘repetitious’ on their assignments (Carol).

This teacher also uses a marking code, as does at least one other:

On the assumption that the fewer but more emphatic points of feedback are more effective, I use a code of no more than half a dozen symbols. I’m reviewing this however ... (Colin).

Much the same point is made by another interviewee:

It is often a good idea to concentrate upon several errors, correct them and then move on to other errors. Picking up too many errors at once can sometimes make a student feel as if their writing is beyond correction (Charles).

This teacher, as well as others, made the point that corrective feedback could also be provided by the students themselves.

## **Conclusion**

The New Zealand teachers in this study, like the British teachers in the 2002 survey, favoured the treatment of grammatical features in complete texts, rather than in isolation. Burgess and Etherington conclude their report (2002, p. 450) by saying “[T]he results paint a picture of the approaches to grammar teaching taken in EAP courses across the UK which may be encouraging to those who advocate a Focus on Form approach.” Consistent with this view, the New Zealand EAP teachers rejected a strictly forms-focussed (FonF) strategy with a pre-determined grammatical syllabus and emphatically preferred to deal with grammatical issues as they arose from the texts used to develop generic EAP skills. To this extent, their approach may be regarded as incidental and generally reactive, although there were some instances where they adopted proactive (planned) FonF strategies – for example, by drawing up forms-focussed worksheets for systematic and perhaps extensive treatment in subsequent classes. However, while following an approach based on FonF, it is also clear that their treatment of the emerging grammatical issues was generally far from the ‘transitory’ remediation suggested by many proponents of a FonF strategy (for example, Long & Robinson, 1998). This is evident from the attention they paid to extensive practice (both structural and otherwise) and in the importance they – and their students – attached to the explicit correction of formal errors, for example by the use of correcting codes. In this respect, too, the range of reactive practices espoused by the New Zealand teachers resembles those of the international respondents (Borg & Burns, 2008, p. 471).

The findings of the present study also align with the portrait of grammar teaching reported in the 2008 international survey as being one “characterised by regular phases of explicit work, a desire to encourage students to discover rules (without discounting the use of direct explanation), and regular opportunities for grammar practice” (Borg & Burns, 2008, p. 477). There was an implication among these international respondents of a broadly-based Presentation-Practice-Production approach to L2 teaching – an approach which was completely absent in the New Zealand interviews; this is almost certainly due to the aims of EAP programmes, rather than general language teaching (which may be inferred from the comments of many of the international respondents). Another interesting finding of the international survey (Borg & Burns, 2008, p. 479) was the absence among the respondents of any technical language, such as ‘focus on form’, despite the generally high level qualification held by the majority of the teachers surveyed. This lack of explicit reference to metalanguage was also noticeable in the at least equally-qualified New Zealand interviewees, although there was use of some terms such as ‘discourse-based approach’, which may have been stimulated by the focus questions. As has been pointed out by Borg and Burns and others, this apparently atheoretical standpoint raises interesting questions about the role of theory in second language teacher development and practice. However, Borg and Burns (2008, p. 479) go on to say: “The largely experiential nature of teachers’ evidence base in justifying their work ... raises questions about the reliability of their judgements about its effectiveness. These are issues which merit continuing study.” The implication here is that professional experience may be unreliable; it may also be the case that some SLA theoretical positions may be invalid. Certainly, it is necessary to investigate the relationship between explicit theories as published in academic works, and the implicit theories developed from teachers’ professional practice.

As previously noted, the New Zealand study was a very limited one. While an attempt was made to contact the entire population of EAP teachers through a search of university websites, there is little doubt that many potential respondents were overlooked. Moreover, only 32 actually completed the questionnaire, and this self-selected sample may not have been typically representative of the wider population – a volunteer bias. Despite this, it is felt that the results which emerged from this survey had some concurrent validity with the findings reported by Burgess and Etherington (2002), Borg and Burns (2008), and several other studies reported by Borg (2006). Another limitation of the present study was the restricted number of interviewees (eventually, nine). Although this was a small sample, it did represent all the institutions covered in the survey and a great deal of useful information about their attitudes and practices were provided, of which only a fraction has been reported in this paper. While thus limited and not amenable to generalisation, the interview data enabled the present researchers to get below the surface opinions revealed in the questionnaire to explore some teachers’ self-reported beliefs and practices in more depth. The usefulness of the findings may be considered in terms of the extent to which the interview extracts reported above may be relatable to the readers’ own professional experiences and background knowledge of the field. Finally, the

inevitable limitation of time-constraints has to be considered. The data were collected over a three-month period, and thus merely reflect attitudes expressed at a particular moment in time; moreover, the research participants were all practising professionals with busy schedules and may not have been able to give as much thought as they would have liked to either the questionnaire or to the interview questions. These points do not necessarily invalidate the findings, but they point to the tentative and partial nature of any implications that are drawn from them.

The project also suggests directions for further research in the investigation of teachers' beliefs and their influence on professional practice, both within the New Zealand academic context, and beyond. In the first place, the present researchers were aware (as were Burgess & Etherington) of some improvements that could have been made to the questionnaire. However, in this case, it was decided not to make any change to the 2002 instrument (other than reduce the five-point scale to four) in order to provide a clear distinction between agreement and disagreement, and to administer the questionnaire online (which proved to be very time-effective and guaranteed complete anonymity to the respondents). It would be appropriate for future survey research in this area to more closely tailor questionnaires to the specific context under consideration, and to consider very carefully the number of scales provided. Secondly, while the decision to interview participants by email was made on the pragmatic grounds of overcoming geographical distance, the present researchers found that asynchronous interviewing in this way provided more opportunities for timely reflection on the issues by both the researchers and the participating teachers than would have been possible in face-to-face contexts. We would recommend this approach in further investigations. Finally, while the interview data provided useful information about teachers' practice, it is evident that – as Borg (2006) and Borg and Burns (2008) have stressed – expressed attitudes need to be triangulated with observed activity. Thus, more fruitful research would seek to identify, and explore, the extent of the convergence and divergence between attitudes, assumptions and knowledge expressed by teachers and their actual classroom behaviour. Finally, as Borg (2006) again emphasises, a fully-fledged investigation into the whole issue of cognition and practice would be to explore the same issues from the learners' perspective and here too make appropriate comparisons. This constitutes a very ambitious research agenda, of which the present study has been merely the first tentative steps, but one which we feel useful and interesting – and, perhaps, essential – if we really wish to understand the realities of classroom learning.

## References

- Borg, S. (1998a). Talking about grammar in the foreign language classroom. *Language Awareness*, 7, 159-175.
- Borg, S. (1998b). Teachers' pedagogical systems and grammar teaching: A qualitative study. *TESOL Quarterly*, 32(1), 9-38.
- Borg, S. (1999). Teachers' theories in grammar teaching. *ELT Journal*, 53(3), 157-167.
- Borg, S. (2003a). Teacher cognition in language teaching: A review of research on what language teachers think, know, believe and do. *Language Teaching*, 36, 81-109.

- Borg, S. (2003b). Teacher cognition in grammar teaching: A literature review. *Language Awareness*, 12(2), 96-108.
- Borg, S. (2006). *Teacher cognition and language education: Research and practice*. London: Continuum.
- Borg, S., & Burns, A. (2008). Integrating grammar in Adult TESOL classrooms. *Applied Linguistics*, 29(3), 456-482.
- Brumfit, C., Mitchell, R., & Hooper, J. (1996). Grammar, language and classroom practice. In M. Hughes (Ed.), *Teaching and learning in changing times* (pp.70-87). Oxford: Blackwell.
- Burgess, J., & Etherington, S. (2002). Focus on grammatical form: Explicit or implicit? *System*, 30, 433-458.
- Clark, C., & Yinger, R. (1977). Research on teacher thinking. *Curriculum Inquiry*, 7, 279-304.
- Clark, C. M., & Peterson, P.L. (1986). Teachers' thought processes. In M. C. Wittrock (Ed.), *Handbook of research and teaching* (3rd ed., pp. 255-296). New York: Macmillan.
- Doughty, C., & Williams, J. (1998). Issues and terminology. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 1-12.). Cambridge: Cambridge University Press.
- Ellis, R. (1992). *Second language acquisition and pedagogy*. Clevedon, England: Multilingual Matters.
- Ellis, R. (2001a). Investigating form-focused instruction. *Language Learning*, 51, Supplement 1, 1-46.
- Ellis, R. (2001b) *Form-focused instruction and second language learning*. Malden, MA: Blackwell.
- Ellis, R. (2002). Does form-focused instruction affect the acquisition of implicit knowledge? A review of the research. *Studies in Second Language Acquisition*, 24(2), 223-236.
- Ellis, R. (2006). Current issues in the teaching of grammar: An SLA perspective. *TESOL Quarterly*, 40(1), 83-108.
- Ellis, R., Basturkmen, H., & Loewen, S. (2001). Learner uptake in communicative ESL lessons. *Language Learning*, 51, 281-318.
- Ellis, R., Basturkmen, H., & Loewen, S. (2002). Doing focus on form. *System*, 34, 419-432.
- Ellis, R., & Gaies, S. (1999). *Impact grammar: Grammar through listening*. Hong Kong: Addison Wesley Longman Asia ELT.
- Fotos, S. (1998). Shifting the focus from forms to form in the EFL Classroom. *ELT Journal*, 52(4) 301-307.
- Freeman, D. (1989). Teacher training, development and decision making model: A model of teaching and related strategies for language teacher education. *TESOL Quarterly*, 23(1), 27-45.
- Freeman, D. (2002). The hidden side of the work: Teacher knowledge and learning to teach. *Language Teaching*, 35, 1-13.
- Grossman, P. M., Wilson, S. M., & Shulman, L. S. (1989). Teachers of substance: Subject matter knowledge for teaching. In M. C. Reynolds (Ed.), *Knowledge base for the beginning teacher* (pp. 23-36). Oxford: Pergamon.
- Isenberg, J. P. (1990). Teachers thinking and beliefs and classroom practice. *Childhood Education*, 66, 322-327.
- Jackson, P.W. (1968). *Life in classrooms*. New York: Holt, Rinehart and Winston,
- Johnson, K. E. (1992a). The relationship between teachers' beliefs and practices during literacy instruction for non-native speakers of English. *Journal of Reading Behavior*, 24(1), 83-108.
- Johnson, K. E. (1992b). Learning to teach: Instructional actions and decisions of preservice ESL teachers. *TESOL Quarterly*, 26(3), 507-535.
- Kennedy, C. (2001). Language use, language planning and EAP. In J. Flowerdew & M. Peacock (Eds.), *Research perspectives on English for academic purposes* (pp. 25-41). Cambridge: Cambridge University Press.
- Long, M.H. (1991). Focus on form: A design feature in language teaching methodology. In K. de Bot, R. Ginsberg, & C. Kramsch (Eds.), *Foreign language research in cross-cultural perspective* (pp.39-52). Amsterdam: John Benjamins.

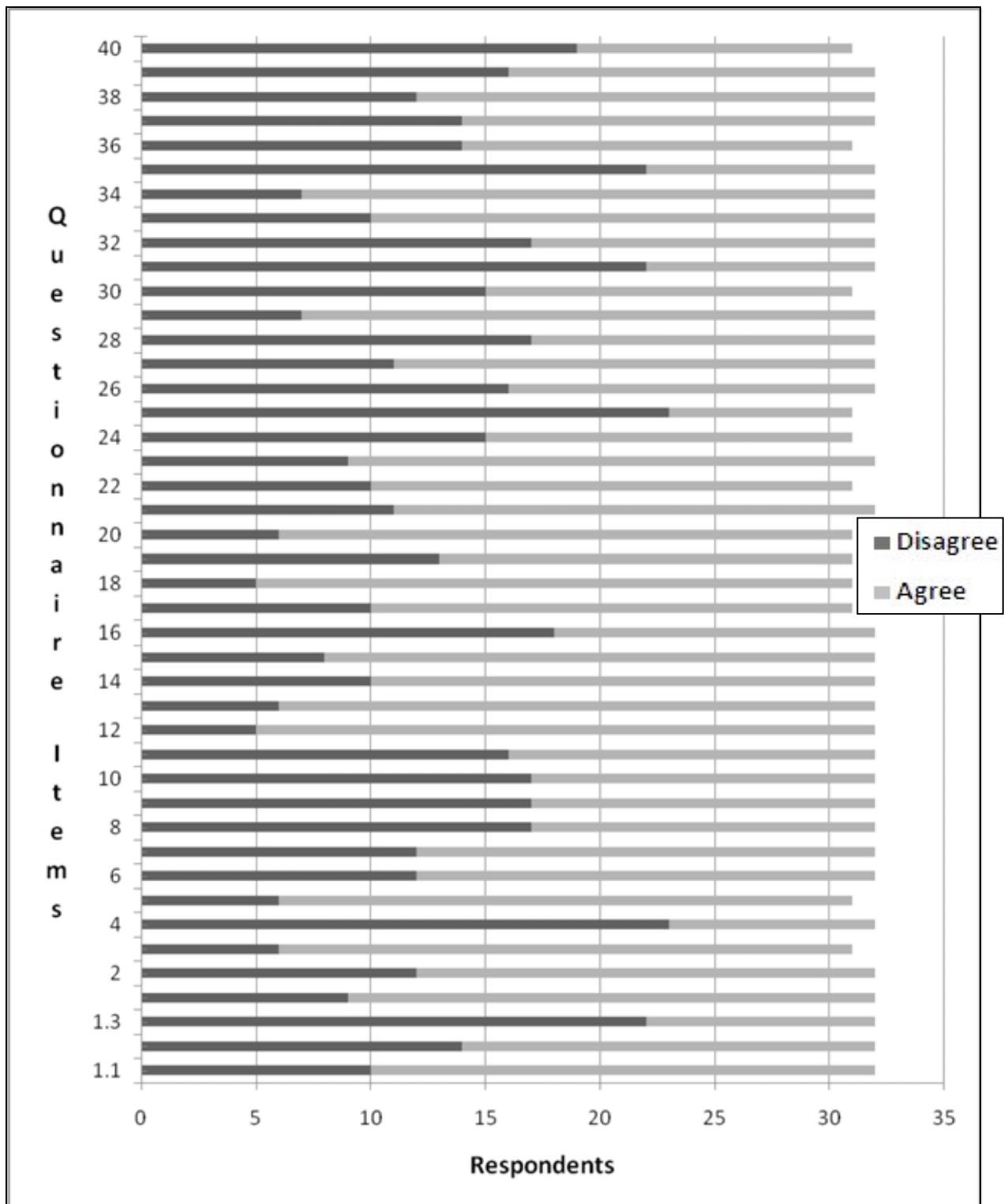
- Long, H., & Robinson, P. (1998). Focus on form: Theory, research and practice. In C. Doughty, & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 15-41). Cambridge: Cambridge University Press.
- Nassaji, H., & Fotos, S. (2004). Current developments in research in the teaching of grammar. *Annual Review of Applied Linguistics*, 24, 126-145.
- National Institute of Education of America. (1975). *Teaching as clinical information processing [microform]: NIE Conference on Studies in Teaching: Panel 6* [N.L. Gage, editor; K. Viehoveer, coordinating editor]. Washington, DC: National Conference on Studies in Teaching 1974.
- Richards, J. C., & Nunan, D. (1990). *Second language teacher education*. Cambridge: Cambridge University Press.
- Sheen, R. (2003). Focus on form: A myth in the making? *ELT Journal*, 57(3), 225-233.
- Wajnryb, R. (1990). *Resource books for teachers: Grammar dictation*. Oxford: Oxford University Press.
- Woods, D. (1996). *Teacher cognition in language teaching*. Cambridge: Cambridge University Press.

**Appendix 1: Questionnaire**

	Strong- ly Dis- agree	Dis- agree	Agree	Strong- ly Agree	N
<b>#1.1</b> <i>The role of grammar in language is as a framework for the rest of the language – a basic system to build everything else on.</i>	0	31.3	46.9	21.9	32
<b>#1.2</b> <i>The role of grammar in language is as the building blocks of language which are combined to form a whole.</i>	6.3	37.5	37.5	18.8	32
<b>#1.3</b> <i>The role of grammar in language is as something which is added on to language proficiency: a refinement of more basic language knowledge.</i>	28.1	40.6	25.0	6.3	32
<b>#1.4</b> <i>The role of grammar in language is as an equal pillar in supporting language proficiency. (Other pillars could be knowledge about pronunciation, appropriacy or culture etc.)</i>	15.6	12.5	43.8	28.1	32
<b>#2</b> <i>Students can learn grammar through exposure to language in natural use.</i>	6.3	31.3	43.8	18.8	32
<b>#3</b> <i>Formal instruction helps learners to produce grammatically correct language.</i>	3.2	16.1	41.9	38.7	31
<b>#4</b> <i>Student use of language does not involve conscious knowledge of the grammatical system and how it works.</i>	25.0	46.9	12.5	15.6	32
<b>#5</b> <i>Students can improve their grammatical accuracy through frequent practice of structures.</i>	6.5	12.9	61.3	19.4	31
<b>#6</b> <i>Students need a conscious knowledge of grammar in order to improve their language.</i>	9.4	28.1	37.5	25.0	32
<b>#7</b> <i>Practice of structures must always be within a full, communicative context.</i>	9.4	28.1	28.1	34.4	32
<b>#8</b> <i>Separate treatment of grammar fails to produce language knowledge which students can use in natural communication.</i>	3.1	50.0	21.9	25.0	32
<b>#9</b> <i>Students need to be consciously aware of a structure's form and its function before they can use it proficiently.</i>	18.8	34.4	28.1	18.8	32
<b>#10</b> <i>The separation of work with a grammar focus from the rest of the language syllabus is useful for students.</i>	18.8	34.4	43.8	3.1	32
<b>#11</b> <i>Decontextualised practice of structures has a place in language learning.</i>	15.6	34.4	37.5	12.5	32
<b>#12</b> <i>Productive practice of structures is a necessary part of the learning process.</i>	0	15.6	43.8	40.6	32
<b>#13</b> <i>Grammar is best taught through work which focuses on message.</i>	6.3	12.5	46.9	34.4	32
<b>#14</b> <i>Participating in real-life tasks with language is the best way for students to develop their grammatical knowledge.</i>	6.3	25.0	43.8	25.0	32
<b>#15</b> <i>Students learn grammar more successfully if it is presented within a complete text.</i>	6.3	18.8	40.6	34.4	32
<b>#16</b> <i>Teachers should only correct student errors of form which interfere with communication.</i>	9.4	46.9	34.4	9.4	32

<i>#17 Comparison and contrast of individual structures is helpful for students learning grammar.</i>	3.2	29.0	41.9	25.8	31
<i>#18 Form-focused correction helps students to improve their grammatical performance.</i>	6.5	9.7	58.1	25.8	31
<i>#19 Grammar is best taught through a focus on individual structures.</i>	9.7	32.3	41.9	16.1	31
<i>#20 Explicit discussion of grammar rules is helpful for students.</i>	0	19.4	35.5	45.2	31
<i>#21 My students find it difficult to transfer their grammatical knowledge into communicative language use.</i>	6.3	28.1	43.8	21.9	32
<i>#22 My students are motivated by problem-solving techniques for learning grammar.</i>	3.2	29.0	54.8	12.9	31
<i>#23 My students expect teachers to present grammar points explicitly.</i>	9.4	18.8	40.6	31.3	32
<i>#24 My students prefer to learn grammar from one sentence examples</i>	3.2	45.2	41.9	9.7	31
<i>#25 My students prefer to find matches between meaning and structure for themselves.</i>	19.4	54.8	19.4	6.5	31
<i>#26 My students find it difficult to handle grammar presented within authentic texts.</i>	6.3	43.8	34.4	15.6	32
<i>#27 My students find authentic texts difficult because of the wide variety of structures which appear.</i>	9.4	25.0	37.5	28.1	32
<i>#28 My students find authentic texts difficult because they are too culture bound.</i>	9.4	43.8	34.4	12.5	32
<i>#29 My students find authentic texts difficult because of the vocabulary used.</i>	9.4	12.5	43.8	34.4	32
<i>#30 My students cannot find form-function matches in authentic texts without explicit direction from teachers.</i>	6.5	41.9	38.7	12.9	31
<i>#31 Teachers find the use of authentic material too time-consuming.</i>	31.3	37.5	21.9	9.4	32
<i>#32 Teachers find it difficult to produce tasks of a suitable level from authentic texts</i>	21.9	31.3	31.3	15.6	32
<i>#33 A lack of explicit grammar teaching leaves my students feeling insecure.</i>	6.3	25.0	40.6	28.1	32
<i>#34 My students find grammatical terminology useful.</i>	9.4	12.5	59.4	18.8	32
<i>#35 Teachers find it difficult to correct student errors of grammar within a written communicative context.</i>	37.5	31.3	15.6	15.6	32
<i>#36 Teachers find it difficult to correct student errors of grammar within a spoken communicative context.</i>	9.7	35.5	41.9	12.9	31
<i>#37 My students find it difficult to improve the accuracy of their grammatical language within a totally communicative writing activity.</i>	6.3	37.5	37.5	18.8	32
<i>#38 My students find it difficult to improve the accuracy of their grammatical language within a totally communicative speaking activity.</i>	0	37.5	40.6	21.9	32
<i>#39 My students find it difficult to use grammatical terminology</i>	3.1	46.9	37.5	12.5	32
<i>#40 My students are frustrated by problem-solving techniques for learning grammar.</i>	6.5	54.8	29.0	9.7	31

## Appendix 2: Responses to online questionnaire.





### Appendix 3: Example of interview questions

Dear

Many thanks for your responses to the first set of interview questions. Before we come on to the second topic (*practice* – below) there are one or two points we'd be grateful if you would expand upon based on your earlier comments:

*Apart from your initial 'heads down' activity, your comments suggest that you tend to adopt (as Ellis and others call it) an approach based on 'incidental focus on form' by dealing with grammar points as they arise. Can you think of a recent example where you did this? For example: what was the grammar 'problem'? Why did you think it should be treated? How did you deal with it? How much time did you spend on the treatment?* For the purposes of these interview questions, we would be glad if you could again focus on the same EAP class you reflected on in the first set of responses

The following statements were in the questionnaire. Of course, we do not know how you personally responded to these questions, so we would be grateful if you could, firstly, let us know the extent to which you agree with the statements by underlining one of the following responses (or by deleting the three inappropriate options):

5. Students can improve their grammatical accuracy through frequent practice of structures.

Strongly agree	Agree	Disagree	Strongly Disagree
----------------	-------	----------	-------------------

12. Productive practice of structures is a necessary part of the learning process.

Strongly agree	Agree	Disagree	Strongly Disagree
----------------	-------	----------	-------------------

17. Comparison and contrast of individual structures is helpful for students learning grammar.

Strongly agree	Agree	Disagree	Strongly Disagree
----------------	-------	----------	-------------------

Would you now expand on these in any manner that you feel appropriate, for example:

- What methods do you employ in the classroom to ensure that students practice current structures?
- Is practice an essential part of your classroom sessions, or do you leave the practice of the target structures to homework, or cohort learning.
- If you can agree that practice is necessary and productive, should it be equally oral and written, or should there be a preponderance of one over the other.

Please feel free to add any further comments about your attitudes towards practice.

## **L1 LITERACY AMONGST GENERATION 1b: A STUDY OF AN AUSTRALIAN AFRIKAANS SPEAKING COMMUNITY**

**Henriette Janse van Rensburg, Aniko Hatoss & Donna Starks**

*University of Southern Queensland, University of Southern Queensland  
and University of Auckland*

### **Abstract**

*Numerous studies have reported on the loss of L1 literacy skills amongst second and third generation immigrants but few have investigated the maintenance of L1 literacy skills amongst Generation 1b, individuals who were born overseas but moved to their new community in the early years of life. This study considers the literacy skills of Generation 1b South African immigrants to Australia, attitudes towards their L1 literacy maintenance, and language use in electronic media. Four findings emerge from the study: there is a general decline in L1 proficiency amongst Generation 1b; this decline is most noticeable in the areas of L1 literacy, family attitudes have relatively little importance in the decline in proficiency levels and L1 literacy is related to language use in electronic media. The findings also point to attitudinal differences towards L1 literacy by mothers and fathers and suggest electronic communication as a useful means of promoting literacy amongst Generation 1b.*

### **Introduction**

Language shift typically refers to the structural or functional reduction in the use of the immigrant language amongst second or third generation immigrant communities (Clyne, 2003; Fishman, 1991). Although less commonly discussed, inter-generational shift can also occur amongst children who were born overseas but who immigrated at an early age. These Generation 1b (Verivaki, 1991; Clyne, 1976) form a vital part of new migrant communities, and are the focus of the present paper.

Literacy skills in a second language are often viewed as secondary to oral/aural skills and as a consequence of this, are typically subject to greater degrees of attrition (see Roberts, 1991 for the NZ Cantonese community; Verivaki, 1991 for the New Zealand [NZ] Greek community). The literature on language maintenance and shift is focussed on oral/aural skills as this is often a reflection of the focus within communities themselves where literacy skills are considered to have very low priority and in some communities viewed as 'not connected to language maintenance in the minds of most of the respondents' (Roberts, 1991, p. 55). Although there is

substantial documentation on inter-generational shift in oral/aural skills, there are far fewer works on loss of L1 literacy.

Language attitudes are often cited as a factor for language shift, yet their effects on language maintenance are less than clear-cut. While negative attitudes impede language maintenance, positive attitudes do not necessarily have a positive effect, partly because attitudes are multiplex and in many cases, there may not be a direct relationship between stated language attitudes and observed language behaviour (Romaine, 1995, 319). This study considers the language attitudes, literacy skills as well as language choices in electronic communication in the first generation Afrikaans speaking community.

## **Context of the Study**

Changes in linguistic social and political climate have led many South Africans to associate greater value and prestige to English (de Klerk, 2001; Rudrick, 2008 etc) with subsequent detrimental effects on South African's other languages (de Klerk, 2001; de Klerk & Barkhuizen, 2002, 2001; Finlayson et al., 1998; Kamwangamalu, 2004, 2002; Louw 2004a, b; Rudrick, 2008; Sonntag, 2003, etc). The social and political changes in South Africa have also resulted in the recent increased emigration of South Africans, with many choosing to immigrate to English speaking countries such as Britain, Canada, USA, Australia and New Zealand. Within the antipodean context, South Africans are one of the fastest growing ethnic groups within New Zealand (see Barkhuizen, 2006, 2005) and Australia. In the latter context they are the fifth fastest-growing ethnic group, where they represent 0.5% of the total Australian population (ABS, 2007). The South African population is located throughout Australia, with large settlements in Western Australia and Queensland. The present study is conducted in Toowoomba, a small regional city in Queensland. The majority of its residents are monolingual with 91% claiming to be speakers of English only. Of those born overseas, most originate from England and New Zealand, communities where English is the first language. Toowoomba's two largest multilingual communities are the South African and the Sudanese. Both communities are relatively small with local populations numbering fewer than 500, and with most recent migration in the past 10 years. Although the majority of Sudanese continue to speak their community languages, fewer than half of the local South African population report using Afrikaans as a home language. Because the majority of South Africans in the Toowoomba region come from families where Afrikaans was their mother tongue, the relatively small number of Afrikaans speakers is a likely sign of incipient language shift.

## **Methodology**

The study considered language proficiency, language use and language attitudes of 54 Afrikaans speaking South Africans from 17 families. All participants immigrated

as families, migrated to Australia in the past nine years, and to live in the local area. All the selected families have children older than 10 years of age. This was to ensure that both children and parents were born outside Australia. 21 participants made contact with the research team, and of these 17 eligible families participated in the study. Volunteering participants gave their written consent in accordance with research ethics regulations and interview and questionnaire data were collected through an Afrikaans speaking immigrant with strong ties in the local community. For simplicity purposes, the questionnaire data was collected in English. All participants had some level of English as many of the younger participants were already more competent in English than in Afrikaans. The questionnaire had eight sections and included the following topics: language proficiencies and preference, language use, social contact with peers and the home country, use of Internet and computer mediated communication for community activities and language maintenance, language attitudes, language and identity, and views about language maintenance. Data was processed by using SPSS data editor and analysed by using descriptive statistical methods.

As part of the data collection process, the parents in the families participated in the interview. The joint one hour interview elicited free speech according to the related themes: migration history; language use and proficiency; attitudes to Afrikaans and English, motivation to use and maintain Afrikaans. The interviews were allowed to develop relatively freely according to the participants' observations and experiences and were conducted in the language of the parents' choice. All were tape-recorded in Afrikaans, transcribed and translated. The data was later categorised according to the main themes that emerged and information extracted into separate files for detailed qualitative analysis.

## **Findings**

Of the seventeen families who participated in the research project, there were 15 two-parent and 2 single-parent families. The parents ranged in age from 30-52 years, the children from 10-20 years. The families had lived in Australia between 1-9 years.

Questionnaire data was elicited on the reported language abilities in each of the four skills: speaking, listening, reading and writing using a five-point Likert scale from 'not at all' (0) to 'very well' (4), summarized in Table 1 below. The findings show that these first generation family members consider themselves to be fluent in both English and Afrikaans. However, signs of incipient shift are beginning to emerge. Parents report slightly better skills in Afrikaans than in English, while the reverse is true for the children, with one exception. The children's mean scores were marginally higher for listening to Afrikaans than for listening to English. Although it is well-known that listening skills are often retained longer than speaking skills, it is rarely the case that the L1 listening skills of Generation 2 immigrants outrank those of their parents. It is likely that Generation 1b, who had weaker overall skills, made

more positive self-report judgments about their listening skills than their parents. Another striking observation from Table 1 is the reported generational differences in the areas of literacy. There is a marked decline in the reported Afrikaans reading and writing abilities of Generation 1b.

**Table 1: Afrikaans and English mean scores for parents and children**

	AFRIKAANS		ENGLISH	
	Adults	Children	Adults	Children
Listening	3.00	2.73*	2.78	2.77
Speaking	3.00	2.55*	2.66	2.73
Reading	2.97	2.05*	2.78	2.73
Writing	2.97	1.77*	2.66	2.68

\*Chi-square<.01

The reported decline in Afrikaans literacy skills is well known within the South African community in Toowoomba. The parent interviews contain commentary on the decline of their children's reading and writing skills in Afrikaans. One excerpt from an interview is presented below.

I want them to speak Afrikaans. Our children are sixteen and nineteen and we have been in Australia for just over two years and already our children have lost the ability to write in Afrikaans [Family 1]

The interviews reveal that for many Afrikaans speaking parents, Afrikaans literacy skills are considered to be of only minor importance. This is reflected in the following comments from the interviews.

I don't think it is important for them to be able to write in Afrikaans, it will just confuse them at this stage. If they later on show interest in languages, we will be more eager to teach them to write Afrikaans, but if they don't show any interest, we will not force them to read and write in Afrikaans [Family 14]

To us, it is important that they speak Afrikaans, but they have never read in Afrikaans before so I do not believe they will be able to write in Afrikaans. It is not important to us that they do [Family 10]

It is important to us that they keep the ability to speak Afrikaans, but they don't really have any need for reading and writing the language in Australia [Family 11]

I don't think it is important for them to be able to write in Afrikaans [Family 14]

Numerous reasons were provided for the lack of interest in maintaining children's L1 literacy. Some parents reported a lack of available resources. A few acknowledge that skills are important but claim such skills are difficult to maintain because of the lack of opportunities available for reading and writing in Afrikaans.

It is important to us but they only have grandparents and us for whom they try to maintain their Afrikaans abilities in writing. They try hard, but it is only a few [situations] they have to do it for [Family 3]

It is important to us, also to be able to maintain their ability to write in Afrikaans as they regularly send emails to family and friends in South Africa. They might lose some of the writing ability in the future, because they don't write Afrikaans on a daily basis anymore [Family 6]

Those parents who were interested in maintaining their children's literacy often felt that it was the responsibility of others.

I do think it is important that our children know how to write in Afrikaans. I would strongly support lessons for him. I don't have the time, so I'll rather pay somebody to teach him [Family 4]

Yes, it is very important to me that they speak Afrikaans. We would like them to be able to read and write Afrikaans as well. After school classes in Afrikaans would certainly be appreciated by many and I do believe many Australian people would also like to learn the language [Family 9]

In order to investigate the children's attitudes, and to determine if their views differed from their parents and children, we separated the responses by generation. One question proved to be particularly revealing. It considered the participant's view about the importance of reading and writing skills in Afrikaans. Differences were observed by generation and gender. Fathers reported little importance while mothers saw more of a need for literacy skills. The view is in line with other research which has considered the role of the mother as the transmitter of the oral/aural L1 skills across generations and the "gatekeeper of language maintenance" (Extra & Verhoeven, 1999, p. 20; see also Fishman, 1991; Winter & Pauwels, 2005). The fathers' attitudes towards literacy provide a possible explanation as to why many of the Afrikaans speaking families who reported an interest in literacy in Afrikaans thought that literacy should be taught outside of the home.

Generation 1b expressed little need for Afrikaans reading and writing skills, although their views were generally more positive than that of their fathers (see Table 2 for details). Children's reported proficiency proved to be unrelated to their attitudes towards Afrikaans. Of the eight children in the study who reported little or no ability in writing Afrikaans, four originated from families where the parents reported an interest in reading and writing Afrikaans and four others originated from families where the children viewed Afrikaans literacy skills as unimportant.

**Table 2: Are reading and writing in Afrikaans important?**

	RESPONSES	PERCENTAGE
FATHERS	1/13	7%
MOTHERS	8/15	53%
CHILDREN	5/22	22%

Another potential factor that emerged as contributing to the maintenance of literacy skills concerns its use within the community. Many of the parents who commented on the usefulness of literacy skills in Afrikaans made reference to electronic correspondence with the homeland. In particular email correspondence from family in South Africa is seen as important for the maintenance of contact with family in South Africa. Most of the parent interviews made reference to this type of communication.

We definitely have other activities where we make use of our mother tongue, such as computer for emails, phone calls in Afrikaans, webcam to friends as well as Skype [Family 1]

We email our family and friends in South Africa and New Zealand in Afrikaans and of course when we speak to them on the phone we also speak in Afrikaans [Family 4]

I use the computer to communicate in Afrikaans by sending emails to family and friends in South Africa. I also use the computer for my work and have lots of communication with Afrikaans speaking South Africans wanting to immigrate to Australia, so I daily write Afrikaans emails [Family 9]

My husband does not use the computer, but I send emails to family and friends as well as using the phone to speak Afrikaans. [Family 12]

My wife makes use of the computer to read Afrikaans magazines and newspapers as well as to send emails to South Africa. We also use the phone and our mobiles to Afrikaans speaking family and friends [Family 13]

When we examined the questionnaire data by generation, it became clear that all seventeen families used email as an important tool for maintaining contact with relatives and friends in South Africa. Although parents reported more frequent electronic interactions than their children, both generations revealed the regular use of email as an important tool for maintaining contact with South African family and friends (see Table 3 for details).



**Table 3: Frequency of electronic communication with friends and family**

	PARENTS		CHILDREN	
	Friends	Family	Friends	Family
Never	3	1	8	3
Monthly	12	13	10	13
Weekly	11	15	4	6
Daily	4	2	0	0

Although studies within the South African context have suggested that English is increasingly becoming the language of electronic communication (see Deumert & Masinyana's 2008 analysis of Xhosa SMS), for both Generation 1 and 1b, Afrikaans is the preferred language in email correspondence (Table 4). It is of note that fathers used Afrikaans as their only language in their email correspondence despite their views on the lack of importance of Afrikaans literacy skills reported in Table 2. The only adult in the survey who did not use Afrikaans was the trilingual mother, who reported the use of German as her language of correspondence. The language choices of Generation 1b were more mixed, with children reporting the use of Afrikaans (12), English (7) and both codes (2). Further investigation revealed that Afrikaans is the language of choice amongst those Generation 1b fluent in Afrikaans (see Table 5). Those who use English for emails appear to do so because they have 'no ability' or 'little ability' in writing Afrikaans. Thus the choice of language is dictated by their literacy skills in Afrikaans.

At present, almost all family members view Afrikaans as the language of correspondence with South African family and friends. It is typically only in cases where Afrikaans skills are inadequate that English is used for electronic communication. This finding suggests that the promotion of regular email correspondence with the homeland may provide an important mechanism for the maintenance of L1 literacy skills about Generation 1b. We noted earlier that fathers report less positive attitudes towards the maintenance of L1 literacy skills. The fathers' use of Afrikaans in email correspondence contrasts with their expressed attitudes towards L1 literacy. It is possible that the fathers' actions provide Generation 1b with positive linguistic input for L1 literacy development. The promotion of regular electronic communication with relatives and friends in Afrikaans could provide a long-term effective means of maintaining literacy skills amongst Generation 1b. The active promotion of regular email correspondence with family in South Africa would have the additional benefit of providing important social links for maintaining and strengthening South African identity amongst Generation 1b.

**Table 4: Language use in electronic correspondence**

	PARENTS	CHILDREN	TOTAL
Afrikaans	31 (96%)	13 (59%)	44
English	0	7 (31%)	7
Both	1 (3%)	2 (9%)	3
Total	32	22	54

**Table 5: Children's reported skills in writing and their language use in electronic correspondence**

	AFRIKAANS	AFRIKAANS/ENGLISH	ENGLISH
very well	4	1	
well	8		1
little	1	1	3
not at all	0		3

## Conclusion

This study reports on the maintenance of L1 literacy skills of 17 Afrikaans speaking families living in a small rural city in Queensland. Although language shift is often considered across generations, these South African families show signs of shift within first generation families, with the strongest shifts in the area of literacy. Roberts' (1991) observation that literacy maintenance is often not seen as important in language maintenance appears to be true in this context. The interviews reveal that parents value L1 oral/aural skills more than literacy skills, and this is reflected in the children's reported abilities in these areas. Studies of language maintenance in immigrant communities note that mothers tend to play an important role in L1 maintenance, and similar results are reported here. Mothers showed more interest in the maintenance of literacy skills in Afrikaans than fathers. Although most Generation 1b report that they can read and write Afrikaans, the children's language abilities varied across the families. Language attitudes about literacy skills also showed considerable variation across the families, but the two variables proved to be unrelated.

These families are relatively new immigrants, and all report the use of email for their correspondence with family members and friends in South Africa. Afrikaans is still

largely the only language for parent correspondence, but the children differ. The language use in the email is directly related to the reported language skills in reading and writing Afrikaans. Those that write 'well' or 'very well' use Afrikaans; others use English, with few reporting the use of two languages. These findings suggest that parents could help to maintain their children's literacy skills in Afrikaans through the promotion of email correspondence with relatives and friends in the homeland.

Given the technological age in which we live, it is easier to communicate across the globe enabling migrant communities to maintain stronger links with the homeland. These Afrikaans speaking parents use email to communicate with their family and friends and these actions appear to have a positive effect on children's email use and their code choice. Given the difficulties of maintaining literacy in an immigrant context, and the current reported patterns of language use in electronic communication in this community, electronic communication with the homeland should be promoted as it is likely to have positive effects on literacy in this and other migrant communities.

## Notes

We would like to express our sincerest appreciation to the Afrikaans speaking South African community in Toowoomba, the local secretary of the South African club as well as Anika Ferreira, for her assistance with this project.

## References

- Australian Bureau of Statistics. (2007). Census. QuickStats. Toowoomba (QLD) (Statistical District). Available at <http://www.censusdata.abs.gov.au/>
- Barkhuizen, G. (2005). Missing Afrikaans: 'Linguistic longing' among Afrikaans-speaking immigrants in New Zealand. *Journal of Multilingual and Multicultural Development*, 26, 216-232.
- Barkhuizen, G. (2006). Immigrant parents' perceptions of their children's language practices: Afrikaans speakers living in New Zealand. *Language Awareness*, 15, 63-79.
- Clyne, M.G. (2003). *Dynamics of language contact: English and immigrant languages*. Cambridge: Cambridge University Press.
- Clyne, M.G. (1976). *Australia talks: Pacific Linguistics D-23*. Canberra: Australian National University.
- de Klerk, V. (2001). Case study: the cross marriage language dilemma: his language or hers? *International Journal of Bilingual Education and Bilingualism*, 4, 197-216.
- de Klerk, V., & Barkhuizen, G. (2002). English in the prison services: a case of breaking the law? *World Englishes*, 21, 9-22.
- de Klerk, V., & Barkhuizen, G. (2001). Language usage and attitudes in a South African prison. *World Englishes*, 21, 9-22.
- Deumert, A., & Masinyana, O. (2008). Mobile language choices – the use of English and isiXhosa in text messages (SMS): evidence from a bilingual South African sample. *English World-Wide*, 29, 117-147.
- Extra, G., & Verhoeven, L. (1999). Immigrant minority groups and immigrant minority languages in Europe. In G.Extra & L. Verhoeven, (Eds.), *Bilingualism and migration* (pp. 3-28). Berlin: Mouton de Gruyter.

- Fishman, J. A. (1991). *Reversing language shift: theoretical and empirical foundations of assistance to threatened languages*. Clevedon: Multilingual Matters.
- Kamwangamalu, N.M. (2002). The social history of English in South Africa. *World Englishes*, 21, 1-8.
- Kamwangamalu, N.M. (2004). Language, social history and identity in post-apartheid South Africa: a case study of the "coloured" community in Wentworth. *International Journal of the Sociology of Language*, 170, 113-129.
- Louw, E. (2004a). Anglicising post-apartheid South Africa. *Journal of Multilingual and Multicultural Development*, 25, 318-332.
- Louw, E. (2004b). Political power, national identity, and language: the case of Afrikaans. *International Journal of the Sociology of Language*, 170, 43-58.
- Roberts, M. (1991). The New Zealand-born Chinese community of Wellington: aspects of language maintenance and shift. In J. Holmes & R. Harlow (Eds), *Threads in the New Zealand tapestry of languages* (pp. 31-70). Wellington: Linguistic Society of New Zealand.
- Romaine, S. (1995). *Bilingualism*. Oxford: Blackwell.
- Rudrick, S. (2008). "Coconuts" and "oreos": English-speaking Zulu people in a South African township. *World Englishes*, 27, 101-116.
- Sonntag, S.K. (2003). *The local politics of global English: case studies of linguistic globalization*. Lanham, Maryland: Lexington Books.
- Verivaki, M. (1991). Greek language maintenance and shift in the Greek community of Wellington. In J. Holmes & R. Harlow (Eds.), *Threads in the New Zealand tapestry of language* (pp. 71-116). Wellington: Linguistics Society of New Zealand.
- Winter, J., & Pauwels, A. (2005). Gender in the construction and transmission of ethnolinguistic identities and language maintenance in immigrant Australia. *Australian Journal of Linguistics*, 25, 153-168.

---

## REVIEWS

---



**Anfara, V. and Mertz, N. (Eds.). (2006). *Theoretical frameworks in qualitative research*. Thousand Oaks, CA: Sage. ISBN 1-4129-1416-7 (pbk.) 207 pp.**

*Theoretical frameworks in qualitative research*, as the title suggests, aims to demystify the role of theory in qualitative research. It is specifically addressed to post-graduate students in the social sciences, who often have trouble “finding a theoretical framework and understanding its pervasive effects on the process of conducting qualitative research” (p. xiii). I believe it will be most valuable for students in post-graduate diploma and Master’s level programmes, but also for teacher-researchers and beginning researchers.

Engaging and accessible to read, the book contains ten detailed analyses of researchers’ experiences incorporating theory in their research projects. While eight of the chapters focus on educational research, the chapters include a variety of theoretical frameworks. The editors introduce the book with an overview of the nature of theory and a literature review on the role of theory in qualitative research. They contend that it is widely held that theoretical frameworks have profound effects on research, but that the literature generally is not specific about the nature of those effects, often leading to confusion for many research students and beginning researchers.

The ten accounts of researchers’ experiences applying theoretical frameworks to their work are the highlight of the book. Each chapter is a retrospective account describing the evolving implementation of the researcher’s theoretical framework for a published research project. Using first person narrative, the authors unpretentiously address personal concerns, challenges, decisions to adopt particular theories and the development of their reasoning in the conduct of their research projects. Each chapter includes a description of the project, the academic and personal rationale for the research, the findings, and a critical discussion of the theoretical models used.

Central to the chapters are the authors’ explanations of how the theoretical frameworks affected their work. They identify several areas: determining the problem, guiding the research questions, focusing the data collection, defining the methodology, and informing the data analysis and interpretation. For example, Chapters 4 (Bettis & Mills) and 5 (Mills & Bettis) report on the same research project with different theoretical frameworks. The authors are colleagues who studied academic staff experiences of departmental restructuring. Both are members of the staff group being researched, and they shared a starting point and methodology. The studies also had a common theoretical position that focused the research. But as the research progressed, their refined, specific theoretical frameworks differently shaped the focus of the study – the literature reviews, data analysis and the outcomes. Similarly, Stubbe, Lane, Hilder, Vine, Vine, Marra, Holmes & Weatherall (2003)

illustrate the changing effects of different theoretical models on the analysis of one workplace text.

While some authors precisely identify theoretical frameworks at the onset of a study (e.g., Lugg, Chapter 10), others tell of processes of selecting theories that match their general epistemological approach, allow them to focus on the issues that drive their research interests, and also serve pragmatic needs. Kearney and Hyle (Chapter 7) for instance, chose to work with the Kubler-Ross grief model to understand how individual emotions relate to organizational change. They found that most studies of organizational change focused on areas such as leadership, institutionalization, or processes, rather than the area of their interest. What's more, because they were using a model outside their discipline, they were more confident with a well-established theory, already tested in a number of research studies. They also took care to present a "thick description" of their data independently of the framework, so that readers could link the data to their own situations.

An important theme in the book is the question of theoretical frameworks becoming overly deterministic. On one hand all the authors affirm the important ways that theory guides their research and understandings. They consider it a vital guide, one that others have used and tested, and one that helps carry them beyond single case descriptions to more powerful, broader explanations. Several draw on metaphors of theory as a "lens", a "sieve", or a "filter" to help researchers see their data in meaningful ways, a vehicle for generalizations and explanations regarding their findings. On the other hand, many recognize the possibility of a framework hiding or distorting the data so that it would fit the theory, usually unconsciously on the researcher's part. However, Fowler (Chapter 3), reporting on a study early in her career, tells of her reticence to incorporate data that conflicted with the theoretical framework of a well-established researcher. Encouraged by a reviewer who identified the discrepancies in her study and the theory, she critiqued the original model, to later learn that the author of her adopted framework was appreciative. Other authors acknowledge that a perfect fit between the data and a particular theory is not always possible or necessary. Harris (Chapter 8), for example, confirms that theories should "guide and inform" rather than "define and force" (p. 145), and that whereas case studies are bound in specific times and spaces, social life is much more dynamic and often unpredictable. Other authors write of incorporating additional theoretical models that dealt with aspects of their research not covered by their major framework (e.g., Karpiak, Chapter 6; Mutch, Chapter 9).

In conclusion, the editors (citing Garrison, 1988) point out that researchers who maintain they are "atheoretical" either "hold the theories tacitly; hold them explicitly"; or use theory covertly (p. 190). In the current climate of funding-driven pressure on teachers to produce research outputs as well as teach, and of policy-driven government and institutional sponsored research, it is more important than ever for academic researchers, as well as students, to work confidently, responsibly



and competently with theoretical frameworks. *Theoretical frameworks in qualitative research* makes an important contribution in this direction.

**Reference**

Stubbe, M., Lane, C., Hilder, J., Vine, E., Vine, B., Marra, M., Holmes, J., & Weatherall, A. (2003). Multiple discourse analyses of a workplace interaction. *Discourse Studies*, 5(3), 351-388.

**JUDY HUNTER, MASSEY UNIVERSITY ALBANY**

---

**Borg, S. (2006). *Teacher cognition and language education*. London: Continuum. ISBN 0-8264-7728-3 (hbk). 314pp.**

As Simon Borg points out in the introduction to this book, over the past fifteen years the study of language teacher cognition has become a legitimate domain of research and theory-building within the field of applied language studies, and is now of sufficient breadth and depth (he draws on over 180 studies from 1976-2006) for a book-length study of its past, present and future.

The volume has two main sections: the first five chapters examine what is currently known about language teacher cognition, while the second five explore different ways in which this topic can be studied. Part 1 opens with a comprehensive overview of the origins of the study of teacher cognition in general education and plots the steady growth and development of research interests and perspectives, beginning with Dunkin and Biddle's process-product model in the early 1970s. Borg then reviews key 1980s cognitivist studies of decision making and thinking (Clark; Yinger; Shavelson & Stern's review; Elbaz; Shulman) before moving on to the broader, richer range of views of the last twenty years which encompass teachers' professional knowledge, beliefs and principles of practice; knowledge and language teacher education; reflective practice and subject-specific teacher cognition. He identifies and comments on the emergence of teachers' professional or practical knowledge (i.e. personal/shared, non-systemic, evaluative knowledge and beliefs) as the main "umbrella" concept in the field from the 1980s onwards, and the shift in alignment of this research area from educational psychology to teacher education during the same period.

Borg warns of the danger of using the knowledge of expert teachers as the basis for offering normative advice for all teachers (as advocated by the "effective teaching" movement), and points out that as yet no research has been conducted on connections between the quality of teacher knowledge and student outcomes. He cites Fenstermacher's 1994 article as the source of two crucial questions in respect of teachers' knowledge: whether commonalities exist in the broad range of concepts described as "knowledge" and whether it is reasonable or feasible to subject this kind of (largely tacit and unanalysed) professional knowledge to examination and critique so that it might gain epistemic status. Unfortunately he does not take up the second issue to consider how evidence of teachers' knowledge might be presented so that it can lay claim to being a variant, but equally legitimate form of disciplinary knowledge.

The next two chapters in the first section of the book examine the cognitions of pre-service and in-service language teachers. Investigations of the first group, most of which has been conducted in North American contexts, have led to conflicting opinions as to the impact of pre-service education on pre-existing beliefs and knowledge, and Borg identifies the need for more wide-ranging and longitudinal studies as well as ones that distinguish between training courses of short duration and year-long programmes of study. Research into the cognitions of in-service teachers has tended to focus on how/how much novices

transfer what they have learned into their classroom practices, and on the relationship between teachers' beliefs and practices once situational and learner factors come into play. As with other research on language teacher cognition, studies in this area feature a diverse range of aims, methodologies and contexts; however Borg notes that there have been as yet very few replications or longitudinal studies, which makes it difficult to draw convincing conclusions or comparisons between studies.

Chapters on teacher cognition in grammar teaching and literacy instruction (the two areas in which most research has been conducted) round off this section of the book. With regard to the former, Borg calls for increased attention to the demands of grammar teaching, and in respect of the latter notes the scarcity of studies investigating teachers' cognitions about reading and writing. In the second half of the book, attention is focused on various ways of studying teacher cognition, with chapters on self-report instruments, verbal commentaries, observation and reflective writing. As well as outlining the frequency, advantages and disadvantages of these methodological choices, Borg evaluates and comments in detail on examples from recent research. As in the first half of the book, he identifies gaps in the field and suggests possible avenues of investigation.

The final chapter draws out eight key issues from the book for comment and review: the nature of language teacher cognition, the scope of research in this area, the relationship between teachers' cognitions and classroom practices; the impact of context on both cognitions and practices; processes of pre-service language teacher learning; the relationship between cognitive and behavioural change in language teachers; the nature of expertise in language teaching; and methodological issues in the study of language teacher cognition. The book concludes with some thoughts on how studies of teacher cognition and second language acquisition research might support and inform each other in identifying connections between teacher cognitions and student learning.

The main aim of the book is to provide a comprehensive, detailed review of what has been achieved so far in this area rather than to establish new territory; however Borg indicates a considerable number of areas of inquiry that are as yet under-researched. Strengths of the book include the clarity and authoritativeness with which the domain is surveyed, the quality and quantity of references to key studies in applied language studies and general education, and the book's clarity and attention to detail. Much-appreciated features are the inclusion of concept maps, lists of definitions to help sort out the plethora of terms used to describe beliefs and knowledge; and boxes, tables and charts to give summaries of research features – all of which enhance the accessibility of the book's contents. The hardback version of the book is expensive, but I am reliably informed that a paperback version is soon to be published. All in all, this volume is a very well-written and informative guide. It will be enormously useful to anyone wishing to learn more about or to embark on a study in this developing area of applied language studies research.

**ROSEMARY WETTE, UNIVERSITY OF AUCKLAND**

---

**Creswell, J.W. (2007). *Qualitative inquiry and research design: Choosing among five approaches* (2<sup>nd</sup> Edition). Thousand Oaks: Sage. ISBN: 1-4129-1607-0. 395pp.**

The blurb at the back of this book says it is “an excellent text for advanced undergraduate and graduate courses in introductory qualitative research methods across the social, behavioral, and health sciences”. In this review, I focus on aspects of this statement to highlight some of the strengths and limitations of the book.

The blurb is spot on with regard to the intended audience. This certainly is an introductory research text. It covers five approaches to qualitative research, and compares these, side-by-side in most chapters, in terms of theoretical underpinnings, study design, data collection and analysis, and report writing. The five approaches are: narrative research, phenomenology, grounded theory, ethnography, and case study. If a beginning qualitative researcher had never heard of these approaches before, they would come away after having read the book with a basic idea of the theory and procedures related to these five approaches. They would also have had the opportunity to read a real, published study conducted in the tradition of each of the approaches (included in a separate chapter at the end of the book). Without much research experience, however, they might be a little confused about their similarities and differences, despite the numerous tables scattered throughout the book distinguishing them. One reason is, as Creswell himself warns, the focus of a particular qualitative study is not always “as clear-cut as it appears” (p. 230). My advice would be to read the book to get an overview of all five approaches, to identify the one that applies most to the study the reader has in mind, and then read further about that approach to better understand its philosophy and procedures.

New researchers, graduate students for example, may find the contents of the book much more meaningful and useful if mediated by an instructor. By this I mean that the instructor would need to interpret with students some of the overly simplistic content of the book and to provide additional examples of studies conducted within the frameworks of the five approaches. More advanced researchers (e.g., PhD students) who have read more widely on qualitative inquiry would find the book less useful, and may even find some of the content questionable. Nevertheless, as an introductory text, it does its job well.

Quite a few sections in the book deal more generally with qualitative research instead of focusing more specifically on each of the five approaches. These provide useful overviews, and include: (a) philosophical frameworks – though this chapter is particularly thin, (b) a chapter on designing a qualitative study, (c) a Table summarizing the characteristics of good qualitative research, (d) a chapter on ways to introduce and focus a study – which includes very useful sections on how to state the research problem, write research questions, and word purpose statements, those tricky sentences which start, “The purpose of this study is to ...”, and (e) a good chapter which deals with issues of standards of validation and evaluation.

The five approaches are distinguished most usefully in the data analysis chapter, though once again, the text is quite basic and does not really get down to the practical nitty-gritty of analysis. This chapter does, however, provide helpful references for further exploration. In fact, all of the chapters have reference lists that point to fundamental texts related to each approach. These are clearly distinguished according to the approaches.

For those who know the first edition, what the second edition does new is the following: (1) It gives a broader coverage of narrative research, expanding the focus from only biography; (2) Interpretive approaches are discussed more broadly, including an emphasis on feminism, ethnicity and critical theory – though very basically; (3) More recent illustrative studies in phenomenology and ethnography are included; (4) Many more examples of studies are included.

What does the book have to offer applied linguists, including language educators? There is nothing specific in the book about our field, focusing as it does more generally on the social, behavioural and medical sciences. It does, nevertheless, cover the basics in qualitative research; content which can easily be adapted to suit the research needs of budding applied linguistics researchers. Until recently, and even now, this has in any case been our practice.

**GARY BARKHUIZEN, UNIVERSITY OF AUCKLAND**

---

**Griffiths, C. (Ed.). (2008). *Lessons from good language learners*. Cambridge: Cambridge University Press. ISBN 978-0-521-71814-1, 324pp.**

*Lessons from good language learners* is an edited collection by Carol Griffiths commemorating Joan Rubin's 1975 article on what good language learners can teach us. It is organized into two sections, the first on learner variables (e.g., motivation) and the second on learning variables (e.g., vocabulary). Each section includes both state of the art and research-based chapters. Each chapter, including the concluding summary, has sections on pedagogical implications and future research directions. Front matter includes a list of contributors, an editor's overview, a prologue, and reflections from Joan Rubin. The list of contributors shows a fairly international mix, with New Zealand strongly represented. The editor's overview presents the background, aims, organization, and uses of the book, as well as a section on terminology that frankly acknowledges the confusion created by 'a bewildering array of terms used in the field of language development' (p. 3-4), such as ESL, EFL, ESOL, and so on. This discussion, done with a light touch, probably did more to stimulate reflection on my own terminological choices than more ideologically-motivated discussions have. The light touch suits the general impression of the book as being 'user-friendly' for the primary audience of pre-service and in-service practitioners.

The volume claims other aims as well, including a review of the literature and research, which is also directed at in-service and pre-service practitioners whose needs will be served by the "wealth of information and references" (p. 5). Unfortunately this aim is unevenly addressed. A look at the reference pages of different chapters suffices: some chapters have four or five pages of references with one-third or more since 2000, whereas others have less than a page of references and fewer than one in ten since 2000. It is difficult to know what to make of a 'state of the art' chapter that relies heavily on research that is more than twenty years old (e.g., the pronunciation chapter), or of research-based chapters using extremely limited self-reported data with few references to recent theoretical and research literature (e.g., the grammar, writing, and errors chapters). These studies may be sound, but the lack of sufficient references to recent literature makes them difficult to locate within the field and the lack of methodological detail makes them difficult to judge. Even when the research studies are better designed (or better reported), they sometimes arrive at vague or obvious conclusions, such as good learners use a variety of strategies. One chapter simply surprised me: readers may not expect a chapter on speaking to focus on an electronic chat programme. It is unfortunate that many of the studies appear to be action research projects, since the practitioner audience could have been well-served by clearly focusing on this. It might have attenuated expectations and added an appealing dimension for practitioners interested in how to learn lessons from their own language learners.

It is also a shame that more of the research-based chapters did not focus on the kinds of issues that Rubin highlights in her reflections, such as which combinations of variables lead to success, how learners self-regulate, how they operationalize strategies, and how strategy training impacts performance. The case study of the effective vocabulary learner that Moir and Nation report on suggests the kinds of insights that are possible. Ehrman's chapter on personality offers an intriguing look at the personality profiles of good language learners and is based on a sound study, even though it is limited to a particular group of adults—over 3000 of them, however. The chapter on reading by Schramm provides interesting and practical pedagogical implications with a 'three level' approach.

Some of the best chapters are state of art overviews. Particularly outstanding is the chapter on motivation by Ushioda which efficiently summarizes decades of motivation research while focusing on the latest directions, including social context and self-regulation. Other effective state of the art chapters include learner beliefs, aptitude, and strategy instruction. The learner beliefs chapter by C. White offers two approaches to restructuring learner beliefs that practitioners as well as researchers will find interesting. The aptitude chapter by Ranta offers clear implications for practitioners even as it points out that we know very little about how teachers actually adapt instruction for learners with different aptitudes. The strategy instruction chapter summarizes three different approaches in a table that makes comparison easy. It also points out that most research has been descriptive, and highlights the need for more research on strategy intervention. I also applaud the summary by Oxford and Lee for not only trying to integrate the many different chapter topics, but in also trying to integrate the pedagogical implications and future research directions. Interestingly, it is motivation and self-regulation that ties pedagogy and research together, giving the reader a sense of the book reflexively spiraling back to Ushioda's opening chapter.

Overall, this is a clearly organized and mostly readable book that unfortunately does not fully deliver on its promise. A more singular vision of aims and a more consistent selection of offerings might have given this volume the kind of unity possessed by earlier collections in the Cambridge Language Teaching Library, such as *Focus on the language classroom* (Allwright & Bailey, 1991) and *Voices from the language classroom* (Bailey & Nunan, 1996). Nonetheless, there are some excellent chapters worth reading here.

## References

- Allwright, D. & Bailey, K. M. (1991). *Focus on the language classroom*. Cambridge: Cambridge University Press.
- Bailey, K. M. & Nunan, D. (1996). *Voices from the language classroom*. Cambridge: Cambridge University Press.
- Rubin, J. (1975). What the 'good language learner' can teach us. *TESOL Quarterly*, 9, p. 41-51.

**ANNE FERYOK, UNIVERSITY OF OTAGO**

---

**Mackey, A. (Ed.). (2007). *Conversational interaction in second language acquisition: A collection of empirical studies*. Oxford: Oxford University Press. ISBN 978-019-4422249-9. 496pp.**

Applied linguists working within the mainstream of Second Language Acquisition (SLA) research owe a considerable debt to Stephen Krashen. His bold claim that comprehensible input and a low affective filter were necessary and *sufficient* condition for language acquisition gave rise to a rich research agenda that is still alive and kicking. Among those who challenged his “hypotheses” can be numbered Merrill Swain, who has for twenty years developed her argument for the importance of comprehensible output, Manfred Peinemann whose Teachability Hypothesis has led to some interesting studies, and Michael Long, who over the years has refined his Interaction Hypothesis, which, in its current version,

includes elements of a hypothesis (an idea that needs to be tested about a single phenomenon), elements of a model (a description of processes or a set of processes of a phenomenon), as well as elements of a theory (a set of statements about natural phenomena that explains why these phenomena occur the way they do (Gass & Mackey, 2006, p. 174, cited in the present book, p. 5).

Certainly, Long’s Interaction Hypothesis has generated a very large number of published reports and many more unpublished theses and dissertations.

This new book edited by Alison Mackey provides a clear overview of the state of research into the role of interaction in SLA, and includes 16 new studies especially written for this book. Altogether, 27 university-based authors contributed to these studies – two in New Zealand (Rod Ellis and Rebecca Adams), two in Japan, and the rest in the USA or Canada, which were also the settings for 10 of the 16 studies, while two others were carried out in Japan, two in South Korea, and one each in Thailand and New Zealand. All of the participants were adults, and the majority university students (most often, it seems, in the same sites as the researchers).

The book begins with an extremely useful introduction by the editor, which among others things, lists and categorises over 70 studies since 1991 investigating the role of interaction in SLA. It then summarises the key points of each of the 16 studies in the collection, and then uses extracts of data from the various chapters to illustrate some of the principle tenets of the Interaction Hypothesis. The book concludes with an Epilogue by Alison Mackey and Jaemyung Goo, “Interaction research in SLA”. This is a meta-analysis and research synthesis of 26 published studies of interaction carried out between 1990 and 2006, as well as some of the new studies reported in the present volume.

These 16 chapters between the Introduction and the Epilogue are divided into three parts. The first six studies consider how interaction might be seen to promote learning



among a range of different types of learners and interlocutors, according to age, gender, motivation, proficiency, etc. Part Two contains five chapters, all of which focus on considering the effectiveness of recasts: Chapters 7 and 8 in terms of learners' working memories, the following two chapters on the perceptions of learners and their instructors about interaction and feedback, and Chapter 11 investigates how different levels of learners' awareness of feedback may be related to learning outcomes. The five chapters in Part Three all focus on directly testing whether, and how, conversational interaction can be associated with language development. Two of the studies investigate different sorts of feedback – recasts (again) and clarification requests. Rod Ellis in Chapter 14 examines the effects of explicit and implicit feedback on two grammatical structures, and Shawn Loewen and Toshiyo Nabei's study focuses on measuring the effect of feedback on L2 knowledge. Chapter 16 investigates the linguistic scope of the interaction hypothesis beyond the tendency of most studies to focus on morphosyntax.

The Epilogue addresses the effectiveness of interaction in promoting the acquisition of linguistic forms, and proceeds to identify the relationships in interaction of various factors. After a very detailed review and through discussion of the wide range of theoretical and methodological issues involved in this aspect of SLA research, the authors conclude (p. 446) that interaction does facilitate the acquisition of both lexis and grammar, and that there is scope for further research within the SLA paradigm. Indeed, their review provides many useful directions – theoretical, contextual and methodological – where future studies could break new ground.

The 16 chapters are each very clearly written and the editor has ensured that each follows a similar organisational structure, making it easy to examine and compare the various studies in detail. There is a comprehensive bibliography (rather than one at the end of each study) and a very adequate index, which also facilitates cross-referencing.

Undoubtedly, this book will be invaluable to those who are currently investigating, or intending to investigate, the role of interaction within SLA. Because it provides many clues as to where research 'gaps' exist, waiting to be filled, it can be highly recommended to students wishing to undertake doctoral or masters theses in this somewhat specific, but evidently still open, area of applied linguistics.

But for those outside the mainstream approach to SLA research, a number of questions might arise as to the terminology adopted within the field generally, and in these studies in particular. The title of this book refers to *conversational* interaction in language *acquisition*. Most of the sixteen studies – and the vast majority of others cited by the authors – actually take place in contrived environments over very limited periods of time. (Indeed, *all* the studies analysed in the final chapter were experimental or quasi experimental in design.) Are such conditions those which actually promote acquisition rather than learning? One does not have to agree with everything Krashen has argued, but his distinction between acquisition and learning,

though hardly a hypothesis, does seem to mark if not a clear boundary, then a distinct continuum, between the two fundamentally different sorts of language development. Therefore, the sort of interaction investigated by most of these case studies – even those which contain some features of genuine communication (e.g. information gap tasks in the chapter by Adams) do not correspond to what many might consider as truly conversational – i.e. open-ended interactions by people in fairly unstructured contexts.

Which brings me to my final point – that many mainstream SLA researchers continue to discount, or even ignore, the essential point that classroom learning (which a number of these studies intend to illuminate) comprises individuals who have unique backgrounds and personalities, and who are in a constant state of developing relationships with one another. This is the fundamental premise of those who view learning and acquisition from a sociocultural point of view. The 2007 international conference in Auckland, “Social and Cognitive Aspects of Second Language Learning and Teaching”, brought together many researchers from both sides – including several of the contributors to this volume – and indeed a measure of common ground was reached among the participants. It is to be hoped that there will be further meeting of minds, and empirical investigations, which will help those actually involved in exploring language teaching to more fully understand the role of interaction in language acquisition and thereby more effectively and practically assist learners to develop competence in this area.

**ROGER BARNARD, UNIVERSITY OF WAIKATO**

---

**Singleton, D. (1999). *Exploring the second language mental lexicon*. Cambridge: Cambridge University Press. ISBN: 0-521-55534-5 (pbk). 341pp.**

Like other books in the Cambridge Applied Linguistics series, *Exploring the mental lexicon* sets out to present the results of recent research that has direct relevance to language learning and teaching, and applied linguistics. This book has a particular focus on L2 lexical acquisition and processing.

Singleton is well-known for his work at Trinity College Dublin with the Modern Languages Research Project (MLRP) which has spawned a number of articles and occasional papers since 1990 informing knowledge of cross-linguistic aspects of language learning at various ages. Prior to that his published research interests were in language transfer, first language interference in second language learners and the effect of age on language learning. This book is an excellent interweaving of all these interests, focussing on the lexical component of language learning and the influences that are brought to bear on lexical choice during the learning process.

The book is divided into four unequal parts. The short introduction deals with concepts and definitions, and a historical overview. The second and largest part is a review of research into lexical development of both first and second languages in experimental psycholinguistics. The third part details the tests – rationale, methodology and results – of the Modern Languages Research Project, while the brief conclusion sums up and points to the future.

Throughout, Singleton's detailed references to and reviews of relevant research in every area before drawing conclusions (or possible conclusions), acquaints even a newcomer to this field with the leaders and milestones in the discipline. I suggest that this is not a newcomer's book, however, but rather suited to a reader already involved in tertiary level applied linguistic research or language teaching.

Singleton starts with a wide-ranging discussion of the work of Firth, Chomsky and Halliday, and the developments derived from it including the COBUILD project. He discusses the effects of collocation and the grammatical equivalent, colligation, acknowledging the oftentimes inseparable interaction of lexicon and grammar. He also touches on the links between lexis and phonology, and lexis and orthography, within the framework of L2 lexical development.

In Part 2, Singleton even includes L1 vocabulary acquisition in the context of the acquisition of literacy skills, as it adds an orthographical dimension to the mental lexicon. There is an interesting comparison of lexical development within naturalistic L2 acquisition vis-à-vis formal instruction. A further comparison is that of L1 lexical development, which is described through different ages and stages, with L2 lexical development. Chapter 3 in this section is devoted to models of lexical organisation and processing in the forms of the modularity hypothesis and connectionism.

I personally found most fascinating the effects of bilingualism and plurilingualism on the mental lexicon. The issue of code-switching and the various pieces of research and hypotheses about it is given with useful examples in several European, Asian and Middle Eastern languages, though precedence is given to English/French. Some discussion deals with cognates, false cognates and imagined cognates. It becomes clear that contextual information comes into play at the phase of lexical selection. Contextual information is able “to reduce levels of activation of lexical representations in the inappropriate language, and/or to raise levels of activation of the lexical representation in the appropriate language” (p. 127). Singleton cites an abundance of research that shows that both in L1 and L2 meaning rather than form is more challenging in lexical acquisition. Further, based on research by Hulstijn & Tangelder (1993), meaning similarity caused more interference than form similarity for both L1 and L2 speakers, and confirms that acquiring semantic characteristics of a word takes much longer than learning its formal characteristics. Take note of this, all language teachers.

In reviewing the work of the MLRP tests, Singleton writes (p 206), “As for the question why the results of what we claim is an essentially lexical test should correlate so well with other measures of general language proficiency, the answer is simply that the lexicon is at the very heart of most aspects of language use.”

**ANNETTE SACHTLEBEN, AUT UNIVERSITY**

---

## GUIDELINES FOR CONTRIBUTORS

*NZSAL* is a refereed journal that is published twice a year. It welcomes manuscripts from those actively involved in Applied Linguistics/Applied Language Studies including second and foreign language educators, researchers, teacher educators, language planners, policy makers and other language practitioners. The journal is a forum for reporting and critical discussion of language research and practice across a wide range of languages and international contexts. A broad range of research types is represented (qualitative and quantitative, established and innovative), including cross-disciplinary approaches.

### 1. Submission of Manuscripts

- 1.1 Articles should be double-spaced in A4 format with generous margins at head, foot and both sides. Pages should be numbered consecutively. Submission of a manuscript implies that it has not been published previously and that it is not under consideration for publication elsewhere.
- 1.2 Articles should normally be between 3000 and 5000 words in length.
- 1.2 Each article should include, on a separate page, an abstract of between 150 and 200 words, which is capable of standing alone as a descriptor of the article. Include the title on the abstract page.
- 1.4 A separate title page should include the following
  - the title of the article
  - author's name, and in the case of more than one author, an indication of which author will receive the correspondence
  - the affiliation of all authors
  - full postal address and telephone, e-mail and fax numbers of all authors
- 1.5 Authors who are submitting a manuscript to the journal for the first time should include a brief autobiographical sketch (50-80 words) on a separate page.
- 1.6 Copies should be submitted as an attachment to Professor John Bitchener, co-editor: [john.bitchener@aut.ac.nz](mailto:john.bitchener@aut.ac.nz)
- 1.7 All relevant articles submitted for publication will be reviewed by members of the Editorial Board or other referees.

## **2. Presentation of Manuscripts**

- 2.1 Sections of the article should be headed but not numbered.
- 2.2 All Figures and Tables should be provided in camera-ready form, suitable for reproduction (which may include reduction) and should require no change. Figures (e.g. charts and diagrams) and Tables should be numbered consecutively in the order to which they are referred. They should be included within the text. All Figures and Tables should have a number and a caption.
- 2.3 Do not use Footnotes. Endnotes should be avoided, but if essential, they should be numbered in the text by means of a superscript and grouped together at the end of the article before the References under the heading Notes.
- 2.4 References within the text should contain the name of the author, the year of publication, and, if necessary, the relevant page number(s), as in these examples:

It is stated by McCloud and Henry (1993, p. 238) that “students never ...” This, however, has not been the case (Baker & Thomas, 2001; Frank, 1996; Smithers, 1985).

- 2.5 The list of References at the end of the article should be arranged alphabetically by authors’ names. References should be given in the following form:

### **Books**

- Lillis, T. M. (2001). *Student writing: Access, regulation, desire*. London: Routledge.
- Wenger, E. (1998). *Communities of practice: Learning, meaning and identity*. Cambridge: Cambridge University Press.

### **Article in book**

- Clark, R. (1992). Principles and practice of CLA in the classroom. In N. Fairclough (Ed.), *Critical language awareness* (pp. 117-140). Harlow: Longman.

### **Journal articles**

- Lea, M. R., & Street, B. V. (1998). Student writing in higher education: An academic literacies approach. *Studies in Higher Education*, 23(2), 157-172.
- Turner, J. (2004). Language as academic purpose. *Journal of English for Academic Purposes*, 3(2), 95-109.

**Unpublished manuscript**

Park-Oh, Y.Y. (1994). *Self-regulated strategy training in second language reading*. Unpublished doctoral dissertation, University of Alabama, USA.

Stein, F. & G.R. Johnson. (2001). *Language policy at work*. Unpublished manuscript.

**Conference presentation**

King, J. & M. Maclagan. 2001, August. *Maori pronunciation over time*. Paper presented at the 14th Annual New Zealand Linguistics Society Conference, Christchurch, New Zealand

For other sources, use APA (American Psychological Association) conventions.

**3. Short reports and summaries**

NZSAL invites short reports on any aspect of theory and practice in Applied Linguistics. Manuscripts could also present preliminary research findings or focus on some aspect of a larger study. Submissions to this section should be no longer than 2000 words, and should follow the submission guidelines for full-length articles (no abstract is required, however).

**4. Reviews**

NZSAL welcomes reviews of professional books, classroom texts, and other instructional materials. Reviews should provide a descriptive and evaluative summary and a brief discussion of the work in the context of current theory and practice. Submissions should generally be no longer than 1000 words.

## ***The Applied Linguistics Association of New Zealand (ALANZ)***

**John Bitchener (President)**

**Rosemary Erlam (Secretary)**

**Janet von Randow (Treasurer)**

The Applied Linguistics Association of New Zealand (ALANZ) is the New Zealand affiliate of AILA (The International Association of Applied Linguistics). Membership of the association provides access to a network of individuals interested and active in applied linguistics research. ALANZ also provides an outlet for publications in applied linguistics through the journal *New Zealand Studies in Applied Linguistics*. Areas of research which are of interest to ALANZ members include foreign and second language education, Maori language, cross-cultural pragmatics, speech and language disorders, language policy and planning, and community languages.

Further information about ALANZ can be obtained by  
writing to the ALANZ secretary:

Rosemary Erlam

Department of Applied Language Studies & Linguistics

University of Auckland

Private Bag 92019

Auckland

New Zealand

Or by visiting the ALANZ website at

<http://www.alanz.ac.nz>