NEW ZEALAND STUDIES IN APPLIED LINGUISTICS

Volume 18 (1) 2012

Contents

ARTICLES

REVIEWS

ISSN 1173-5562

ARTICLES

New Zealand Studies in Applied Linguistics, 2012, 18 (1) pp – pp

MODIFIED OUTPUT, CLARIFICATION REQUESTS AND DEVELOPMENTAL PROGRESS OF LEARNER LANGUAGE: THE CASE OF NEGATION OF ADJECTIVES IN L2 JAPANESE

Masayoshi Ogino

University of Canterbury
masayoshi@xtra.co.nz

Abstract

This study explores the relationship between modified output produced in response to clarification requests and second language (L2) learning. Recently, studies on output have focussed on how the production of modified output (i.e., learners' reformulation of their own utterances) promotes L2 learning through describable developmental stages of grammatical structures. The target language and linguistic features of such studies, however, have been mostly limited to question formation in L2 English, and this line of research needs to be expanded to other linguistic features and beyond English to identify whether the findings can be generalized. This paper reports the results of a quasi-experimental study that applied developmental progress as a measure to L2 Japanese, targeting the negation of adjectives in the non-past tense. An analysis was made of a total of 1,011 cases of negation of adjectives, elicited by computer-administered oral production tasks from 28 learners of Japanese in a New Zealand university. Despite some limitations found in facilitating inflectional change and destabilizing an overgeneralized form, the results suggest modified output following clarification requests facilitated developmental progress, thus lending further empirical support to the claim of Swain's output hypothesis (1985).

Keywords: *modified output, clarification requests, developmental progress, Japanese as a foreign language*

Introduction

Second language (L2) learners occasionally reformulate their own utterances following corrective feedback, in the form of *modified output*. Such output has been seen as an important component of interaction with input and corrective feedback. It is also one of the main constructs in the interaction hypothesis (e.g., Long, 1983, 1996) and the output hypothesis (Swain, 1985, 1995). The central debate on the role of output in L2 learning has been whether output is necessary (Long, 1996; Swain, 1985), only facilitative (VanPatten, 2004), or unnecessary (Krashen, 1998). Accordingly, there is debate on whether modified output (MO) facilitates L2 learning. Despite numerous empirical studies on modified output, the number of interaction studies targeting L2 Japanese is still limited. Also, few

studies on modified output in L2 Japanese have employed developmental progress of learner language as a measure. The present study investigates whether the production of modified output in response to clarification requests facilitates L2 development. It seeks to extend output-focused research by employing developmental progress of interlanguage as a measure in Japanese as a foreign language (JFL).

Background

The importance of pushed and modified output in L2 learning has been investigated in interaction research (e.g., McDonough & Mackey, 2006). The theoretical framework for these studies is the output hypothesis (Swain, 1995), which claims the following three functions besides improving fluency: (1) the *noticing function* proposes that the activity of production may prompt L2 learners to consciously recognize some of their linguistic problems; (2) the *hypothesis-testing function* refers to learners' trial run for comprehensibility or linguistic well-formedness; and (3) the *metalinguistic function* claims a mediational role for the use of language to reflect on language (see Muranoi, 2007 for a summary of research on each function). Empirical evidence from previous research, however, does not yet appear to confirm that the production of modified output facilitates L2 learning (e.g., Takashima & R. Ellis, 1999). Moreover, a meta-analysis of interactional research up to 2006 (Mackey & Goo, 2007) reported the provision of opportunities for modified output did not show any difference in language development, and suggested the need for more research specifically designed to examine its effects. These results are not surprising because the production of modified output is a complex phenomenon involving a number of variables, which include data elicitation tasks, the type of interlocutor in interactions, and learner variables such as the capacity of working memory and developmental level.

Measurements of the impact of modified output

Measurement of the impact of modified output may be another contributing factor to the different results. Improvement in grammatical accuracy has been operationalized as evidence of the positive impact of modified output (e.g., Izumi, 2002; Takashima & R. Ellis, 1999). This measurement is reasonable because the output hypothesis originated from the findings of weak grammatical accuracy in input-rich French immersion programmes in Canada (Swain, 1985). However, grammatical accuracy has some limitations as a measurement because it is based on target language norms and overlooks the development of learner interlanguage (i.e., Bley-Vroman's (1983) *comparative fallacy*).

Recently, a growing body of interaction research has been carried out with a focus on how interaction involving the target feature advances learners through describable developmental stages. The analysis of interlanguage development (R.

Ellis & Barkhuizen's (2005) interlanguage analysis) attempts to describe "the underlying developmental system of the L2 learner" (Spada & Lightbown, 1993, p. 208) and avoid the comparative fallacy. In this analysis, the various linguistic devices that learners use to express a particular grammatical structure are identified at different points in the learners' development, which is followed by determining the stage of acquisition (R. Ellis, 2008). Nevertheless, only a handful of studies on modified output (e.g., McDonough, 2005) have used developmental progress as a measurement, in contrast to research on input and corrective feedback. Hence, the application of interlanguage analysis into research on modified output is of importance in order to capture possible changes in learner language, which may not be revealed by grammatical accuracy.

Targeted language and developmental progress as a measurement

Interaction-development research has largely targeted question formation in L2 English (e.g., Loewen & Nabei, 2007). One of the reasons is the availability of well-established developmental stages (Pienemann, 1998; Pienemann & Johnston, 1987). Although the choice of the same target feature among studies has its merits, there is a need to target linguistic features of languages other than English to make the findings more generalizable. L2 Japanese is a good candidate for this. There has been little interaction research in L2 Japanese that has used developmental progress as a measure of the interactional component, possibly due to the lack of studies on the developmental sequence, which did not start until the 1990s. However, research to date has demonstrated that some interlanguage structures are developmental (e.g., negation and causative constructions), and there are sufficient grounds for conducting interactional research using developmental progress as a measure (see Mori & Mori, 2011, for a review of research between 2000-2010). Also, it should be noted that although there are a number of interaction studies targeting L2 Japanese (e.g., Egi, 2010; Iwashita, 2010), few studies have specifically investigated the impact of modified output on the interlanguage of L2 Japanese, so this is an area for further investigation.

Manipulation of the production of modified output

The use of corrective feedback has been one of the options to elicit modified output, and the clarification request, a type of implicit corrective feedback, has been used (e.g., McDonough, 2005; Nobuyoshi & R. Ellis, 1993). This has two functions as feedback: it signals misapprehension or inappropriateness, and lodges a plea for reformulation (Saxton, Houston-Price, & Dawson, 2005). Explicit corrective feedback, such as recasts, provides positive evidence, which would result in adding more variables in an investigation centering on the impact of modified output. Clarification requests, on the other hand, are ambiguous in terms of the location of problems in the utterance and do not provide linguistic information that learners can incorporate into the production of modified output (i.e., no positive evidence). Thus, they are considered to be the least intervening as a trigger of output modification. Also, clarification requests have been

demonstrated to be effective in leading to modified output, compared to recasts or confirmation checks (Lyster, 2004).

Purpose of the study

Against the background above, the present study expands the application of interlanguage analysis to L2 Japanese, and explores the relationship between *modified output produced in response to clarification requests* (MO hereafter) and L2 learning. This study, which was part of a larger study into the role of modified output and corrective feedback in L2 learning (Ogino, 2009) investigates whether there is a positive relationship between the production of MO and developmental progress in the target linguistic feature. Qualitative and quantitative changes of interlanguage were used as measures to investigate the relationship, and hypotheses on two aspects of interlanguage were tested.

Hypothesis 1: There will be a difference in the changes of the use of types of non-target-like form across the tests between the experimental group and the control group.

Hypothesis 2: The experimental group will outperform the control group in terms of the use of developmentally more progressed forms.

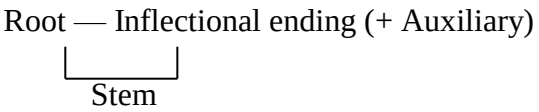
The theoretical significance of the current study lies in testing the output hypothesis, in a broad sense, with interlanguage analysis in L2 learning. This study targets a grammatical feature in JFL, and thus expands interaction-development research beyond commonly targeted linguistic features in ESL. On the pedagogical side, exploring whether the production of MO contributes to L2 learning is of particular importance in the context of error response strategies in language classrooms. It should be noted that the results are not attributed to the sole impact of modified output, but to the combined impact of clarification requests and modified output. McDonough (2005) is one of the few studies that measure the sole impact of modified output by not giving learners an opportunity to modify their erroneous output following feedback, by way of the interlocutor continuing to talk. This technique was not chosen in the present study because of the possible unnaturalness of the tasks used in the study, the target linguistic feature, and the participants' proficiency level.

Method

Targeted linguistic feature

The targeted linguistic feature is the negation of predicate regular adjectives in L2 Japanese. Adjectives in Japanese are categorised into two types on the basis of the morphological classes: regular adjectives and nominal adjectives. Regular adjectives end with a non-past tense morpheme *i*, while nominal adjectives require *na* when

they are used as a noun modifier. The current study investigates only regular adjectives and the term adjectives will be used hereafter to refer to them. Japanese adjectival phrases consist of a root and an inflectional ending which forms the adjective stem, together with optional auxiliaries (see Figure 1 below). The predicate alone can constitute a complete sentence, with the subject and other elements omitted when understood from context (Kanagy, 1991, p. 28), and, therefore, *samu-i* (cold-NONP) is a complete sentence meaning '(It) is cold'. Like verbs and auxiliary verbs, adjectives are inflected according to tense, mood, negation and politeness levels.



(adapted from Shibatani, 1990, p. 224)

Figure 1: Adverbial Phrases

Negation in Japanese is post-verbal while that in English is pre-verbal. The negative construction in Japanese involves a negative morpheme that is always bound to other morphemes, such as tense (non-past or past) and politeness level. Following previous studies on negation (Kamura, 2001; Kanagy, 1991, 2001), the term *negation pattern* is used to refer to the bound morphemes of negation and others. The negation patterns are different among the predicate categories. For example, *ku-na-i* (hereafter *kunai*) is used for negating adjectives in an informal style, while *ja-na-i* (hereafter *janai*) is used for nominal adjectives and nouns.

The construction of the negation of a predicate adjective follows three steps: (1) the identification of *samu-i* as an adjective; (2) the replacement of the *i* (non-past tense morpheme) with an adverbial inflector *ku*; and (3) the addition of a negation pattern *nai* (a negative morpheme *na* + a non-past tense marker *i*). The correct negation form is *samu-kunai*, '(It) is not cold' shown in Figure 2. Several variations have been reported in both L1 and L2 studies, including non-target-like patterns caused by overgeneralizing use of a negation pattern to other predicate categories (e.g., **samu-i-janai*) or by failing to inflect the stem (**samui-kunai*).

samu-ku + *na-i* (cold-Adv.+ NEG-NONP, '(it) is not cold')

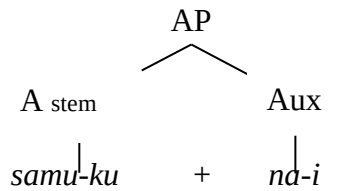


Figure 2: Tree structure of 'samu-ku + na-i'

Research has revealed that the acquisition of negation in Japanese passes through predictable developmental stages in both first language (Clancy, 1985) and L2 (Kamura, 2001; Kanagy, 1991, 2001; Kawaguchi, 2009), and thus its predictive hierarchy is available for investigation. The developmental stages of negation of

adjectives, based on findings from previous studies, are characterized by the reduction of variation of the non-target-like negation pattern in the order of **nai* > **janai* > target-like use.

Participants

Participants were recruited from among students who were enrolled in beginner and intermediate Japanese courses at a national university in New Zealand. A total of 52 participants took the pre-test, and the spoken data of 28 participants (14 each in the experimental and control groups) were analyzed for the study; 12 participants who scored 90% or higher in the pre-test and another 12 participants who did not complete all of the required sessions or withdrew half way through were excluded from the data analysis. The first languages of the participants were English, Chinese, Croatian and French. Twelve participants reported having been to Japan previously, and two of the participants had stayed in Japan for longer than two months. The average length of Japanese language learning at secondary school was one and a half years for both groups (SD = 2.14 for the experimental group and 2.07 for the control group), and that at university was 8.9 months (SD = 6.47) for the experimental group and 5.6 months (SD = 4.60) for the control group. The average length of study of the control group is shorter than that of the experimental group, and the potential difference in their general L2 proficiency could enlarge the gap between the two groups in terms of their performance on the post-tests. However, they were grouped based on the accuracy scores in the pre-test and considered to be compatible in terms of the command of the targeted structure despite the difference in the length of learning.

Data collection procedure

Design

The study employed a quasi-experimental design, including a pre-test, post-test and delayed post-test. Twenty-four participants were matched on the basis of the accuracy scores1 on the target form in the pre-test, and each member of the pair was randomly assigned to an experimental group or to a control group. The experimental sequence of the present study was over a period of two months, totalling approximately 44 hours. A post-test was conducted one week after the second treatment, and a delayed post-test was held approximately four weeks after the first post-test.

Testing instruments

Computer-administered picture description speaking activities were used for the pre-test, post-test and delayed post-test. These activities were carried out in a language laboratory equipped with an individual microphone and a computer screen for each participant. They were instructed to reply orally to 40 questions (20 questions to elicit negation of adjectives as a target form, 10 distracter questions to elicit affirmative adjectives, and another 10 questions to elicit non-adjectives). All the instructions were provided in English, but participants heard the questions in Japanese as pictures were presented on a computer screen. Participants could go to the next question by clicking the mouse, but the computer automatically presented the next question and

picture after one minute. The activities consisted of vocabulary and structures that had been taught in class prior to the pre-test.

Instructional treatments

Two instructional treatment sessions were conducted, the first treatment session being held two weeks after the pre-test, and the second treatment session one week after the first treatment session. Two different elicitation instruments, an interview (20 minutes) as a meaning-focused activity and mechanical drills (15 minutes) as a form-focused activity, were employed for the purpose of maximizing the frequency of production of modified output. Mechanical drills are a feature of the audio-lingual method with its theoretical foundations in behaviorism and structural linguistics. A mechanical drill was chosen based on the pilot study results, where participants used antonyms of the adjective instead of using negation. Although their use does not fit into the interactionist view of communicative interaction, the opportunity to negotiate meaning while performing a drill is not necessarily precluded. It at least provides an opportunity for interaction between the participant and the interlocutor to occur, in which participants are required to produce output, receive feedback, consider the intent of the clarification requests, reflect on their own utterance, and respond to the feedback in some way to complete the task. Whatever its limitations as a language teaching method (it has been pointed out that this kind of exercise limits learner output in terms of length and complexity (e.g., R. Ellis, 2005) and does not lead to the output necessary for L2 learning), it met the data collection needs of this study.

When participants in the experimental group used non-target-like negation patterns, they received clarification requests by means of a standard formula – for example, *Sumimasen* (Excuse me?), or *E* (Pardon?) – in a way that did not provide further linguistic information about the target form to the participant (see Nobuyoshi & R. Ellis, 1993, about the methodologically-focused communication task). The control group did not receive such feedback.

Analysis

Each response by the participants in the audio-recorded tests and treatment sessions was transcribed and individually coded by the researcher and a research assistant. The stages predicted within the Processability Theory (Pienemann, 1998) are not used for analysis of the test. One of the reasons is that the Processability Theory does not predict the relative difficulty among the combination of adjectives and non-target-like negation patterns (i.e., which pattern, *janai* or *kunai*, emerges or disappears first). This is because their processing procedures fall within the same stage (i.e., phrasal procedure). Another reason is that 'emergence' as a criterion as evidence of reaching a higher stage is not applicable to the negation of adjectives, as JFL learners are found to use target-like forms of negation of adjectives as well as a variety of non-target-like forms from the onset, suggesting that learners are not constrained by processing procedures. Hence, the change of negation patterns within a stage in which participants had already demonstrated emergence was analyzed.

Each non-target-like negation pattern identified was first classified into five

categories (i.e., *nai, masen, janai, kunai* and *others*). *Non-target-like* is operationalized as the incorrect choice of a linguistic form from among the range of negation patterns that exist in Japanese, while *target-like* refers to the correct choice of a Japanese negator and an associated morpheme, and the correct inflectional change on any given adjective stem. Formal and informal forms (e.g., *samu-ku arimasen* versus *samu-kunai* for 'it is not cold') were coded into the same category. The inflected and uninflected stems were not differentiated except for *kunai,* which is the correct negation pattern, but classified into the same category because of the small sample size in the current study.

Results

Oral production data in the treatment sessions

The oral production of participants in the treatment sessions was coded according to the number of attempted uses of negation, target-like patterns and non-target-like patterns. Descriptive statistics are provided in Table 1. It also displays the number of clarification requests given to non-target-like patterns in the experimental group.

Table 1: Oral production data from the treatment sessions

	Experimental group (n = 14)				Control group (n = 14)			
	Total	M	SD	Mdn	Total	M	SD	Mdn
Attempted use	220	15.7	4.86	15.0	246	17.6	4.85	19.0
Target-like pattern	139	9.9	5.90	10.5	131	9.4	5.26	10.5
Non-target-like pattern	81	5.8	5.78	4.5	115	8.2	4.81	8.0
Clarification requests	70	5.0	4.79	3.5	n/a	n/a	n/a	n/a
Modified output	29	2.1	2.23	1.5	n/a	n/a	n/a	n/a

The treatment activities elicited a total of 220 negations of adjectives in the experimental group and 246 in the control group. The distribution of scores was normal and the mean was used as the measure of central tendency. Alpha was set at .05 for all statistical tests in the analysis of treatment data and testing data. A *t*-test was carried out to compare the mean of attempted use of negation for each group, and the results[2] indicated that there were no significant differences between the groups, t (26) = 1.013, p = .321. Clarification requests were given to 70 out of a total of 81 non-target-like patterns produced by participants in the experimental group, and 41.6% of clarification requests triggered the production of MO. Even though clarification requests did not provide participants with any linguistic information, 62.1% of MO was target-like.

Testing data

A total of 1,011 negations were elicited in the pre-test, post-test and delayed post-test. Descriptive statistics are provided in Table 2. The distribution of scores was not normal and the median was used as the measure of central tendency. A non-

parametric Mann-Whitney test was carried out to confirm the pre-treatment equivalence of the two groups, and the results indicated the difference in the median accuracy scores was not significant (Z = -.407, p = .701) at the point of pre-test. Also, further Mann-Whitney tests revealed that the differences in the number of attempted uses in each session between the groups were not significant (Z = -1.684, p = .093; post-test, Z = -.531, p = .595; delayed post-test, Z = -1.346, p = .178).

Table 2: Numbers of negations produced in the tests

	Experimental Group (n = 14)					Control Group (n = 14)				
	Total	M	SD	Mdn	Rang	Total	M	SD	Mdn	Range
Pre-test	99	7.1	4.12	7.5	1-13	148	10.6	5.73	11.0	2-20
Post-test	187	13.4	3.89	14.0	5-18	190	13.6	5.35	15.5	4-20
Delayed post-test	186	13.3	3.71	13.5	5-19	201	14.4	5.69	17.0	4-20

Figures 3 and 4 graphically present the distribution of each type of non-target-like pattern across the tests by group. The results are reported only using descriptive statistics because of the small sample size.

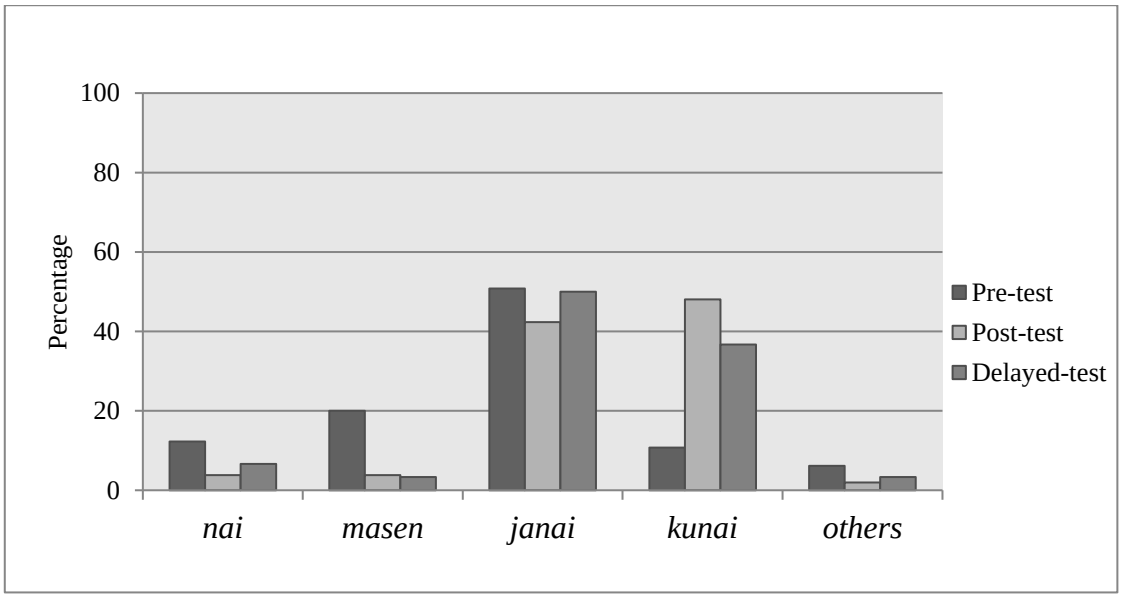


Figure 3: Percentage of each type of negation pattern in the experimental group

In the experimental group, there was a marked change in the distribution of *kunai*, which increased from 10.8% in the pre-test to 48.1% in the post-test and outnumbered *janai*. The increase was caused mostly by the decrease in *nai*, *masen* and *janai*. The distribution of *kunai* remained relatively high (36.7%) in the delayed post-test. *Janai* showed the largest distribution in the pre-test and delayed post-test. Both *nai* and *masen* decreased in the post-tests although the distributions of each type were small. In the control group, there was little change in the percentage of each type of pattern across the three tests. The use of *janai* was the most frequent throughout the tests (72.6% in the pre-test, 85.7% in the post-test and 79.7% in the delayed post-test). *Nai* was the second largest (12.6%), following

janai in the pre-test, but it almost disappeared in the post-tests. The use of *kunai* only increased slightly, but the distribution itself was small, from 5.3% in the pre-test to 9.1% in the post-test and 9.5% in the delayed post-test.

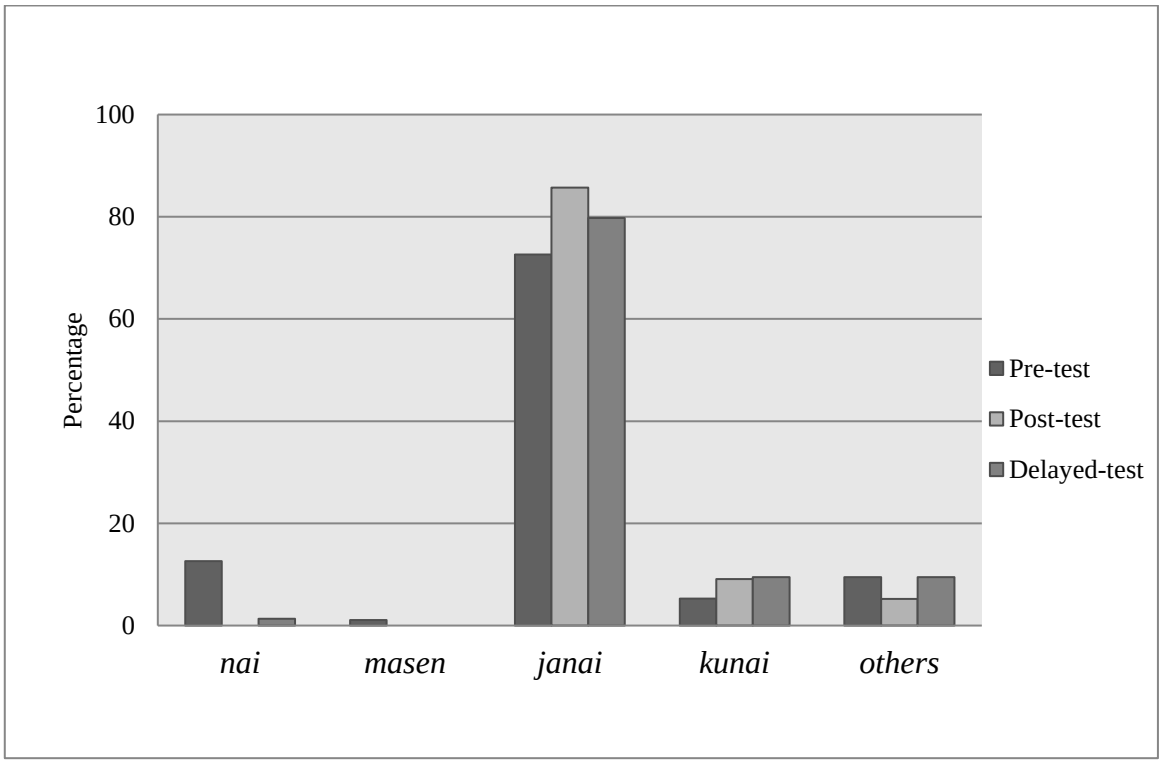


Figure 4: Percentage of each type of negation pattern in the control group

There were some noticeable differences in the types of pattern in interlanguage development reflected by the frequency of non-target-like patterns between the experimental group and the control group, which supports the first hypothesis. The difference was highlighted by the use of *kunai* after the instructional treatments in the experimental group. *Kunai* is considered to be a developmentally more progressed form, and the second hypothesis was also supported.

Discussion

The research question concerned the impact of MO on L2 learning, which was measured by the changes in developmental progress. Both of the hypotheses were supported, as there was a difference in the changes of the use of types of non-target-like form across the tests between the experimental group and the control group (Hypothesis 1), and the experimental group outperformed the control group in terms of the use of developmentally more progressed forms (Hypothesis 2). Hence, the results of the present study showed a positive relationship between the production of MO and developmental progress, thus extending partial support for the output hypothesis in JFL. It should be remembered, however, that these results are not based on the sole impact of modified output but the combined impact of clarification requests and modified output.

It should be pointed out that the experimental group received only a small number of clarification requests (median = 3.5, range = 0-16), and produced a smaller number of instances of modified output (median = 1.5, range = 0-8). Despite the limited nature of the treatment, which could be insufficient for demonstrating a difference between the experimental group and the control group, some noticeable differences were observed in interlanguage development between them. These findings support the notion that modified output could be associated with L2 learning even though its frequency is low (Shehadeh, 2002). A greater frequency of MO than that in the present study may demonstrate a stronger impact on learner language; however, this remains inconclusive and requires further studies that manipulate the frequency of production of modified output in finer detail.

The results also suggest two possible limitations of the impact of MO in facilitating the developmental progress of learner language. Firstly, MO might be limited in facilitating inflectional changes in interlanguage. The considerable increase in *kunai* in the experimental group indicates that participants chose the target-like negation pattern *kunai*, rather than non-target-like negation patterns such as *nai* or *masen*. This change suggests the production of MO facilitated the developmental progress of learner language. However, MO did not necessarily lead to the production of correct usage that requires the inflection of *i* (non-past morpheme) to *ku* (adverbial inflector) before adding a negative morpheme *nai* to the adjective stem.

The non-target-like form (e.g., **samui-kunai*) and the target-like form (*samu-kunai*) differ in the information processing operations within the stages proposed in Processability Theory (Pienemann, 1998). The choice of the correct negation pattern, *kunai,* does not involve information processing at the phrase level (i.e., a category procedure) and the procedural skills required for choosing *kunai* do not differ from choosing other non-target-like negation patterns (e.g., *janai* or *masen*). On the other hand, the inflectional changes for producing the target-like form require information processing in the phrasal level (i.e., a phrasal procedure), and therefore they are considered to be developmentally more advanced.

This limitation of MO suggests speculation that the effectiveness of MO might lie in "noticing", which has been claimed to be important in L2 learning (e.g., Skehan, 1998; Swain, 1995). Clarification requests that triggered modified output in the present study did not provide the source of the problem or linguistic information. Nevertheless, if learners already have a mental presentation of the targeted form in their interlanguage system (Egi, 2010), it is still possible that they played a major role in the change toward the correct choice of negation pattern. The implicitness of clarification requests could force learners to reflect on their own utterances and to "engage in deeper cognitive processing" (Loewen & Reinders, 2011, p. 24), thus leading to their noticing a problem (e.g., meaning, form or pronunciation). However, simply "noticing the gap" may be insufficient to achieve the inflectional change, and it may require more MO or other triggers such as more input on the target form, or provision of positive evidence by way of

explicit feedback. This supports the importance of the process of the production of modified output (Mackey, 2007; Swain, 2005), and validates the significance and necessity of research on the internal cognitive processes between the provision of clarification requests and the production of modified output, which has been seen in recent research (e.g., Egi, 2010; Gass & Lewis, 2007).

Another possible limitation of MO is that it may be ineffective in reducing the frequency of an overgeneralized form, such as the non-target-like negation pattern, *janai*. It is a negation pattern used for negating both nouns and nominal adjectives, but learners seem to perceive it as a general negation marker irrespective of the predicate category (i.e., overgeneralization) even though they have explicit grammatical knowledge about it (Kamura & Sakoda, 2001). The results show that *janai* was predominantly used across the tests in both groups, consistent with the findings that *janai* was the most persistent non-target-like pattern (Kamura, 2001; Kanagy, 1991). This suggests that MO did not facilitate the shift towards the reduction of this overgeneralized form, although it facilitated the shift towards the choice of correct negation pattern.

One of the possible explanations for this is the lack of positive evidence in the production of MO. Recasts are reported to be ineffective in changing stabilized interlanguage beyond a certain point (Nicholas, Lightbown, & Spada, 2001). One of the major differences between recasts and clarification requests in MO is that recasts provide positive evidence while clarification requests do not. Clarification requests, which could implicitly prompt noticing, may be more effective at facilitating interlanguage development (Ammar, 2008) and destabilizing certain forms (Lyster & Mori, 2006), but might be insufficient to destabilize the use of the overgeneralized form. This raises another question about what could trigger the destabilization of such interlanguage, and future research needs to investigate the relationship between such overgeneralized forms and the impact of modified output with positive or negative evidence. Another explanation is related to the amount of input of the target feature (frequency effects in language processing, N. Ellis, 2002). Participants might have been exposed to more input of *janai,* which might overpower the impact of MO. There is no research available on the ratio of negation patterns by parts of speech JFL learners are typically exposed to, but the data obtained from a corpus of interviews of L1 Japanese speakers show that the use of noun/nominal adjective + *janai* was 28%, following verb + *nai* (67%), while there was only 5% of adjective + *kunai* (Moriyama & Naidan, 2009). Although the ratio of negation patterns JFL learners is typically exposed to could be very different, if "noticing" plays a major role, it could trigger the shift to a form that learners are not so exposed to and which, therefore, may be more marked than other negation patterns. On the other hand, it may be difficult to stop learners from using *janai* for negating adjectives simply by noticing the gap through the process of production of MO, because it is the form that learners are more exposed to in comparison with *kunai*, and the overgeneralized use of *janai* to

negate adjectives is a product based on natural logic from the learners' viewpoint. In short, MO might be insufficient to overcome the effect of input. However, this remains unclear and awaits future research.

Limitations and future research

There are a number of limitations in the present study, which include the grouping of participants, the small sample size and the limited nature of the treatment. Another limitation is the design, which did not differentiate the impact of modified output and clarification requests, so that it is unclear whether the change was due to clarification requests, modified output or the combination of both. Although modified output is not produced without a trigger (typically corrective feedback), measuring the impact of clarification requests and modified output respectively without unnatural intervention continues to be a methodological challenge.

One of the possible directions for future research is to seek to target other structures shown to be developmental in nature for a range of languages. The present study used the type of linguistic device based on previous empirical research instead of developmental stages within the conceptual framework of the Processability Theory. It applied interlanguage analysis to interaction-development studies, which is useful for future research in that it investigates the developmental changes of a grammatical feature within a single stage in which learners had already demonstrated emergence. Another direction is to target structures whose developmental stages are well predicted within the Processability Theory, which will make it possible to share the same framework with previous studies in terms of developmental sequence based on processing procedures. The combination of these two directions in future research could enhance our understanding of L2 learning. Furthermore, limitations of MO need to be explored, as well as what component in interaction can facilitate inflectional change and destabilize the overgeneralized form.

Conclusion

Kanagy (2001) presented a question, "What exactly 'triggers' a change in learner interlanguage, causing them to shift from one way of expressing an L2 form, to another way?" (p. 85). Notwithstanding its limitations, the present study suggests that the production of modified output in response to clarification requests can be one of the triggers. Swain's output hypothesis was originally framed in terms of the impact of output on grammatical accuracy, but the findings suggest that MO may facilitate interlanguage development towards target-like use in JFL, thus lending partial empirical support to the claim. It was also suggested that MO could be ineffective in facilitating inflectional change and destabilizing the use of overgeneralized forms, and that the main function of MO lies in "noticing". In light of this outcome, the present study validates the significance and necessity of research that investigates the relationship between modified output, internal cognitive processes and L2 learning.

Acknowledgements

I would like to express my special thanks to Anne Feryok and two anonymous reviewers as well as to Kon Kuiper, Jenefer Philp, Ian Bruce and Roger Barnard. An earlier and shorter version of the research reported in this paper was published in the Proceedings of the 2011 Spring Conference for the Society for Teaching Japanese as a Foreign Language (written in Japanese).

Notes

1. Interactional research employing interlanguage analysis often assigns participants to groups based on the targeted developmental stage (McDonough, 2005), but the present study, as part of a study with a larger scope, chose accuracy for the purpose of comparison with studies that have used accuracy as a measurement for the impact of MO (Nobuyoshi & R. Ellis, 1993; Takashima & R. Ellis, 1999). Nevertheless, the distribution of negation patterns in the pre-test showed some similarity, except *masen* and *janai,* between groups.

2. Alpha was set at .05 for all statistical tests in the analysis of treatment data and testing data. All the statistical tests were performed using the *Statistical Package for the Social Sciences* (SPSS) version 12.1.

References

Ammar, A. (2008). Prompts and recasts: Differential effects on second language morphosyntax. *Language Teaching Research, 12*(2), 183–210.

Bley-Vroman, R. (1983). The comparative fallacy in interlanguage studies: The case of systematicity. *Language Learning, 33*, 1-17.

Clancy, P. (1985). The acquisition of Japanese. In D. Slobin (Ed.), *The crosslinguistic study of language acquisition, Vol 1, The data* (pp. 373-524). Hillsdale, NJ: Erlbaum.

Egi, T. (2010). Uptake, modified output, learner perceptions of recasts: Learner responses as language awareness. *Modern Language Journal, 94*, 1-21.

Ellis, N. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition, 24*, 143-188.

Ellis, R. (2005). Principles of instructed language learning. *System, 33*, 209-224.

Ellis, R. (2008). *The study of second language acquisition* (2nd ed.). Oxford, UK: Oxford University Press.

Ellis, R., & Barkhuizen, G. (2005). *Analysing learner language*. Oxford, UK: Oxford University Press.

Gass, S. M., & Lewis, K. (2007). Perceptions about interactional feedback: Differences between heritage language learners and non-heritage language learners. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 79-99). Oxford, UK: Oxford University Press.

Iwashita, N. (2010). The effect of intensive recast treatments on the long-term development of less salient structures in Japanese as a foreign language. *Journal of Japanese Language Teaching, 146*, 18-33.

Izumi, S. (2002). Output, input enhancement, and the noticing hypothesis: An experimental study on ESL relativization. *Studies in Second Language Acquisition, 24*, 541-577.

Kamura, N. (2001). Nihongo no hiteikei no shuutoku: Chuugoku bogowasha ni taisuru juudantekina hatsuwa chousa ni motozuite [Developmental sequences of negation in Japanese by adult Chinese-speaking learners]. *Acquisition of Japanese as a Second Language, 4*, 63-80.

Kamura, N., & Sakoda, K. (2001). Gakushuusha no goyoo o umidasu gengo shori no sutoratejii (2) - Hiteikei 'jyanai' no baai [A language processing strategy that produces learners' errors (2) - The case of the negative form 'JYANAI'-]. *Bulletin of the Department of Teaching Japanese as a Second Language, Hiroshima University, 11*, 43-48.

Kanagy, R. (1991). *Developmental sequences in the acquisition of Japanese as a foreign language: The case of negation.* Doctoral dissertation, University of Pennsylvania.

Kanagy, R. (2001). Developmental sequences, second language acquisition and Japanese language pedagogy. In H. Nara (Ed.), *Advances in Japanese language pedagogy* (pp. 47-92). Columbus, OH: National East Asian Languages Resource Centre.

Kawaguchi, S. (2009). Acquiring causative constructions in Japanese as a second language. *Japanese Studies, 29*, 273-291.

Krashen, S. (1998). Comprehensible output? *System, 26*, 175-182.

Loewen, S., & Nabei, T. (2007). Measuring the effects of oral corrective feedback on L2 knowledge. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 361-377). Oxford, UK: Oxford University Press.

Loewen, S., & Reinders, H. (2011). *Key concepts in second language acquisition*. New York, NY: Palgrave Macmillan.

Long, M. (1983). Native speaker/non-native speaker conversation in the second language classroom. In M. Clarke & J. Handscombe (Eds.), *On TESOL '82* (pp. 207-225). Washington, DC: TESOL.

Long, M. (1996). The role of the linguistic environment in second language acquisition. In W. Ritchie & T. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413-468). San Diego, CA: Academic Press.

Lyster, R. (2004). Differential effects of prompts and recasts in form-focused instruction. *Studies in Second Language Acquisition, 26*, 399-432.

Lyster, R., & Mori, H. (2006). Interactional feedback and instructional balance. *Studies in Second Language Acquisition, 28*, 169-300.

Mackey, A. (2007). Introduction: The role of conversational interaction in second language acquisition. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 1-26). Oxford, UK: Oxford University Press.

Mackey, A., & Goo, J. (2007). Interaction research in SLA: A meta-analysis and research synthesis. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 407-452). Oxford, UK: Oxford University Press.

McDonough, K. (2005). Identifying the impact of negative feedback and learner's responses on ESL question development. *Studies in Second Language Acquisition, 27*, 79-103.

McDonough, K., & Mackey, A. (2006). Responses to recasts: Repetitions, primed production and linguistic development. *Language Learning, 56*, 693-720.

Mori, Y., & Mori, J. (2011). A language in focus: Review of recent research (2000-2010) on learning and instruction with specific reference to L2 Japanese. *Language Teaching, 44*, 447-484.

Moriyama, S., & Naidan, B. (2009). Relations between acquisition and input frequency of Japanese negative forms in verbs and adjectives. *Proceedings of Ninth Annual Meeting of the Japanese Cognitive Linguistics Association, 9*, 480-487.

Muranoi, H. (2007). Output practice in the L2 classroom. In R. DeKeyser (Ed.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology*. New York, NY: Cambridge University Press.

Nicholas, H., Lightbown, P., & Spada, N. (2001). Recasts as feedback to language learners. *Language Learning, 51*, 719-758.

Nobuyoshi, J., & Ellis, R. (1993). Focused communication tasks and second language acquisition. *ELT Journal, 47*, 203-210.

Ogino, M. (2009). *Modified output in response to clarification requests and second language learning.* Doctoral dissertation, University of Waikato, Hamilton, New Zealand.

Pienemann, M. (1998). *Language processing and second language development: Processability theory*. Amsterdam: John Benjamins.

Pienemann, M., & Johnston, M. (1987). Factors influencing the development of language proficiency. In D. Nunan (Ed.), *Applying second language acquisition research* (pp. 45-141). Adelaide, Australia: National Curriculum Resource Centre, Adult Migrant Education Program.

Saxton, M., Houston-Price, C., & Dawson, N. (2005). The prompt hypothesis: Clarification requests as corrective input for grammatical errors. *Applied Psycholinguistics, 26*, 393-414.

Shehadeh, A. (2002). Comprehensible output, from occurrence to acquisition: An agenda for acquisitional research. *Language Learning, 52*, 597-647.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford, UK: Oxford University.

Spada, N., & Lightbown, P. (1993). Instruction and the development of questions in L2 classrooms. *Studies in Second Language Acquisition, 15*, 205-224.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. M. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235-253). Rowley, MA: Newbury House.

Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics: Studies in honour of Henry G. Widdowson* (pp. 125-144). Oxford, UK: Oxford University.

Swain, M. (2005). The output hypothesis: Theory and Research. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 471-483). Mahwah, New Jersey: Lawrence Erlbaum.

Takashima, H., & Ellis, R. (1999). Output enhancement and the acquisition of the past tense. In R. Ellis (Ed.), *Learning a second language through interaction* (pp. 173-188). Amsterdam: John Benjamins.

VanPatten, B. (2004). Input and output in establishing form-meaning connections. In B. VanPatten, J. Williams, S. Rott, & M. Overstreet (Eds.), *Form-meaning connections in second language acquisition* (pp. 29-48). Mahwah, NJ: Lawrence Erlbaum.

New Zealand Studies in Applied Linguistics, 2012, 18 (1) pp – pp

JAPANESE AS A FOREIGN LANGUAGE EXTENSIVE READING AND SELF-REGULATION: CASE STUDIES OF HIGH SCHOOL LANGUAGE LEARNERS

Ryoko de Burgh-Hirabe & Anne Feryok

University of Otago
deburgh-hirabe@xtra.co.nz & anne.feryok@otago.ac.nz

Abstract

In general, studies on extensive reading (ER) have focused on its effects, overlooking the process through which students do ER. Also, much of the research on ER uses large-scale surveys that may not reveal what learners actually do when reading extensively. One learner process, self-regulation, appears to be important to ER, as several recent studies on motivation for ER have implied. However, self-regulation of behaviour and context has not yet been the focus of a study. This study uses qualitative methods to examine self-regulation of behaviour and context by nine high school learners of Japanese as a foreign language in an ER project. The findings showed that the students who met both study and ER expectations were those who could self-regulate their behaviour and context. The findings also suggested that external demands, especially the national exam, distracted students from doing ER in their own time. The implications are that reading in class would secure a certain amount of reading for all students, and this could also provide support for developing self-regulatory abilities. For some students, ER may have potential as a means for developing self-regulation.

Key words: *extensive reading, self-regulation, Japanese as a foreign language*

Introduction

Extensive reading (ER) typically involves larger amounts of reading than intensive reading through both longer texts and longer periods of time devoted to reading them (Grabe, 2009). This is facilitated by texts that are attractive, interesting, and within the linguistic competence of students (Day & Bamford, 1998; Grabe, 2009). Through ER, learners can develop reading fluency (Day & Bamford, 1998; Taguchi, Takayasu-Maass, & Gorsuch, 2004) and incidentally learn vocabulary (Nation, 2007; Waring & Takaki, 2003). For these reasons, ER is one of the typical activities in the meaning-focused input strand in second language (L2) learning (i.e., learners' focus on understanding, gaining knowledge and /or enjoyment) (Nation, 2007).

It is not surprising that in the past three decades a substantial number of studies have reported on the benefits of ER (e.g., Elley, 1991; Elley & Mangubhai, 1981; Hafiz &

Tudor, 1989; Hitosugi & Day, 2004; Leung, 2002; Mason & Krashen, 1997; Robb & Susser, 1989; Taguchi et al., 2004; Tanaka & Stapleton, 2007). In general these studies focus on the outcomes of ER, overlooking the process through which students do ER. One learner process in particular, self-regulation, appears to be important to ER, as several recent studies on motivation for ER (e.g., Leung, 2002; Tabata-Sandom & Macalister, 2009; Takase, 2004, 2009) have implied. In particular, self-regulation of behaviour and context has not yet been the focus of a study. This study focuses on the self-regulation of behaviour and context in an extensive reading project, conducted outside of class-time, with a group of high school learners of Japanese as a foreign language.

Self-regulation in L2 research and educational psychology

Self-regulation is a multidimensional construct that involves "*cognitive*, *affective*, *motivational*, and *behavioral* components that provide the individual with the capacity to adjust his or her actions and goals to achieve desired results in light of changing environmental conditions" (Zeidner, Boekaerts, & Pintrich, 2000, p. 751). According to Dörnyei (2005), self-regulation has become an important focus because of its broad perspective on the processes (as opposed to the products) of self-regulation.

This study adopts Pintrich's general framework (2000, 2004) because of its explicit categorisation of self-regulation in areas relevant to this study, in particular behaviour and context. As Pintrich points out in his discussion of post-secondary settings, "much of the learning that goes on takes place outside the college lecture hall or classroom" (p. 400). His framework is designed to include the types of self-regulation associated with such environments, making the framework particularly apt for this study that took place outside of class-time.

Pintrich's (2000, 2004) divides self-regulation into four areas, cognition, motivation and affect, behaviour, and context. Each of these is sub-divided into four phases, forethought, monitoring, controlling, and evaluating. The areas can be analysed separately, even though in practice they are intertwined (Wolters, Benzon, & Arroyo-Giner, 2011). This study focuses specifically on two areas, behaviour and context, because of their apparent relevance in the ER studies mentioned above.

Pintrich (2004) maintains that the ability to regulate behaviour, particularly effort and persistence, is crucial. Planning to monitor and manage time and effort occurs in the forethought phase. The second and third phases involve monitoring and controlling the time (e.g., by making schedules) and effort (e.g., by forming intentions) (Pintrich, 2000). In particular, research has shown that self-regulated learners and high achievers engage in time management activities (Zimmerman, 1998; Zimmerman & Martinez-Ponz, 1986). In the final phase actual behaviours are reflected on and judged.

Context is closely related to behaviour because it concerns participant behaviour while accomplishing a task in a specific environment. The forethought phase concerns task perceptions (e.g., procedures) and its environment (e.g., physical conditions). Context is monitored and controlled through strategies that shape the learning environment (e.g., removing distractions). In the final phase these are evaluated (Pintrich, 2000).

Studies on self-regulation in the L2 field

Self-regulation studies of L2 learning are still relatively new. Most rely on self-report surveys. In what appears to be the first L2 self-regulation study, Tseng, Dörnyei, and Schmitt (2006) developed an instrument to measure the underlying self-regulatory capacity that leads to strategy use, arguing that it is relevant to second language learning. Of particular relevance is Hirata's (2010) survey investigating self-regulation and motivation of 381 Japanese as a foreign language (JFL) students at New Zealand universities learning *kanji*. The results suggested the interdependence of self-regulation and motivational beliefs; self-efficacy was important because it appeared to facilitate behavioural regulation. Although the present study does not focus on *kanji* learning, similar motivational beliefs may affect the extent to which the JFL learners self-regulate in ER.

ER motivation studies

Although they do not directly address self-regulation, several studies on motivation for ER have implications for the challenge that learners face in self-regulating behaviour and context. Nishino (2007) investigated two Japanese secondary school learners' motivation for extensive reading in English. The participants managed to control their behaviour and context while reading for two and a half years, although their motivation fluctuated because of the interaction of various influences (e.g., materials, sociocultural influences). However, the participants' ability to self-regulate for such an extended period of time may have been helped because the researcher, who acted as a tutor in the study, was the mother of one participant and the aunt of the other.

Two studies in JFL settings discuss the difficulty university JFL learners had in sustaining the effort to read extensively outside a classroom setting, where the participants had to manage the ER process for themselves. Leung (2002), the researcher-participant in her 20-week diary study, reported that she found it difficult to find "discipline, and time to read" (p. 76) due to other commitments. Tabata-Sandom and Macalister (2009) also mention that their participant found it difficult to read regularly during a three-month ER study due to external demands.

Other studies indicate that time management is challenging if ER is done outside the classroom—even when it is a course requirement. Takase's survey studies on the motivational influences of university (2009) and high school (2004) EFL students showed that students who did ER during class read far more than those who did not, with busy schedules reported as the main influence. Those lacking self-regulatory ability may need support. Robb (2002) argues that students need "a clear follow-up or

tracking mechanism to hold them responsible for their work" (p. 147). These studies imply that self-regulation of behaviour in self-directed contexts is important for the success of ER.

Research Questions

Much of the research on ER, like the self-regulation studies mentioned above, uses large-scale surveys. While surveys are relatively easy to administer to large groups, they may not reveal what learners actually do when reading extensively (e.g., Arnold, 2009; Mohd Asraf & Ahmad, 2003; Leung, 2002; Nishino, 2007; Tabata-Sandom & Macalister, 2009). The present study considers high school JFL learners' self-regulation for ER using qualitative methods. It aims to answer the research questions:

1. Did the participants read extensively?

2. Did self-regulation of behaviour and context influence their extensive reading?

Methodology

Participants and research setting

Participants familiar with the Japanese writing system were recruited at two New Zealand high schools. At the schools' request, classroom teachers made the initial solicitation for participation, and then the first researcher explained and answered questions about the project. Five senior male students at one school and four senior female students at the other school volunteered. See Appendix A for profiles of participants, (using pseudonyms). In both schools Japanese was an elective subject with four one-hour classes each week.

In New Zealand, high school students work toward the National Certificate of Educational Achievement (NCEA) Levels One, Two, and Three during their senior years. NCEA involves internal and external assessments based on standards measuring subject skills and knowledge. For each passing assessment, students receive credit, with high achievement recognized with endorsements of merit or excellence. In Japanese, students have three internal assessments and an end-of-year external exam. Each NCEA level provides the sentence structure, vocabulary and *kanji* standards that must be demonstrated in order to be awarded credit.

The ER programme

Due to classroom time constraints, ER was an out-of-class self-directed activity organized for the research project. It was initially planned for five months across the first two terms, but ran seven months for four students (Jane, Josie, Jack, Nick) who voluntarily continued in Term Three.

At the outset, the first researcher met the participants in their classroom, introduced the available materials, and demonstrated the reading record and journal. She asked

students to select an easy and interesting book and to read as much as possible, emphasising that one book weekly was desirable. Participants then chose a book.

In Terms One and Two the first researcher made on average 17 near-weekly visits to each school to provide on-going support. She asked the participants how their reading was going and gave them encouragement and advice when they returned and borrowed the ER books at the beginning of the class. In Term Three, she made four visits to each school.

Materials

The first researcher collected 80 graded readers (Levels One to Four) and 67 children's books, which were graded into four levels using the criteria in Hitosugi and Day (2004). In Term Three six newly published graded readers at the lowest level were provided.

Procedures

The present study is part of a larger qualitative case study that explored JFL learners' perceptions about ER and motivational change and influences (de Burgh-Hirabe, 2011) that included reading records, interviews, journals, questionnaires, and think-alouds. This study uses the reading records, interviews and journals.

Interview and journal data.

Three 30-minute to one-hour interviews were held with each participant at the beginnings of Terms One, Two and Three; for those who continued another was held at the end of Term Three. Interview topics included strategies, progress, motivation, and challenges. All interviews were audio recorded and transcribed by the first researcher. In addition, journal entries from each participant were collected each term between the first and the second interviews and the second and the third interviews.

For this study a deductive (Ellis & Barkhuizen, 2005) iterative analysis was performed after the inductive iterative analysis for the original study (de Burgh-Hirabe, 2011) revealed evidence of self-regulation. Codes and categories were based on Pintrich's (2004) framework of self-regulated learning. For example, the comment, "It's hard to find the time" was coded *time management* in the category *behaviour*. The second researcher checked the coding and the categories. All discrepancies were discussed and issues resolved.

Findings

Because this study focuses on behaviour and context, only these areas will be addressed in the case studies. A brief overview of other areas is provided here as a general background.

The original data analysis (de Burgh-Hirabe, 2011) showed that all of the participants were motivated to do ER because they perceived that ER was useful to achieve the

goal of improving their JFL. In most cases this goal was connected to two other goals that motivated them: going to Japan and doing well on exams. All of the participants displayed a range of cognitive self-regulation strategies for ER. Most of the participants displayed motivational and affective self-regulation strategies; where relevant these will be mentioned below.

The case studies are roughly organized from the most self-regulating to the least self-regulating. See Appendix B for exact figures on each participant's reading.

Josie

In Term One Josie set a goal of reading over 50 books before the end of Term Two, but during Term Two her goal changed to finishing Level Three when she realized that her initial goal was unrealistic because the higher-level readers took longer to read.

In Term One, Josie finished Level One and moved to Level Two. At the project start, although internal assessments and volleyball prevented Josie from reading much, she wrote that she was developing a good routine for fitting reading in around other commitments. After the journal comment, an injury forced her to pull out of sports and probably contributed to Josie having more time to read. Josie doubled her reading in Terms Two and Three. She described two relevant features of her usual time management skills: she finished her assignments a week before they were due so she could revise them, and she included time to relax each day. Prior to the project she relaxed by reading in English, and so for the project she simply used part of that time for reading in Japanese.

Josie met her revised goal. The fact that Josie explicitly revised it suggests the extent of her ability to monitor and control her goals. Although the extra time available no doubt contributed to Josie's increased reading, it was clear that Josie had developed good time management skills managing her school and sports commitments which she transferred to ER.

Tracey

Tracey's goal was to read one book per day. It was not clear if she meant she would complete a book every day or merely read some every day.

In Term One, Tracey read the Level One books and moved to Level Two. In the interview Tracey said that she was not very busy in Term One. Tracey reported continuing reading regularly (although not as much) in Term Two, when she began reading Level Three books. (Unfortunately, she lost her record sheet.) Although Tracey was busier, in her journal and interview she reported making a timetable for herself. However, she also said that she read less when aerobics practice started.

Tracey did not meet her goal if it was to complete one book every day, but she did report reading three books a week for most of the weeks in Term Two. She showed her self-regulation of behaviour through her time management with a schedule for planning and monitoring her activities, and her awareness of the impact competing activities had on her time.

Alan
Alan's goal was to read one book each week.

In Term One Alan roughly met his reading goal. His journal entry in the fourth week indicated that extensive reading was becoming part of his routine, although he pointed out that it was difficult to get used to it. He reported that he wrote himself a note and left a book out to help him remember. In Term Two Alan read more books. He wrote that he was trying to read more. In the interview he said that he could fit in ER around his other commitments, and he no longer needed to remind himself. Alan mentioned that he had been going to tutorials that taught him how to study effectively and prepare for exams across subjects.

Alan met his goal, perhaps because it was a rather limited goal. He demonstrated self-regulation by recognising the need to create a new habit, which he did by self-monitoring himself with reminders. Alan successfully applied his knowledge of and skills in using self-regulatory strategies, which were supported by his study tutorials, to extensive reading.

Emma
Emma's goal was to "try and finish Level Two".

In Term One, Emma read extensively and regularly. She found the Level One books enjoyable and easy. In Term Two, Emma read only half as much. Although she did not enjoy the Level Two books because they were "sad", she attributed the decline to extracurricular activities, saying "this term it's been really busy". She became anxious about the NCEA exams and felt she needed to devote time to them. Emma liked the idea of making extensive reading compulsory, saying in the interview that she needed pressure to read. As she pointed out when asked why she had not dropped some of her extracurricular activities, "When I dropped them then I was able to do something else that I have been missing out … so it's just the same still".

Although Emma did not meet her goal because she read only half of the Level Two books, clearly she read extensively. Her comments about making ER compulsory suggest she was aware of her struggle to self-regulate. It is highlighted by her difficulty in finding time to do ER even when she had dropped other activities to do so, and the fact that anxiety pushed her to study, which shows her regulating her behaviour largely in response to external demands.

Jane
Jane's goal was reading at least 32 books by the end of Term Two.

In Term One Jane read about one and a half books each week. Jane quickly moved to the second level readers, and even read three third level books. However, in Term Two Jane read the least, saying, "My schoolwork is the problem". Since Japanese was her strongest subject, Jane decided to study only her other subjects until late at night, saying "my free time merged in with my study time." When she did take time off, she listened to music or watched movies in Japanese. Although Jane had remained in the project until the end of Term Three, she read no ER books.

Jane did not meet her goal. Jane attributed her decreased reading to external demands, which she perceived as beyond her control. Her description of her free time and study time as "merged" suggests that she did not recognise the need to establish boundaries for her behaviour. The fact she chose to not study her strongest subject, Japanese, in an effort to improve in other subjects, suggests external demands motivated her behaviour, but also that it was a conscious choice to some extent.

Nick

Nick's goal was to read 20 books by the end of Term Two.

In Term One, Nick struggled to read regularly. Nick's low proficiency level made the Level One books very challenging. Although Nick allocated time for ER, he said "I'm too tired before I go to bed". He described how homework, English language books, and a new girl friend left him little time to read in Japanese. He pointed out that he would have done more reading if there was "punishment" for not doing the reading, and jokingly said during an informal chat that the first author would have to "taser him". However, in Term Two Nick doubled both the number of books and amount of reading time. In part this was because he had broken up with his girl friend, he had not found any English books to read, and he had adjusted to the demands of homework. However, Nick also said that structures he was learning in Japanese class were helping him understand the books and he started using the glossary that had been provided. He also started reading at school instead of before going to bed. In Term Three, although Nick spent less time reading, he again doubled the number of books. In that term Nick benefited from the newly published Level Zero graded readers, which he reported as easy to read without glossaries.

Nick did not meet his goal. However, Nick was aware of his struggle to self-regulate. Without the right task (i.e., easy and interesting books) and environment (i.e., no distractions), Nick might not have been able to read more. In Term Two he took active steps to support his reading before the easy readers became available and while distractions were lessening. These facts suggest Nick was learning how to self-regulate during ER.

Jack

Jack set a goal of reading "a minimum of 16 books (one book a week)" by the end of Term Two.

In Term One, although Jack read spent more time reading than any other participant, he did not read very many books. In the interview, he said it was easy for him to read with the other extension work he was already doing for Japanese. He did not report any distracting influences or obstacles.

However, Jack did not sustain this amount of reading in Terms Two and Three. In Term Two, Jack said that schoolwork and new friends kept him busy. In Term Three, despite saying he wanted to read, Jack was busy preparing for the upcoming exams; he spent his time on practising letter writing and revising vocabulary for the NCEA exams. As he said, "It's kind of hard to fit it all in."

Jack did not meet his goal. Jack attributed his decreased reading to external demands. Since Jack continued to read, however, it is unclear to what extent he succumbed to external demands as opposed to consciously choosing to read less and study more.

Drew

Drew's goal was reading 30 books.

For the first three weeks in Term One, Drew managed to read one book a week, but then read only one more. He wrote in his journal that he was trying to make ER a routine, but was unable to because he had been busy with cricket. Drew said in the interview that he used no particular strategies to remind himself to read. In Term Two Drew read the same amount. He said in the interview that he was busy with other commitments, but halfway through the term wrote in his journal that he planned to read more. However, he repeatedly said in the third interview that he could not make extensive reading into a routine. Drew agreed that other commitments were part of what prevented him from reading, but said that the major reason was his inability to fit reading into a routine, saying "I didn't really set time to do it." Drew welcomed the idea that extensive reading be compulsory.

Drew did not meet his goal. He attributed his lack of reading to his inability to self-regulate his behaviour. His awareness did not seem to impact his behaviour.

Ben

Ben did not set a reading goal.

Ben read several books in both Term One and Term Two. In the second interview, Ben said that he had no concerns or problems with how his reading was going, although he seemed regretful about how little he had read. He said he enjoyed it and found the books interesting. In the last interview, Ben said that he was not studying Japanese much because he had no exams coming up. When asked about other

extension activities in Japanese (e.g., watching movies), Ben said that they were not important and he would rather do other things.

Since Ben did not set a goal, he neither met nor failed to meet it. As he said in the interview, "I just read when I feel like it." He did not display awareness of or skills in using self-regulatory strategies during the project. Since the voluntary and non-credit-bearing nature of the project provided no external demands, Ben chose to do other things that he enjoyed more.

Discussion

The answer to the first research question "*Did the participants read extensively?"* is that it varied among the participants. Josie both read the most books and spent the most time doing so. Of the other participants, Tracey and Emma read more books, but Alan and Jack (whose proficiency levels were lower) spent more time reading. Some participants (Jack, Jane, Emma) read more books and spent more time earlier in the project, while two others (Alan and Nick) did so later in the project. Two participants (Drew and Ben) read so little that they cannot be said to have read extensively.

The answer to the second research question *"Did self-regulation of behaviour and context influence their extensive reading?"* is that self-regulation of behaviour and context appeared to influence the amount of extensive reading that participants did.

Several of the participants clearly demonstrated that their self-regulation through goal-setting influenced their reading. Alan and Tracey had weekly goals, and Josie displayed how she monitored and controlled her goal by revising it. Zimmerman (2011) points out, "The goal systems of self-regulated students are often organized with short term goals serving as proximal paths to more distal goals" (p. 56). In fact, Alan mentioned that meeting his goal was a reason to increase his reading in the second term. Only one other student, Jack, mentioned a weekly goal, but it was in a parenthetical comment. It may have been only incidental to his 30-book minimum because it mapped onto the number of weeks in two terms; Jane also had distal goal that mapped onto a weekly goal.

For those who purposefully set them, weekly goals appeared to work in concert with time management. Alan, Tracey, and Josie managed their time so that they read regularly by timetabling it and thereby establishing a routine. Once the routine was established, they did not find it difficult to maintain.

The two participants who did not read extensively, Drew and Ben, did not display or describe self-regulating behaviour. Ben did not set any goals. Drew set a goal of 30 books over five months—something only one participant, Josie (who read 34 books), achieved—making it stand out as an example of the "vague, distal goals" that

Zimmerman (2011, p. 56) describes as typical of poorly self-regulated students. Neither Drew nor Ben managed their time either. In fact, neither of them blamed their busy schedules, only their lack of effort to control their time.

Between these extremes are the other four participants. Their ability to self-regulate appeared to be in formation. When there were no problems with the task and no potentially conflicting demands in the environment, they read extensively. When there were problems with the task and potentially conflicting demands in the environment, they read less. This general pattern shows that these participants, unlike Drew and Ben, were able to self-regulate under optimal conditions. However, to varying extents they lost this ability when conditions were less than optimal. In this respect they resemble other participants in ER studies who, despite their intentions, struggled to find time in their busy schedules (Takase, 2004, 2009; Leung, 2002).

However, one difficulty with interpreting the extent of self-regulation under challenging conditions is the extent to which the participants were conscious of how external demands influenced their extensive reading. Jack, Jane, and Emma appeared to weigh the increasing demands made by schoolwork, in particular the NCEA, and decided to study instead of read. Jack appeared to have the most control over this process, managing to continue reading even though it was "hard to fit it all in". As mentioned above, Jack set a weekly goal, and this may have helped him monitor and control his time, particularly since Jack took a great deal of time—nearly an hour—to read one book. Jane and Emma, however, both struggled with managing their time in the face of increased study demands, despite reading quite regularly and extensively in Term One (and, unlike Jack, needing only 15 to 20 minutes to read one book). The few hours they devoted to reading seemed to slip out of their control by being "merged" with study time (Jane) or filled with "something else that I have been missing out" (Emma). This suggests a lack of monitoring as well as a lack of control that may well have been exacerbated by their long-term goals.

Nick is the only participant to have significantly increased the amount of reading he did over three terms, even if his best term and overall totals were markedly less than those of others. In terms of reading quantity and environmental distractions, Nick's first term of ER most closely resembles the last term of Jane and Emma, with very limited reading due to external demands and a lack of self-regulation of behaviour and environment. The difficulty of readers that were above his proficiency level cannot be underestimated, yet Nick persisted, which is crucial to self-regulation (Pintrich, 2004). The turning point was in the second term, when Nick began to exercise more self-regulation over the task (applying knowledge from classes and using the glossaries) and the environment (reading in school, not bed). These positive steps were then reinforced with fewer distractions in the second term and more suitable books in the third term. In the third term, even though Nick faced the same study demands the other students faced, as they read less, he read more (although, since the books were easier, he could spend less time doing so). Although it is clear that other factors—his strong motivation, the suitability of the readers—were also powerful influences on Nick's extensive reading, it is also clear that Nick's

increased self-regulation played a role. ER appeared to push Nick to recognise his need to monitor and control his behaviour and the environment by taking advantage of the tasks and tools used in the ER project.

Conclusion

Research on self-regulation has not been conducted in the ER field, because ER in institutional settings is often an "other regulated" (Zeidner et al., 2000, p. 752) requirement. However, the importance of self-regulation should not be overlooked; this study expands on Robb's (2002) point that students cannot be relied on to self-regulate, by showing there is a range of self-regulatory behaviours in students. It shows that those who self-regulated were more successful in meeting their goals, which suggests that self-regulation may be a powerful means to succeed in learning in general as well as ER and L2 learning, both when it is and is not a requirement (Takase, 2004, 2009).

Limitations of the study

This study is a small-scale study in a particular setting, making it difficult to generalise. However, the setting shares features with some studies on ER in Japan and elsewhere, because the external demand that most distracted students from doing ER in their own time was a national exam. The findings might be useful for researchers and teachers in such settings.

Pedagogical implications

This study has shown that the students who met both study and ER expectations were those who could self-regulate their behaviour and context. In places where exam preparation appears to govern student free time as much (or more than) it has in this study, there is a danger that students may not learn how to self-regulate, but only respond to the external demands, like some of the participants in this study.

This suggests two implications. The less obvious one is that ER may be a means of helping students develop the ability to self-regulate their behaviour, much as it appears to have done for Nick. The more obvious one is that reading in class, as advocated by Takase (2009), would secure a certain amount of reading for all students. This could also provide support for developing the self-regulatory abilities. Social cognitive researchers in educational psychology (e.g., Zimmerman, 2000) have promoted the use of parental, instructor or peer models to develop self-regulation. ER in class could enable students to learn from peers who cope better with reading, possibly motivating them to emulate others (Zimmerman, 2011). The teacher could assist students with goal setting and time management. This is something that can be taught and practiced in class—as occurred with Alan, who apparently was taught such techniques.

Suggestions for future research

Future research could further investigate the role of ER and behavioural and contextual self-regulation by more closely examining how students manage tasks, time, and external demands. These studies should be conducted in various settings, especially with students in different age groups in order to see how self-regulation on specific types of tasks (such as ER) develops with age. Studies with a larger number of L2 learners using mixed methods (e.g., questionnaire and follow-up interviews) may also be fruitful. Also, a longitudinal classroom study that compares ER programmes with different degrees of assistance in self-regulation may be interesting. Such a study could inform us what the teacher can do to help students sustain effort for ER on their own. In addition, given Nishino's (2007) findings and her role in her participants' lives, it might also be interesting to explore the impact of significant others on motivation to read in ER and if it contributes to developing self-regulation.

In conclusion, it appears that research on self-regulation involved in ER has great potential to inform researchers and teachers how they can help L2 learners self-regulate, especially in the area of behavioural regulation, because teachers are not always there to make their students read whether ER is assigned or voluntary. It is hoped that this study inspires them to explore self-regulation in the ER field.

References

Arnold, N. (2009). Online extensive reading for advanced foreign language learners: An evaluation study. *Foreign Language Annals, 42*, 340-366.

Day, R. R., & Bamford, J. (1998). *Extensive reading in the second language classroom*. Cambridge: Cambridge University Press.

de Burgh-Hirabe, R. (2011). *Extensive reading and L2 reading motivation in Japanese as a foreign language: A case study of New Zealand high school students.* Unpublished doctoral dissertation, University of Otago, Dunedin, New Zealand.

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Mahwah, NJ: Lawrence Erlbaum.

Elley, W. B. (1991). Acquiring literacy in a second language: The effect of book-based programs. *Language Learning, 4*. 375-411.

Elley, W. B. & Mangubhai, F. (1981). *The impact of a book flood in Fiji primary schools*. Wellington, New Zealand: New Zealand Council for Educational Research.

Ellis, R., & Barkhuizen, G. (2005). *Analysing learner language*. Oxford: Oxford University Express.

Grabe, W. (2009). *Reading in a second language*. New York: Cambridge University Press.

Hafiz, F. M., & Tudor, I. (1989). Extensive reading and the development of language skills. *ELT Journal, 43*(1), 4-13.

Hirata, A. (2010). *An exploratory study of motivation and self-regulated learning in second language acquisition: Kanji learning as a task focused approach.* Unpublished master's thesis, Massey University, Palmerston North, New Zealand.

Hitosugi, C. I., & Day, R. R. (2004). Extensive reading in Japanese. *Reading in a Foreign Language, 16*(1), 20-39.

Leung, C. Y. (2002). Extensive reading and language learning: A diary study of a beginning learner of Japanese. *Reading in a Foreign Language, 14*, 66-81.

Mason, B., & Krashen, S. (1997). Extensive reading in English as a foreign language. *System, 25*, 91-102.

Mohd Asraf, R., & Ahmad, I. S. (2003). Promoting English language development and the reading habit among students in rural schools through the Guided Extensive Reading program. *Reading in a Foreign Language, 15*, 83-102.

Nation, P. (2007). The four strands. *Innovation in Language Learning and Teaching, 1*(1), 1-12.

Nishino, T. (2007). Beginning to read extensively: A case study with Mako and Fumi. *Reading in a Foreign Language, 19*, 76-105.

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 451-502). San Diego, CA: Academic Press.

Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review, 16*(4), 385-407.

Robb, T. N. (2002). Extensive reading in an Asian context: An alternative view. *Reading in a Foreign Language, 14*, 146-147.

Robb, T. N., & Susser, B. (1989). Extensive reading vs. skills building in an EFL context. *Reading in a Foreign Language, 5*, 239-251.

Tabata-Sandom, M., & Macalister, J. (2009). That 'eureka feeling': A case study of extensive reading in Japanese. *New Zealand Studies in Applied Linguistics, 15*(2), 41-60.

Taguchi, E., Takayasu-Maass, M., & Gorsuch, G. J. (2004). Developing reading fluency in EFL: How assisted repeated reading and extensive reading affect fluency development. *Reading in a Foreign Language, 16*, 70-96.

Takase, A. (2004). Investigating students' reading motivation through interviews. *Forum for Foreign Language Education, 3*. Institute of Foreign Language Education and Research, Kansai University. Osaka: Naniwa Press.

Takase, A. (2009). The effects of SSR on learners' reading attitudes, motivation and achievement: A quantitative study. In A. Cirocki (Ed.), *Extensive reading in English language teaching* (pp. 547-560). Munich, Germany: LINCOM.

Tanaka, H., & Stapleton, P. (2007). Increasing reading input in Japanese high school EFL classrooms: An empirical study exploring the efficacy of extensive reading. *The Reading Matrix, 7*, 115-131.

Tseng, W., Dörnyei, Z., & Schmitt, N. (2006). A new approach to assessing strategic learning: The case of self-regulation in vocabulary acquisition. *Applied Linguistics, 27*, 78-102.

Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language, 15*, 130-163.

Wolters, C. A., Benzon, M. B., & Arroyo-Giner, C. (2011). Assessing strategies for the self-regulation of motivation. In B. J. Zimmerman & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 298-312). New York, NY: Routledge.

Zeidner, M., Boekaerts, M., & Pintrich, P. R. (2000). Self-regulation: Directions and challenges for future research. In M. Boekaerts, P. R. Pintrich & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 749-768). San Diego, CA: Academic Press.

Zimmerman, B. J. (1998). Academic studying and the development of personal skill: A self-regulatory perspective. *Educational Psychologist, 33*, 73-86.

Zimmerman, B. J. (2000). Attaining of self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13-39). San Diego, CA: Academic Press.

Zimmerman, B. J. (2011). Motivational sources and outcomes of self-regulated learning and performance. In B. J. Zimmerman & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 49-64). New York, NY: Routledge.

Zimmerman, B. J., & Martinez-Pons, M. (1986). Development of a structured interview for assessing student use of self-regulated learning strategies. *American Educational Research Journal, 23*, 614-628.

Appendix A

Study participants and background information

Pseudonym	Gender	L1	Age &Year	Visits to Japan	Participation
Jane	F	English	16, Year 13	Nil	Terms 1, 2, 3
Emma	F	English	17, Year 13	2	Terms 1, 2
Josie	F	English	17, Year 13	2	Terms 1, 2, 3
Tracey	F	English	17, Year 13	2	Terms 1, 2
Nick	M	English	14, Year 11	Nil	Terms 1, 2, 3
Jack	M	English	15, Year 11	Nil	Terms, 1, 2, 3
Alan	M	English	16, Year 12	2	Terms 1, 2
Ben	M	English	16, Year 12	1	Terms 1, 2
Drew	M	English	16, Year 12	1	Terms 1, 2

Appendix B

Reading amount and time during Term One, Two and Three

	Term One (8 weeks)	Term Two (10 weeks)	Term Three (9 weeks)	Total
Josie	17 (3.60 hr)	11 (6.40 hr)	6 (6.20 hr)	34 (16.40 hr)
Tracey	16 (4.03 hr +)	10 (-)	----------	26 (-)
Emma	15 (2.53 hr)	8 (1.21hr +)	----------	23 (3.74 hr +)
Alan	6 (3.23 hr)	10 (6.84 hr)	----------	16 (10.07 hr)
Jane	13 (2.84 hr)	2 (0.4 hr)	0 (0 hr)	15 (3.24 hr)
Nick	2 (1.25 hr)	4 (3.75 hr)	8 (2.68 hr)	14 (7.43 hr)
Jack	6 (7.25 hr)	4 (3.6 hr)	2 (1.5 hr)	12 (12.35 hr)
Drew	5 (1.15 hr)	3 (0.5 hr +)	----------	8 (1.65 hr +)
Ben	3 (0.75 hr)	4 (1.08 hr)	----------	7 (1.83 hr)

Note. The students who did not participate in Term Three have ---------- in the Term Three cell. The – indicates Tracey lost her reading record in Term Two. The + indicates that the students did not fill in the reading record completely, so the reading time could not be accurately calculated.

New Zealand Studies in Applied Linguistics, 2012, 18 (1) pp - pp

COMPREHENSION MONITORING IN READING ENGLISH AS A FOREIGN LANGUAGE

Feifei Han

University of Sydney
feifei.han@sydney.edu.au

Abstract

Comprehension monitoring plays an important metacognitive role in constructing a mental representation of a text. Although considerable research has been done in L1 reading, less has been conducted in FL reading. This study investigated comprehension monitoring of two kinds of errors (i.e., external and internal) by 126 Chinese EFL readers with different FL reading proficiency using a one-way ANOVA. In addition, two high and two low proficiency English readers explained the errors they had detected and possible corrections for the errors in retrospective interviews analysed through content analysis. The results suggested that: high proficiency FL readers performed better on comprehension monitoring than low proficiency FL readers; although low proficiency FL readers could recognize that a problem existed in the text, they sometimes failed to retrieve the contradictory information.

Keywords: *foreign language reading, comprehension monitoring, reading proficiency, reading errors, Chinese EFL learners*

Introduction

The primary goal of reading is to construct a mental representation of meaning from a text (Grabe & Stroller, 2002). This sense-making activity involves coordination of multiple levels of sub-component processes, in which readers use a variety of skills and strategies (Block, 2004; Graves, Juel, & Graves, 1998). Among these available skills and strategies, comprehension monitoring plays a crucial role (Alexander & Jetton, 2000; Auerbach & Paxton, 1997; Pressley & Afflerbach, 1995; Wray, 1994; Zinar, 2000).

Although comprehension monitoring has been widely researched in L1 reading with speakers of English as their native language (Baker & Anderson, 1982; Baker & Brown, 1984; Flavell, 1981; Garner & Anderson, 1982; Markman, 1981; Pressley & Ghatala, 1990; Yuill & Oakhill, 1991; Zabrucky & Ratner, 1992; Zinar, 2000), it has received comparatively less attention in foreign language (FL) reading (Han & Stevenson, 2008). The present study investigated comprehension monitoring in English reading by Chinese students learning English as a foreign language (EFL). More specifically, the study examined whether comprehension monitoring differed among Chinese EFL learners with different proficiency.

Literature Review

Comprehension monitoring

Comprehension monitoring is defined as "a metacognitive process…essential for competent reading, which directs the reader's cognitive processes as he/she strives to make sense of incoming textual information" (Wagoner, 1983, p. 328). Comprehension monitoring involves a series of activities which can be categorized into three behaviours: evaluation, planning and regulation (Baker, 1985; Otero, 1998; Paris & Myers, 1981). Evaluation allows readers to evaluate their current understanding of the text and helps them decide whether there is a need to take compensatory actions. If there is, it requires planning to select strategies relevant to the comprehension problem. Regulation then implements the appropriate strategies to fix up comprehension breakdown (Casanave, 1988). For proficient readers, they may re-allocate attention, slow down the speed of reading, re-interpret certain chunks in the text, re-evaluate the hypothesis they have made, and move backward or look ahead in the text in order to resolve the ambiguity (Casanave, 1988; Otero, 1998; Palincsar & Brown, 1984).

Studies in L1 comprehension monitoring

There are a great number of studies on comprehension monitoring in L1 reading. These studies typically required readers to detect the errors or inconsistencies embedded in a text (Garner, 1987). Previous studies examined comprehension monitoring with children (e.g., Zabrucky & Ranta, 1992) and adult readers (e.g., Pressley & Ghatala, 1990); comparing good and poor readers (e.g., Zabrucky & Moore, 1989); using different text types, such as narratives (e.g., Ruffman, 1996) and expositions (e.g., Knudsen, 2001); and manipulating and creating different types of errors (e.g., Kinnuen & Vaurus, 1995), including lexical errors (nonsense words), external errors (information that contradicts general world knowledge); and internal errors (the text contains contradictory information) (Han & Stevenson, 2008; Oakhill, Hartt, & Samols, 2005).

The results of these studies indicate that comprehension monitoring develops over time (Baker, 1984; Garner & Taylor, 1982; Kolić-Vehovec & Bajšanski, 2001). Younger and poorer readers do not monitor their comprehension successfully (Flavell, 1981; Garner, 1990; Garner & Anderson, 1982; Markman, 1981; Yuill & Oakhill, 1991; Zabrucky & Moore, 1989; Zabrucky & Ranter, 1992). There was a general trend that younger and less skilled readers noticed more lexical errors than the other types since they tended to evaluate their comprehension on a lexical level (Westby, 2004). Poorer readers had particular difficulty in the detection of internal errors (e.g. Ehrlich, 1996; Ehrlich, Remond, & Tardieu, 1999), and this could be attributed to the increased difficulty of detecting internal errors rather than external errors, as detection of internal errors requires readers to compare the incoming information with a recently constructed representation of the text, which is less stable than readers' general knowledge about the world (Baker & Zimlin, 1989; Ehrlich,

1996; Ehrlich, Remond, & Tardieu, 1999; Markman, 1985; Oakhill, Harrt, & Samols, 2005; Otero & Kintsch, 1992).

Studies in FL comprehension monitoring

Compared with the amount of research on comprehension monitoring in L1 reading, only scant attention has been paid to comprehension monitoring in FL reading. Most FL comprehension monitoring studies adopted a process orientation using a think-aloud method. These studies compared how native and non-native readers monitored comprehension (Block, 1992); compared bilingual readers' comprehension monitoring processes in reading in their stronger language and weaker language (Jiménez, García, & Pearson, 1996); compared what proficient and less proficient EFL readers did to monitor comprehension in English reading (Yang, 2002); and compared how high reading proficiency and high metacongitive awareness (HPHM) readers and low reading proficiency and low metacognitive awareness (LPLM) readers monitored their comprehension to detect internal errors during English reading (Khonamri & Kojidi, 2011).

Block (1992) found that comprehension monitoring was more likely influenced by reading proficiency than by language background because both proficient L1 and FL readers tended to monitor their comprehension efficiently. Similarly, Jiménez et al. (1996) showed that of 8 bilingual participants, the successful readers carefully monitored their comprehension, whereas the less successful readers employed fewer strategies in resolving comprehension difficulties. In terms of EFL readers' comprehension monitoring, Yang (2002) examined 12 Chinese EFL learners' comprehension monitoring processes and found that good readers displayed more competency in monitoring their ongoing thinking process and were able to integrate the information they previously encountered to interpret the meaning. The poor readers seemed to only process information sporadically and did not know how to integrate it. In a more recent study with Iranian EFL learners, Khonamri and Kojidi (2011) used a reading test and the Metacognitive Awareness of Reading Strategies Inventory (MARSI) (Mokhtari & Reichard, 2002) to select two groups of students. One group of five students (HPHM) obtained higher scores in the reading test and also reported using more reading strategies, while the other group of five students (LPLM) performed poorly in the reading test and its members appeared to be less metacognitively aware via self-reporting to the MARSI. The students were asked to think-aloud when reading an English text embedded with some internal errors. The participants' use of reading strategies measured by think-aloud protocols was analysed qualitatively. The results showed that HPHM readers used all kinds of reading strategies more frequently than did LPLM readers, and in particular, HPHM readers employed nearly double the amount of comprehension monitoring strategies when compared to LPLM readers. Although the study revealed that HPHM readers used more monitoring strategies than LPLM readers did, it did not examine how successfully the readers monitored their comprehension (i.e. the rate of error detection) and whether their reading proficiency affected the rate of error detection.

Two studies compared L1 and FL comprehension monitoring outcomes. Morrison (2004) compared Canadian English-speaking undergraduates' comprehension monitoring in English and French (FL) by asking students to detect lexical errors and information errors (i.e., internal and external errors). She found that students were more successful in the detection of information errors than lexical errors. Han and Stevenson (2008) pointed out that the poorer performance on the detection of lexical errors compared to information errors could be partly attributed to FL readers' limited lexical knowledge. In order to truly examine FL readers' comprehension monitoring, Han and Stevenson (2008) used only information errors (i.e., internal and external errors) to compare Chinese EFL readers' comprehension monitoring in Chinese and English reading. They found that participants performed significantly better in comprehension monitoring in L1 reading than in FL reading, with FL reading comprehension monitoring possibly explaining around 5% of variance in FL reading proficiency.

The present study

While L1 reading research indicated that comprehension monitoring outcomes differ between more and less proficient readers (e.g. Baker & Anderson, 1982; Baker & Brown, 1984; Pressley & Ghatala, 1990), and FL reading studies showed some qualitative differences in comprehension monitoring processes among readers of different proficiency, there is a lack of research on how FL reading proficiency may influence FL comprehension monitoring outcomes. The present study adopted a mixed-method design investigating comprehension monitoring of FL readers with different FL reading proficiency (high, medium and low proficiency level).

This study addresses the question:
How does comprehension monitoring performance in FL reading differ among FL readers with different reading proficiency as reflected in an error detection task and retrospective interviews?

Method

Research design

The study adopted a mixed-methods design. Cohen and Manion (1985) pointed out that reliance on only one kind of research method is very likely to result in a one-sided representation, whereas the use of two or more methods from multiple perspectives has the capability to represent more fully the complexity of issues being researched. In the present study, quantitative data were collected for the outcome of comprehension monitoring, whereas qualitative data were collected to reflect the process of monitoring comprehension, such as how learners evaluated their comprehension, recognised the breakdown of their comprehension, identified the sources of breakdown, and corrected problems.

Setting and participants

The study was conducted in the Vocational Department in a provincial university in Shaanxi Province, China. Forty-one male and 85 female students from in year 2, participated in the study. Their mean age was 18 years old and on average they had received 4.5 years English education.

Participants' mid-term examination scores in an extensive reading course were collected as indicators of their English reading proficiency. The examination used a variety of formats to check students' comprehension for reading five short English texts, including multiple-choice questions, true or false, and short-answer questions. Students' scores ranged from 35 to 94 out of 100.

The error detection task

Participants' FL reading comprehension monitoring was measured by using an error detection task, which asked students to read two English narratives and to underline information errors in them. The error detection task was adopted from Han and Stevenson (2008) (see Appendix 1 for a sample text). Text one (T1), *The poor man and his three sons,* was a Philippine folktale (601 words), and text two (T2), *Anya's garden,* was an Indian folktale (580 words) (Bedjos, 1993). The reason for choosing stories from Asia was that reading Asian stories would not require novel culture knowledge in order to be understood. The decision was made to use narratives because this level of participant reads English narratives more often than other text types. In order to make sure that the chosen texts were suitable for the participants' English reading proficiency, their English teachers were consulted about the difficulty of linguistic features of the two texts.

As mentioned earlier, because using lexical errors could examine FL readers' vocabulary knowledge rather than their comprehension monitoring, the error detection task only used information errors, namely external and internal errors. One example of an external error was "He eats water*"*, in which "eats water*"* was contradictory to general world knowledge. An example of an internal error was "Alice could not open the kitchen door because her key was bent. She walked through the kitchen door*"*, in which "walked through the kitchen door*"* violated the information "could not open the kitchen door*"*. Five external errors and 5 internal errors were scattered throughout each text for a total of 10 external and 10 internal errors. Each correct error detection received 1 point, and the maximum scores for the error detection task were 20. The reliability analysis showed that the task was quite reliable, with the Cronbach's alpha coefficient .82.

Retrospective interviews

Retrospective interviews require learners to report their thoughts after they have completed the task (Cohen, 1996). Since the lag between verbalisation and tasks may alter genuine thinking processes (Ericsson & Simon, 1993), the shorter the time interval between the retrospection and the tasks, the more reliable they are (Yamashita, 2002). In consideration of the reliability of the retrospective interviews,

the participants were selected and contacted before they carried out the error detection task to ensure that the interviews were conducted on the same day as the error detection task was.

Four students from those who were going to take part in the error detection task were selected and contacted in advance to obtain their permission to participate in retrospective interviews. The selection of the participants was based on scores in the mid-term exam in their extensive course. Students A and B both scored 92 and were regarded as proficient English readers, whereas students C and D scored 40 and 49 respectively and were categorized as poor English readers. The information on the participants is summarised in Table 1.

Table 1 Information on the four participants in the retrospective interviews

Students	**English reading scores**	**Scores in English error detection task**	
		External	Internal
A	92	4	9
B	92	7	7
C	40	4	4
D	49	2	2

The retrospective interview used only one text, *The poor man and his three sons,* as using two texts would cause fatigue for the participants. In the interview, the researcher presented each participant's own answers for the error detection task and asked them (1) why they underlined the sentence; (2) what was wrong with it; (3) how they could correct it to make sense of it.

Data analysis

Quantitative data analysis was performed with SPSS 18.0 for descriptive statistics, one-way ANOVA and post-hoc analysis. Before the analysis, participants were first divided into high, medium and low proficiency FL readers. Scores of 70 and 60 were used as cut-off points to group the students, as the score of 70 is average and the score of 60 is passing, according to the common Chinese academic standard. Students who obtained a score equal or above 71 were grouped as high proficiency readers, those whose scores were between 70 and 60 were classified as medium proficiency readers, and low proficiency readers included students whose scores were below 60. Table 2 presents the distribution of the participants by levels of English reading proficiency.

Table 2 Distribution of participants by levels of English reading proficiency

Groups	**Number of students**	**Minimum score**	**Maximum score**	**Mean**
High	43	71	94	80.98
Medium	50	60	70	63.36
Low	33	35	56	45.91

Content analysis was used to analyse the qualitative data. Berg (2007) suggested that content analysis is "a careful, detailed and systematic examination and interpretation" of unstructured word-based data in order to identify "patterns and themes" (p. 303). The content analysis was conducted following two steps proposed by Brown (2002). The first step is to identify and label each individual idea unit, which is "a response referring to a single concept, idea or feature in statements" (Pereira, 1991, p. 54). The second step was to form and name categories. This was done in a cyclical manner for categories to emerge, and the categories were then labelled.

The content analysis revealed that participants' retrospective reporting could be categorised into error identification, sources of error clarification, and correction of errors. For error identification, the participants could successfully identify the error (e.g., "'*Eat'* seems not right here"), or they could realise a comprehension breakdown was occurring (e.g., "There must be something wrong*'*"), or they failed to recognise inconsistent information (no comment). When they were trying to locate the source of a problem, they could sometimes realise that the information was against general knowledge (e.g., "Human beings cannot sleep in the river"); or they might find one sentence contradicted a previous sentence (e.g., "In the previous text, it said 'the man had three sons'"); or they felt something was wrong using their intuition (e.g., "I judged on my intuition"). With regard to the correction of errors, some readers could offer a few reasonable corrections, whereas others failed to do so. Examples of the participants' comments are listed in Appendix 2.

Results and Discussion

The descriptive statistics of high, medium and low proficiency FL readers' overall comprehension monitoring scores, scores of comprehension monitoring of external errors and internal errors in FL reading are displayed in Table 3. The results of the one-way ANOVA are displayed in Table 4.

Table 3 Descriptive statistics of FL comprehension monitoring

Variables	**FL reading proficiency levels**	**M**	**SD**
Overall	High	10.77	4.20
	Medium	8.60	4.49
	Low	7.67	4.20
External	High	6.07	2.38
	Medium	4.86	2.63
	Low	4.24	2.31
Internal	High	4.70	2.17
	Medium	3.74	2.22
	Low	3.42	2.41

Table 4 Results of one-way ANOVA

Variables	**df**	***F***	***p***	**η^2**
Overall	2,123	5.37	.01	.08
External	2,123	5.55	.01	.08
Internal	2,123	3.47	.03	.05

Table 4 showed that statistically significant differences in overall comprehension monitoring in English reading were found among the three groups of readers ($F(2,123) = 5.37$, $p = .01$, $\eta^2 = .08$). Moreover, the three groups of readers also differed in comprehension monitoring of external errors ($F(2,123) = 5.55$, $p = .01$, $\eta^2 = .08$) and internal errors ($F(2,123) = 3.47$, $p = .03$, $\eta^2 = .05$). The effect size values of eta-squared ranged between .05 to .08.

Tukey's HSD post-hoc tests were then performed and results are displayed in Table 5. The results show firstly, that high proficiency English readers performed better than medium (MD = 2.17, $p = .04$) and low proficiency (MD = 3.10, $p = .01$) English readers in overall comprehension monitoring in English reading. However, no difference was found between medium and low proficiency English readers in overall comprehension monitoring. Secondly, in terms of comprehension monitoring of external errors, high proficiency readers scored higher than low proficiency readers, with the mean difference being 1.83 (MD = 1.83, $p = .01$). However, no difference was found either between high and medium proficiency English readers or between medium and low proficiency English readers. Lastly, for comprehension monitoring of internal errors, high proficiency English readers also obtained higher scores than low proficiency counterparts, and there was 1.27 mean difference between the two groups of readers (MD = 1.27, $p = .04$). But this did not appear to be different either between high and medium proficiency English readers or between medium and low proficiency English readers.

Table 5 Results of Tukey's HSD post-hoc tests

Variables	**Contrasts**	**MD**	**Standard error**	***p***
Overall	High-Medium	2.17	.90	.04
	High-Low	3.10	.99	.01
	Medium-Low	.93	.97	.60
External	High-Medium	1.21	.51	.06
	High-Low	1.83	.57	.01
	Medium-Low	.62	.55	.51
Internal	High-Medium	.96	.47	.11
	High-Low	1.27	.52	.04
	Medium-Low	.32	.51	.81

The content analysis for the qualitative data from the retrospective interviews showed that comprehension monitoring of the higher proficiency English readers was different from that of lower proficiency English readers in English reading. The

differences were found mainly in: error identification and sources of error clarification.

In the error identification phase, participants A and B, the two high proficiency English readers, recognised more problems than participants C and D, the low proficiency English readers. In addition, participants C and D appeared to be too concerned with grammatical problems. For example, participant C underlined the sentence "But I do have something for you*"*, which in fact does not contain any error. She explained that there should be something between *do* and *have*, such as the word *not* in order to make it grammatical correct. She further commented that because there is no need to use an auxiliary verb *do* in a positive sentence, and the auxiliary verb *do* is only used in a negation sentence.

In terms of sources of error clarification, participants A and B could provide reasonable explanations for each sentence they underlined, and correct them even though sometimes the researcher had not yet asked them for corrections. From their answers, it could be seen that they could hold the content of the preceding part of the text in their memory and compare this with the newly encountered information. Each time they explained an internal error, they located the contradictory sentence in the previous text without hesitation. For instance, with regard to the sentence "The sons thanked their father for the gold*"*, student A not only was able to remember the exact text but was also able to elaborate on it. She stated, "The previous text said he *'*had no gold*'*, and in fact 'the man gave his sons a cat, a rooster and a…'." In a similar vein, student B also held the previous textual information in her memory and quickly located the sentence and explained, "Here said that *'*he had no gold'." When participants C and D realised something was wrong, they quite often attributed it to their intuition. Participant D considered the sentence "He slept for days and day*"* was problematic by commenting, "I felt something wrong here. I felt it shouldn't be like that".

In summary, the results from the quantitative analysis showed that the outcomes of FL readers' comprehension monitoring is similar to that obtained in L1 comprehension monitoring studies in that more proficient L1 readers performed better on comprehension monitoring than less proficient L1 readers. In the present study, more proficient FL readers performed better on overall comprehension monitoring, internal error detection and external error detection. The qualitative analysis also revealed that low proficiency FL readers sometimes appeared to be concerned with grammatical structures in English reading. Even if they realised a problem in the text, they sometimes failed to retrieve the contradictory information, and they often attributed the error detection to their own intuition. The findings of this study support both previous L1 and FL reading comprehension monitoring studies.

Conclusion

This study has shown that proficiency levels make a difference in comprehension monitoring in FL reading of narratives by examining both quantitatively and quantitatively analyzed data.

Some limitations of the study should be pointed out. First and foremost, the present study only used an error detection task and a very limited number of retrospective interviews to examine participants' comprehension monitoring. No information on participants' concurrent comprehension monitoring process has been collected. Secondly, only a single text type, narratives, were used in the study.

These limitations suggest directions for future research, which could combine information from larger numbers of both readers' on-line comprehension monitoring (e.g., think-alouds) and off-line comprehension monitoring (e.g., error detection tasks). Additional text types could be used to compare FL readers' comprehension monitoring in different text types. Such studies would add to the knowledge of FL reading comprehension, just as this study has done for narrative.

References

Alexander, P. A., & Jetton, T. L. (2000). Learning from text: A multidimensional and developmental perspective. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (pp. 285-310). Mahwah, NJ: Lawrence Erlbaum Associates.

Auerbach, E., & Paxton, D. (1997). "It's not the English thing": Bringing reading research into the ESL classroom. *TESOL Quarterly, 31,* 237-261.

Baker, L. (1984). Spontaneous versus instructed use of multiple standards for evaluating comprehension: Effects of age, reading proficiency and types of standard. *Journal of Experimental Child Psychology, 38,* 289-311.

Baker, L. (1985). How do we know when we don't understand? In D. L. Pressley, G. E. MacKinnon, & T. G. Waller (Eds.), *Metacognition, cognition, and human performance: Theoretical perspectives* (pp. 155-206). Orlando: Academic Press.

Baker, L., & Anderson, R. (1982). Effects of inconsistent information on text processing: Evidence for comprehension monitoring. *Reading Research Quarterly, 17,* 281-294.

Baker, L., & Brown, A. L. (1984). Metacognitive skills and reading. In P. D. Pearson, R. Barr, M. L. Kamil, & P. Mosenthal (Eds.), *Handbook of reading research* (pp. 353-394). White Plains, NY: Longman.

Baker, L., & Zimlin, L. (1989). Instructional effects on children's use of two levels of standards for evaluating their comprehension. *Journal of Educational Psychology, 81,* 340-346.

Bedjos, J. B. (1993). *Festival of folktales: Stories for whole language learning.* San Diego, CA: Dominie Press.

Berg, B. L. (2007). *Qualitative research methods for the social sciences.* Boston, MA: Pearson Education, Inc.

Block, C. C. (2004). *Teaching comprehension: The comprehension process approach.* Boston, MA: Pearson Education Inc.

Block, E. (1992). See how they read: Comprehension monitoring of L1 and L2 readers. *TESOL Quarterly, 20,* 319-343.

Brown, J. D. (2002). *Doing second language research.* Oxford: Oxford University Press.

Casanave, C. (1988). Comprehension monitoring in ESL reading: A neglected essential. *TESOL Quarterly, 22,* 283-302.

Cohen, A. D. (1996). Verbal protocols as a source of insights into second language learner strategies. *Applied Language Learning, 7,* 5-24.

Cohen, L., & Manion, L. (1985). *Research methods in education.* Sydney: Croom Helm.

Ehrlich, M. (1996). Metacognitive monitoring in the processing of anaphoric devices in skilled and less-skilled comprehenders. In C. Cornoldi & J. Oakhill (Eds.), *Reading comprehension difficulties: Processes and remediation* (pp. 221-249). Mahwah, NJ: Lawrence Erlbaum Associates.

Ehrlich, M., Remond, M., & Tardieu, H. (1999). Processing of anaphoric devices in young skilled and less skilled comprehenders: Differences in metacognitive monitoring. *Reading and Writing: An Interdisciplinary Journal, 11,* 29-63.

Ericsson, A., & Simon, H. (1993). *Protocol analysis: Verbal reports as data.* Cambridge, MA: The MIT Press.

Flavell, J. H. (1981). Cognitive monitoring. In W. P. Dickson (Ed.), *Children's oral communication skills* (pp. 35-60). New York: Academic Press.

Garner, R. (1987). *Metacognition and reading comprehension.* Norwood, NJ: Ablex.

Garner, R. (1990). When children and adults do not use strategies: Toward a theory of settings. *Review of Educational Research, 60,* 517-529.

Garner, R., & Anderson, J. (1982). Monitoring-of-understanding research: Inquiry directions, methodological dilemmas. *Journal of Experimental Education, 50,* 70-76.

Garner, R., & Taylor, N. (1982). Monitoring of understanding: An investigation of attentional assistance needs at different grade and reading proficiency levels. *Reading Psychology: An International Quarterly, 3,* 1-6.

Grabe, W., & Stoller, F. (2002). *Teaching and researching reading.* Harlow: Longman.

Graves, M. F., Juel, C., & Graves, B. B. (1998). *Teaching reading in the 21st century.* Needham Heights, MA: Allyn and Bacon.

Han, F. & Stevenson, M. M. (2008). Comprehension monitoring in first and foreign language reading. *University of Sydney Papers in TESOL, 3,* 73-110.

Jiménez, R., García, G., & Pearson, P. (1996). The reading strategies of bilingual Latina/o students who are successful English readers: Opportunities and obstacles. *Reading Research Quarterly, 31,* 90-112.

Khonamri, F., & Kojidi, E. M. (2011). Metacognitive awareness and comprehension monitoring in reading ability of Iranian EFL learners. *Profile, 13,* 99-111.

Kinnunen, R., & Vaurus, M. (1995). Comprehension monitoring and the level of comprehension in high- and low-achieving primary school children's reading. *Learning and Instruction, 5,* 143-165.

Knudsen, E. V. (2001). *Understanding reading comprehension difficulties at the university level.* Unpublished doctoral dissertation, University of Western Ontario, London, Ontario, Canada.

Kolić-Vehovec, S., & Bajšanski, I. (2001). Children's metacognition as predictor of reading comprehension at different developmental levels. *ERIC,* ED45422.

Markman, E. (1981). Comprehension monitoring. In W. P. Dickson (Ed.), *Children's oral communication skills* (pp. 61-84), New York: Academic Press.

Markman, E. (1985). Comprehension monitoring: Developmental and educational issues. In F. Chipman, W. Segal, & R. Glaser (Eds.), *Thinking and learning skills: Research and open questions* (pp. 275-291). Hillsdale, NJ: Lawrence Erlbaum Associates.

Mokhtari, K., & Reichard, C. A. (2002). Assessing students' metacognitive awareness of reading strategies. *Journal of Educational Psychology, 94,* 249-258.

Morrison, L. (2004). Comprehension monitoring in first and second language reading. *The Canadian Modern Language Review, 61,* 77-106.

Oakhill, J. V., Hartt, J., & Samols, D. (2005). Levels of comprehension monitoring and working memory in good and poor comprehenders. *Reading and Writing: An Interdisciplinary Journal, 18,* 657-686.

Otero, J. (1998). Influence of knowledge activation and context on comprehension monitoring of science texts. In J. Hacker, J. Dunlosky, & C. Graesser (Eds.), *Metacognition in educational theory and practice.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Otero, J., & Kintsch, W. (1992). Failures to detect contradictions in a text: What readers believe versus what they read. *Psychological Science, 4,* 229-235.

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction, 1,* 117-175.

Paris, S. G., & Myers, M. (1981). Comprehension monitoring, memory, and study strategies of good and poor readers. *Journal of Reading Behaviour, 13,* 5-22.

Pereira, L. J. (1991). *The role of the situation model in the reading comprehension processes of grade seven readers.* Unpublished doctoral dissertation, the University of Victoria, Victoria, Canada.

Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Pressley, M., & Ghatala, E. S. (1990). Self-regulated learning: Monitoring learning from text. *Educational Psychologist, 25,* 19-33.

Ruffman, T. (1996). Reassessing children's comprehension-monitoring skills. In C. Corroldi & J. V. Oakhill (Eds.), *Reading comprehension difficulties: Processes and intervention* (pp. 33-67). Mahwah, NJ: Lawrence Erlbaum Associates.

Wagoner, S. (1983). Comprehension monitoring: What it is and what we know about it. *Reading Research Quarterly, 18,* 328-346.

Westby, C. (2004). A language perspective on executive functioning, metacognition, and self-regulation in reading. In C. A. Stone, E. R. Silliman, B. J. Ehren, & K. Apel (Eds.), *Handbook of language and literacy: Development and disorders* (pp. 398-427). New York: The Guilford Press.

Wray, D. (1994). Comprehension monitoring, metacognition and other mysterious processes. *Support for Learning, 9,* 107-113.

Yamashita, J. (2002). Reading strategies in L1 and L2: Comparison of four groups of readers with different reading ability in L1 and L2. *I.T.L. Review of Applied Linguistics, 135-36,* 1-35.

Yang, Y. F. (2002). Reassessing readers' comprehension monitoring. *Reading in a Foreign Language, 14,* 18-42.

Yuill, N., & Oakhill, J. (1991). *Children's problems in text comprehension: An experimental investigation.* Cambridge: Cambridge University Press.

Zabrucky, K., & Moore, D. (1989). Children's ability to use three standards to evaluate their comprehension of text. *Reading Research Quarterly, 24,* 336-352.

Zabrucky, K., & Ranter, H. (1992). Effects of passage type on comprehension monitoring and recall in good and poor readers. *Journal of Reading Behaviour, 24,* 373-391.

Zinar, S. (2000). The relative contributions of word identification skill and comprehension-monitoring behaviour to reading comprehension ability. *Contemporary Educational Psychology, 25,* 363-377.

Appendix 1 Sample of the error detection task and solutions

Bold indicates an external error; shaded indicates an internal error. The correct answer is given in brackets after the inconsistent word(s).

The Poor Man and His Three Sons

Long ago in the Philippines, a poor farmer lived with his three sons. The man had worked hard all his life. But he had no gold. His riches were his three sons.

One day, the farmer was very sick. He knew that he was dying. He called his sons to his bed. “My sons, I will die soon. I have no gold. But I do have something for you.”

The sons listened sadly. They didn’t want their father to die.

First, the father gave the oldest son a cat. Then he gave the second daughter (son) a rooster. And to the youngest son he gave a scythe. The father said, “Travel far and wide. Find **a school (the right place)** to trade the rooster, the cat and the scythe. Then you will be happy.”

The sons thanked their father for the gold (gifts). The father spoke no more. He had died.

Months passed. Life was hard for the sons. The farm was on poor **trees (land)**. There were no horses or oxen on the farm. The sons had little **rubbish (food)** to eat.

“Father said to trade the cat,” said the oldest son. “But who wants to trade for a cat? Everyone on the island has a cat.”

“Father said to travel far and wide,” said the youngest son. “I will start tomorrow.” The youngest son left his brothers. He took the scythe with him. He slept (travelled) for days and days. At last he came to a small island. There he saw some farmers picking rice. He was surprised. They were picking the rice with their hands.

“May I help you? I can cut the rice fast.” The youngest son started cutting the rice with his scythe.

The men were surprised. They had never seen a scythe before. “You can cut rice so quickly!” the men said, “What is that in your hand?”

The youngest son said, “It is a scythe. My father gave it to me.”

The men talked quietly together. Then they said, “We want to **eat (trade for)** your scythe. We will give you gold.”

The youngest son thought for a moment. Then he said, “I will trade with you.” He gave the scythe to the men. Then the young man travelled home. He showed his father (brother) the riches.

“I can’t believe it!” said the middle son. “Father was right. You need to find the right place to trade.”

The next day, the middle son travelled far and wide. He took his rooster with him. At last he found a village without roosters. He slept in the **river (street)** with his rooster. When the sun came up, the rooster crowed. People looked out their windows. “What is that?” they asked.

The middle son answered, “It is my rooster. It is crowing. It wakes me every morning.” The people wanted the rooster for their village. They offered gold to the middle son. Happily, the young man took the gold. He returned home with the rooster (rich and happy).

The oldest son said to his brothers, “I don’t think I can trade my cat. But my father wanted me to try.” He travelled for days and months. At last he came to an island with no cats. The island had problems with rats. The son put his cat down. The cats ran after the rats. Soon there were no more rats.

The people said, “We must have this cat! Will you trade for it?”

“Yes, I will,” said the son with a smile. The people gave the young man a bag of gold.

The oldest son went home. He showed his riches to his brothers. The three sons lived happily for many years.

Appendix 2 Examples of participants' comments in retrospective interview

Phases	**Situations**	**Examples**
error identification	explicitly identifying an error	*How does it become **mother**?*
		***Eat** seems not right here.*
		*Not **repair** here.*
	realising a comprehension breakdown	*It seems not practical.*
		There must be something wrong.
		Haha…funny!
	failing to recognise an error	No comment.
Sources of error clarification	realising the information is against general knowledge	*It is impossible for them to eat rubbish.*
		Human beings cannot sleep in the river.
		*When they were in famine, how could they **ate and drank extravagantly**?*
	realising the information contradicted to other textual information in the text	*In the previous text, it said **the man had three sons**.*
		*The previous text said he **had no gold**, and in fact **the man gave his sons a cat, a rooster and a …**sorry I cannot pronounce it.*
		*How could he **slept for days and days** but then **came to a small island**?*
	using one's intuition	*I judged it on my intuition.*
		I forget why I underlined it, can't remember.
		I felt it should not be like that, am I right?
correction of error	offering one or more reasonable corrections to error	*So it must be the second son, not **the second daughter.***
		Maybe arrows are on the floor or in other rooms, but not in the sky.
		I think they were hungry and had nothing to eat.
	failing to provide correction to error	No comment.

Words in **bold** indicate the exact words from the texts in the error detection tasks.

New Zealand Studies in Applied Linguistics, 2012, 18 (1) pp - pp

LEARNER AUTONOMY IN LANGUAGE LEARNING: HOW TO MEASURE IT RIGOROUSLY

Le Thi Cam Nguyen

Victoria University of Wellington
le.nguyen@vuw.ac.nz

Abstract

This article reports on the application of three principles used to measure learner autonomy. The three principles involved offer a clear definition of learner autonomy, investigating it from different perspectives and validating research tools. Particular attention is given to the description of the methodology in order to show how learner autonomy can be rigorously investigated in a three-phase study. At the macro-level, both qualitative and quantitative approaches to exploring learner autonomy were used. At the micro-level, there was an ongoing refinement of the research instruments. A variety of tools was used to encourage university EFL participants to provide rich and reflective accounts of learner autonomy in the sociocultural setting of Vietnam.

Keywords: *learner autonomy, measuring learner autonomy, promoting learner autonomy, learner autonomy and language proficiency, metacognitive training.*

Introduction

There has been growing interest in the role of learner autonomy in language teaching and learning. A number of studies have been conducted to investigate the strengths of learner autonomy and different approaches to promoting it. However, most of them are descriptive in nature. Research methods for investigating learner autonomy have included teachers' observations, interviews and students' learning journals (Tagaki, 2003), students' self-assessment and peer assessment (Nachi, 2003; Natri, 2007), students' feedback or evaluation sheets (Nicoll, 2007; Sert, 2006), oral interviews and questionnaires (Pickard, 1995, 1996), learner logs and evaluation of learning (Pearson, 2004), teachers' diaries and students' evaluation (Dam, 1995), students' reflective writing (Smith, 2001), students' portfolios (Nunes, 2004; Rao, 2005; Shimo, 2003) and questionnaires (Chan, 2001, 2003; Spratt, Humphrey, & Chan, 2002). Many researchers have claimed that learners in their studies became more autonomous. Their statements have been based on learners attending class more regularly (Tagaki, 2003), actively engaging in classroom activities (Dam, 1995; Natri, 2007; Nunes, 2004; Rao, 2005), demonstrating a high level of reflection (Kohonen, 2000, 2001; Mizuki, 2003; Shimo, 2003), and accepting responsibility for their own learning (Cunningham &

Carlton, 2003; Stephenson & Kohyama, 2003). These studies show considerable insight into learners' autonomous behaviours, but they are often not strong on providing empirical evidence of the tangible benefits of learner autonomy. Several studies have demonstrated the link between learner autonomy and language learning outcomes (Champagneet al., 2001; Dam & Legenhausen, 1996; Vickers & Ene, 2006). However, due to a lack of compatibility among groups of participants, for example, they have not produced sufficiently strong and convincing evidence. There is a need for evaluating and measuring learner autonomy more rigorously, which, if done properly, could provide persuasive evidence of the advantages of learner autonomy for language learning.

Important factors affecting the measurement of learner autonomy are discussed by Benson (2001). Firstly, learner autonomy is a multidimensional construct. It is possible to identify and list behaviours that display learners' control over their learning such as self-accessing their learning, reflecting on the value of activities they initiate to improve their learning or designing their own learning programmes. However, there are not sufficient grounds to conclude that autonomy consists of any specific combination of those behaviours. Also, the extent and the degree to which learners are autonomous depend on a range of variables such as the cultural context, the particular situation, the stage of learning, the individuals and their experiences. Secondly, learners may possess autonomy as a capacity but not necessarily exercise these skills. They know how to control and manage their learning but do not use this knowledge. Thirdly, learners may acquire autonomy as a result of developmental processes. The more mature they become, the more autonomy they gain. Benson (2001, p.188-190) also suggests various ways of measuring learner autonomy in language learning including: 1) finding out whether learners make and use a learning plan, take part in classroom decisions, reflect upon their learning, and initiate changes in a target language; and 2) looking at whether learners are able to create situations of learning for themselves and to monitor and self-access their own performance. Measuring learner autonomy is a difficult matter (Benson, 2001; Mynard, 2006) due to the variety of factors affecting it and the complexity of the construct.

How to measure learner autonomy rigorously

Learner autonomy is a complicated phenomenon. It can be measured using three principles. The first principle is having a clearly-defined notion of learner autonomy based on which any accounts of learner autonomy can be analysed and measured. The second principle is looking at learner autonomy from a variety of points of view and employing both qualitative and quantitative methods to collect data as each can supply equally valuable, but different, data. The third principle is ensuring that the tools are carefully developed, piloted, and validated so they can do the best job possible. Multi-item questionnaires were used in this study to measure learner autonomy so it is important to validate them to ensure that each

item on a scale correlates with the other items and with the total scale score.

Having an operationalised definition of learner autonomy

Within the learner autonomy field, a large number of terms are used to refer to an almost identical concept. This causes confusion, especially to novice researchers and practitioners. Additionally, in the existing notions of learner autonomy, the coverage is too broad and general on the scale of evaluation and measurement of learner autonomy, which makes it difficult for researchers to measure learner autonomy. Language education tends to explore psychological aspects of learner autonomy, which focus on learners' abilities and the internal changes that they make in the learning process. Holec's (1981) definition is the one most often cited. There are four main characteristics in his definition. Firstly, autonomy is an ability to take charge of one's own learning. Secondly, this ability is not innate but is necessarily acquired through systematic and purposeful learning. Thirdly, autonomy is a potential capacity to act in a learning situation, and not the actual behaviour of an individual in that situation. The fourth feature is related to learners' ability to take charge of their learning by becoming responsible for the decisions made in the learning process, including deciding the objectives, identifying the contents and progressions, selecting methods and techniques, and monitoring the procedure of acquisition, and evaluating what is acquired (Holec, 1981, p. 3). Being the most popularly accepted, the traditional notion of autonomy conceptualised by Holec (1981) has served as a fundamental description of learner autonomy. While it is agreed that learner autonomy occurs universally, an operationalised definition seems to be essential to researching learner autonomy in a particular educational setting.

The operational concept of learner autonomy used in Nguyen's (2008, p. 68) study is illustrated in Figure 1 below. In this conception of learner autonomy, the two basic elements of self-initiation and self-regulation are closely connected. Self-initiation is learners' volition and willingness to learn without any kind of coercion, persuasion or external initiation. It is broken into *reasons for learning* and *making efforts to learn*. While the former indicates the cause or motive for learning, the latter implies acts of initiating learning activities and behaviours to support learning. Self-regulation involves the metacognitive skills of *planning*, *monitoring*, and *evaluating*.

Both elements of learner autonomy involve the interaction between the learner and the task. The self-regulation component in essence represents a set of learning strategies. It is skill-focused and could possibly be improved through training. The self-initiation is learner-driven. This definition should work in any context where learners are not in a position to take control over the content of the learning, one of the three levels of control discussed by Benson (2001). Within the classroom, learners are encouraged to use the self-regulatory skills of planning, monitoring,

and evaluating to perform any task given.

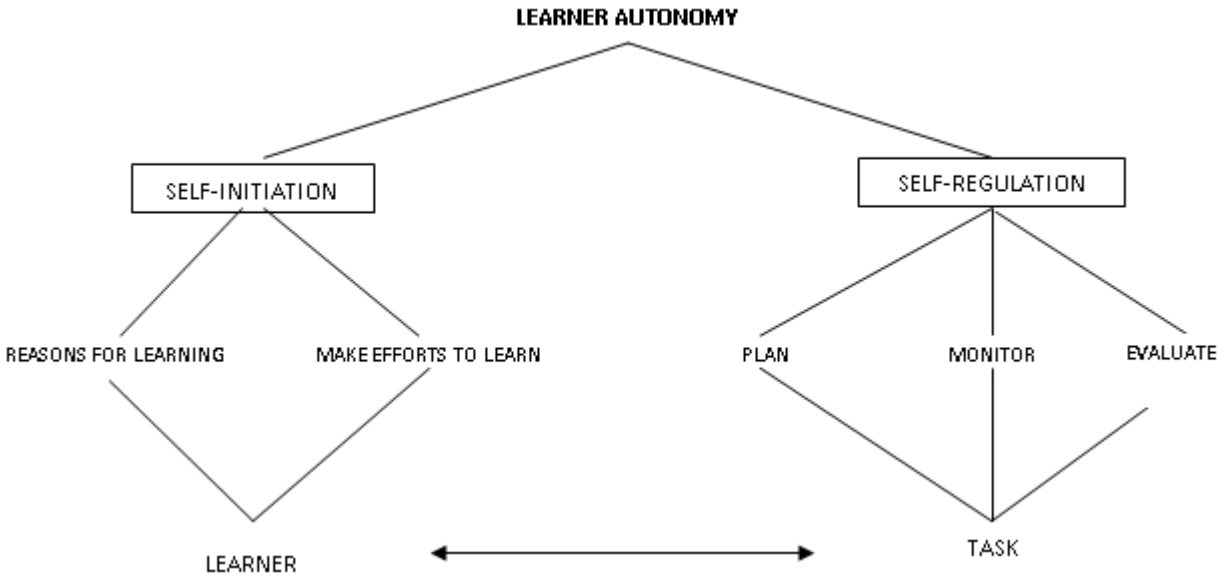


Figure 1: Operational concept of learner autonomy

Looking at learner autonomy from different perspectives using different tools

The main purpose of this study was to explore aspects of learner autonomy demonstrated by Vietnamese students at a university in Vietnam and to find an appropriate approach to promoting it. The study, carried out in three phases including a pilot study and two main phases, followed the four main ways of evaluating learner autonomy proposed by Sinclair (1999), including collecting feedback from teachers and learners, logging learners' behaviours, researching the effects of strategy training, and monitoring learners' gains in proficiency in the target language. Table 1 below illustrates research purposes, research questions, research instruments and the number of learners of English as a Foreign Language (EFL) who volunteered their participation at each stage of the study. The participants were the same cohort of learners who were at their first year of university when the pilot study started and became second-year students at Phase One and third-year students at Phase Two of the study. Table 1 also indicates the link between Phase One and Phase Two of the study. Phase One was intended to investigate the relationship between learner autonomy and language proficiency, which was statistically tested through correlations. Based on the results of Phase One, which demonstrated a positive connection between learner autonomy and language proficiency (Nguyen, 2008), Phase Two looked at the effectiveness of metacognitive training aimed at fostering learner autonomy.

To measure learner autonomy rigorously, both quantitative and qualitative tools were used to collect different points of views of learner autonomy for the study.

Quantitative data originated from questionnaires, which were used in both Phase One and Phase Two of the study, and the pre- and post- writing tests in Phase Two. Qualitative data came from interviews, learners' learning logs, learners' diaries, and classroom observations in the pilot study and Phase Two as indicated in Table 1.

Table 1: Summary of research purposes, questions, and instruments

Phase	Research purpose	Research questions	Research instruments	Number of participants
Pilot study	- testing research instruments - looking for initial indications of the nature of the relationship between learner autonomy and language proficiency	1. Is there a relationship between learner autonomy and language proficiency? 2. Are there differences in learner autonomy among students of different year levels? 3. Are there differences in the number of activities and the amount of time devoted to learning English by learners of different levels of academic achievement?	Questionnaire Interview and Questionnaire Learner learning logs	389 6 students 181 students 15 students
Phase one	- investigating the relationship between learner autonomy and language proficiency - exploring learner autonomy in a Vietnamese educational context	1. Are Vietnamese undergraduate students of English autonomous learners? 2. What are the most popular learner self-initiated out-of-class and in-class activities performed by these Vietnamese students? 3. What is the relationship between learner autonomy and these Vietnamese students' English language proficiency?	Questionnaire	177 students
Phase two	- conducting an experiment to train students in metacognitive skills - examining relationship between the metacognitive training and learner autonomy	1. Does training in metacognition lead to improved written English? 2. Will improvements in written English be maintained? 3. Does training in metacognition techniques result in higher learner autonomy? 4. Does metacognitive training in the context of English learning and teaching result in the transfer of metacognitive skills to other areas of language learning?	Metacognitive training package; Writing tests (pre-, post-, delayed) Questionnaires (pre + post) Learner diaries; Classroom observations; Interviews (one-on-one, group, email)	94 students: 1 experiment group of 33 students, 2 control classes of 25 students each; 2 teachers; 1 researcher; 11 students from experimental group

The qualitative data in the pilot study came from interviews and learning logs.

Interviews were employed to add more items to the questionnaire, the main research instrument of the study. Learning logs were used to record the number of activities and the amount of time students devoted to learning English both in class and during self study outside the classroom context. The purpose of the learning logs was to explore autonomous behaviours demonstrated by the students. Two different forms of the learning logs were used, out-of-class and in-class, and learners were asked to keep the logs for two weeks. For the out-of-class logs, learners were requested to write about any English-learning related activities they performed outside the classroom. For the in-class logs, learners were required to write about writing-related activities they undertook in the two writing lessons during a two-week period. The learning logs played an important role in revising the questionnaire. After students' learning logs were analysed, a few more questions were added.

The qualitative data of Phase Two consisted of interviews, diary entries, and classroom observations. The aims of using interviews in this phase included obtaining information from both student and teacher perspectives on the application, the effectiveness, and the transfer of metacognitive strategies in students' English learning. Two different forms of interview were conducted including one-on-one interviews, group interviews as well as interviews via email. In Phase Two diary entries about metacognitive training sessions were provided by learners in the experimental group on a voluntary basis. In their diary entries learners were asked to write about a metacognitive strategy learned at the session; the application of the metacognitive strategy in writing and in other language areas of listening, speaking and reading; and their reflections/comments on the strategy. The purpose of the diary was 1) to gather information on the use of metacognitive skills by students in their writing and the transfer of metacognitive strategies to other skills such as speaking, listening, and reading; 2) to raise learners' awareness about metacognitive strategies; and 3) to provide immediate feedback for the researcher to adjust the training sessions that would help learners better comprehend and apply the metacognitive skills that were taught. The diary entries were submitted to the researcher on a weekly basis for feedback. The students were provided with the researcher's comments on their reflections. Their attention was drawn to the most common grammatical mistakes. This was done to encourage the students to submit their diary entries more frequently. The diary entries were used to explain or back up the data originating from the questionnaires. Also in Phase Two, classroom observations were conducted to capture any differences in the way learners in the experimental group applied metacognitive skills in writing, and the way the two teacher participants conducted their writing lessons in the two control classes. The observations were carried out across all three classes in the first, third, and the last weeks of the course.

Developing and validating research instruments: the two questionnaires

Based on the two main elements of learner autonomy (self-initiation and self-regulation) outlined in the operationalised definition of learner autonomy, two questionnaires (self-initiation and self-regulation) were developed and used across the pilot study and the two phases of the research. The two sub-elements of self-initiation include reasons for learning and making efforts to learn. In this study reasons for learning were associated with motivation. Therefore, the self-initiation questionnaire covered questions categorised into 1) learners' motivation to learn English and 2) activities learners initiate to improve their English. The first version of the self-initiation questionnaire (Table 2) was composed of 91 questions. Questions in the motivation section were based on motivation types in Gao, Zhao, Cheng, and Zhou (2004, 2007). Questions in the activities section were designed based on activities developed by Spratt et al. (2002). The original version of the self-regulation questionnaire (Table 3) had 55 questions falling into three main categories: *planning*, *monitoring*, and *evaluating*. The questionnaire asked students about steps they undertook before, during, and after writing.

Table 2: Summary of self-initiation questionnaire

Sections	Subsections	Questions	Number of questions
Activities to improve English	Learning English outside the classroom	Q1-Q13	13
	Using English outside the classroom	Q14-Q37	24
	Overt learning behaviours	Q38-Q54	17
	Covert learning behaviours	Q55-Q63	9
Motivation	Immediate achievement	Q64-Q66	3
	Information medium	Q67-Q68	2
	Individual development	Q69-Q75	7
	Social responsibility	Q76-Q78	3
	Going abroad	Q79-Q81	3
	Intrinsic interest	Q82-Q87	6
	Learning situation	Q88-Q91	4
		Total	91

Since questionnaires became the main research instrument of the study, they were carefully developed and validated.

Questionnaire development process

Questionnaire items came from three main sources including (1) adaptations of existing questionnaires; (2) original design; and (3) results of the pilot study. The questionnaires were developed in three steps: (1) piloting, (2) revising, (3) trying out and incorporating that feedback into a final version of the questionnaire. The pilot study played an important role in revising the questionnaire items and the way the questionnaire should be carried out. In fact, it was expected that planning

activities for a writing task, such as doing concept mapping of the topic, organising ideas or thinking of the possible vocabulary, would result in students writing better pieces than the ones for which no planning activities were conducted. To put it another way, it was expected that there would be a positive relationship between learner autonomy and students' language proficiency. However, the pilot study showed that there was a negative relationship between planning and language proficiency. The unexpected results encouraged the researcher to identify the weaknesses of the questionnaire. It was found that although the pilot questionnaire asked subjects about specific learning behaviours, they were not given any task to perform. Therefore, it was difficult for them to produce valid answers about specific planning behaviours. The pilot study provided the researcher with hands-on experience of how to design a questionnaire that would work with these participants. In the second step, the revised questionnaire went through several rounds of revisions. All items were re-categorised. Section One (Activities to improve English) in the original questionnaire was re-classified as *Out-of-class activities* and *In-class activities*. This section was further broken down to include *Learning English* and *Using English* as well as *Covert learning* and *Overt learning*. In addition to the categorisation of questionnaire items, care was taken to ensure that each item covered only one feature. For example, the following question was broken down into two separate items:

Original item: I consider assessment criteria set by teachers or comments made by other people to judge how well I have written the paper.

New item 1: I consider assessment criteria set by teachers to judge how well I have written the paper.

New item 2: I consider comments made by other people to judge how well I have written the paper.

The third step was to try out the revised questionnaire after it was translated into Vietnamese, randomised, and proofread. In the questionnaire distributed to the participants all headings such as *Out-of-class activities*, *In-class activities*, *Learning English*, *Using English* etc. were removed and all the items were randomised so that the items that had been under each heading were distributed throughout the questionnaire. The avoidance of a large number of related items occurring together would improve consistency of responses to related questions. The think-aloud protocol, which is a process where participants report while doing a task, was then used for receiving feedback from students trying out the questionnaire because the researcher was interested to know which item(s) of the questionnaire did not work, why it (they) did not work, and how long it would take to complete the questionnaire as well as the writing task students are required to perform prior to answering the questionnaires. Both the students and the researcher went through each question and the students were asked to tell the

researcher about the questions that did not make sense. The questions were then revised according to the suggestions made by the students. The second trial was conducted and followed the same procedures. All comments and suggestions were then incorporated into a final and polished version of the questionnaire.

Table 3: Summary of self-regulation questionnaire

Sections	Subsections	Questions	Number of questions
Before writing	Goal setting	Q1-Q3	3
	Pre-writing	Q4-Q6	3
	Task knowledge	Q7-Q12	6
	World knowledge	Q13-Q17	5
	Rhetorical knowledge	Q18-Q21	4
	Linguistic knowledge	Q22-Q24	3
	Audience knowledge	Q25-Q26	2
	Self knowledge	Q27-Q28	2
During writing	Monitoring task progress	Q29-Q31	3
	Monitoring strategies	Q32-Q35	4
	Monitoring language problems	Q36-Q40	5
	Monitoring feeling	Q41	1
	Monitoring task concentration	Q42	1
	Monitoring knowledge	Q43	1
	Monitoring task performance	Q44	1
After writing	Evaluation of goal achievement	Q45-Q46	2
	Evaluation of strategies	Q47-Q48	2
	Evaluation of resources use	Q49	1
	Evaluation of assessment criteria	Q50-Q52	3
	Evaluation of mistakes	Q53-Q54	2
	Evaluation of self-modifying	Q55	1
		Total	54

Questionnaire validation

Questionnaire validation was conducted to ensure internal consistency by using Cronbach's alpha, which involves the provision of a precise internal consistency estimate. If the items are scored as continuous variables, the alpha provides a coefficient to estimate consistency of scores on an instrument. The questionnaire validation consisted of two parts. First, exploratory and confirmatory factor analyses were conducted on the 146 questions from the two original questionnaires (Table 2 and Table 3) in order to check the construct validity of the subconstructs of learner autonomy. Item analysis was also performed to obtain the internal consistency reliability of each subconstruct and to determine which items were problematic. The purpose of this was to produce a better version of the questionnaire that had fewer items but covered similar constructs with satisfactory levels of internal consistency reliability, while retaining as much of the original information as possible.

In the original version of the self-initiation questionnaire there were three constructs: *Out-of-class activities*, with sub-constructs *Learning English outside*

class (13 questions) and *Using English outside class* (24 questions); *In-class activities,* with sub-constructs of *Overt language learning behaviours* (17 questions) and *Covert language learning behaviours* (9 questions); and *Motivation,* with sub-constructs of *Instrumental motivation* (12 questions), *Integrative motivation* (12 questions) and *Situational motivation* (4 questions). In order to obtain the most reliable data, factor analysis was conducted with each construct by looking at those variables that clustered together in a meaningful way. This was done, following Field (2005), by finding variables that correlated highly with a group of other variables but did not correlate with variables outside that group. The factor loading in the factor analysis provided the relative contribution that a variable made to the factor. Immediately after the variables (items in the questionnaire) under each factor were formed, the reliability analysis of those items was run to ensure no item would cause a substantial decrease in alpha. Items whose values of "alpha if item deleted" were higher than .8 were deleted. As a result of this process, under *Out-of-class activities*, in each of the sub-constructs of *Learning English outside classroom* and *Using English outside classroom,* six questions remained with alphas of .756 and .815 respectively. For *In-class activities*, each sub-construct of *Overt language learning behaviours* and *Covert language learning behaviours* had five questions left. Their alphas were .813 and .850 respectively. As for the motivation construct, the factor loading formed three sub-constructs of *Individual development, Intrinsic interest* and *Going abroad*. Each of these sub-constructs had three questions and their alphas were .673, .774, and .783 respectively.

Tables 4, 5, 6, and 7 provide a detailed account of the steps the researcher undertook to ensure the reliability of the 13 items categorised as *Learning English outside the classroom*.

The alpha of the 13 questions is .809 (Table 4). However, if Q12 is deleted, the alpha will be .814 (Table 5).

The fewer items but the higher alpha would imply a greater level of reliability for the questionnaire. Therefore, the researcher gradually dropped Q12, Q4, Q3, Q11, Q7 and Q10, which caused considerable decreases in the alpha of the items being examined (Table 6). Finally, six questions (Q1, Q2, Q5, Q6, Q8 and Q13) remained and their alpha is .756 (Table 7).

Table 4: Reliability statistics of 13 items of learning English outside class

Cronbach's Alpha	Cronbach's Alpha based on standardised items	Number of items
.809	.816	13

Table 5: Item-total statistics of 13 items of learning English outside class

	Scale mean if item deleted	Scale variance if item deleted	Cronbach's Alpha if item deleted
Q1	21.72	53.421	.796
Q2	21.72	53.433	.796
Q3	20.72	54.603	.803
Q4	21.77	53.688	.805
Q5	21.66	52.026	.794
Q6	21.54	50.881	.788
Q7	22.31	53.349	.802
Q8	21.77	50.308	.788
Q9	22.34	50.765	.785
Q10	21.90	50.663	.792
Q11	22.26	49.808	.795
Q12	21.62	52.038	.814
Q13	21.59	51.261	.791

Table 6: Gradual dropping items causing substantial decrease in alpha

Cronbach's Alpha	Cronbach's Alpha based on standardised items	Number of items	Items to be deleted
.809	.816	13	
.814	.817	12	Q12
.815	.819	11	Q4
.811	.816	10	Q3
.806	.809	9	Q11
.795	.798	8	Q7
.779	.782	7	Q10
.756	.761	6	

Table 7: Reliability statistics of 6 items of learning English outside class

Cronbach's Alpha	Cronbach's Alpha based on standardised items	Number of items
.756	.761	6

In the initial version of the self-regulation questionnaire there were three constructs, namely *planning*, *monitoring* and *evaluating*. Under each construct there were originally many other sub-constructs with one or more questions. However, for the *planning* construct, the factor analysis came up with two factors of *general planning* and *task-specific planning*. The number of items of each factor was four and the alphas were .751 and .786 respectively (Table 8). After the factor analysis for the *monitoring* construct was performed, it was very difficult to decide which items to retain because the factor loading indicated items that did not closely match the sub-constructs in the original questionnaire. The researcher therefore decided to conduct a reliability analysis for all items in the *monitoring* construct in order to exclude those items whose 'alpha if deleted' was the highest each time. Reliability analysis was carried out until there were seven items left (out of 16 items in the original questionnaire). The alpha of the *monitoring* construct was .838 (Table 8). As for the *evaluating* construct, the exploratory analysis showed that all 11 questions in the original questionnaire loaded on one

factor. The researcher used the same strategy of reliability analysis as was applied in the *monitoring* construct to eliminate items in the *evaluating* section. As a result, seven items were kept and their alpha was .873 (Table 8).

Table 8: Reliability statistics of the final questionnaire

Learner autonomy constructs			Cronbach's Alpha	Number of items
Self-initiation	Reasons for learning English	Individual development	.673	3
		Intrinsic interest	.774	3
		Going abroad	.783	3
	Making an effort to learn English	Learning English out of class	.756	6
		Using English out of class	.815	6
		Overt language learning in class	.813	5
		Covert language learning in class	.850	5
Self-regulation		General planning	.751	4
		Task-specific planning	.786	4
		Monitoring	.838	7
		Evaluating	.873	7
			Total	53

The resulting new version of the questionnaire (Table 8), which combines both the self-initiation and the self-regulation questionnaires (Table 2 and Table 3), did not aim to include all possible items related to learner autonomy, but only those showing high internal validity. This resulting questionnaire had only 53 items categorised under five distinct elements: *reasons for learning English, making an effort to learn English, planning, monitoring* and *evaluating*. Each item under each category correlates with the other items. To put it another way, compared with the original number of items in the self-initiation and self-regulation questionnaires, the number of items in the resulting new version was smaller but its internal consistency reliability was greater.

Data analysis

Throughout the study, the operationally defined concept of learner autonomy served as a framework on which analyses of learner autonomy were conducted. This study drew on two types of triangulation among the seven different categories of triangulation summarised by Brown (2001, p. 228). These included data triangulation and methodological triangulation. The former entails the use of multiple sources of data to look into the phenomenon from different perspectives. The latter requires the employment of several data collection procedures. When making interpretations of a phenomenon or drawing conclusions about a particular finding, additional sources of information were referred to. Attempts were made to avoid relying solely on the results supplied by the scores of learners' writing tests or the questionnaires.

Pilot study

The research instruments, the interviews, the learning logs and the questionnaire provided the researcher with different sources of data from which several new categories and questions were generated. The newly created questions were of value to the questionnaire, the main research instrument for the following stages of the study. While several supplementary questions were added to the questionnaire, some items were deleted because they were either too general or irrelevant to Vietnamese educational practice. On the basis of the information collected from the interviews and the learning logs, the questionnaires were revised before being used in Phase One and Phase Two of the study.

Phase One

Research questions one and two (Table 1) were addressed mainly by observing the mean scores of the major learner constructs and those of out-of-class and in-class activities respectively. To answer research question three (Table 1), correlation coefficients between learner autonomy constructs and EFL proficiency measures were analysed. The data from the resulting new questionnaire were submitted to Pearson's correlation analysis, which examined the relationship between learner autonomy and language proficiency. Being the sole research instrument in Phase One, questionnaires seemed to be useful for collecting data on the degree to which Vietnamese learners were autonomous, the activities they initiated, and the relationship between learner autonomy and language proficiency.

Phase Two

To answer each research question (Table 1), both sets of data were analysed. The quantitative data coming from the questionnaires (pre- and post-) and tests (pre-, post- and delayed) were submitted to SPSS 16.0 for analysis. Each research question was answered by observing the mean scores of the writing tests which comprised four components: content, organisation, language, and grammatical accuracy, and the mean scores of learner autonomy constructs including reasons for learning English, making an effort to learn English, planning, monitoring, and evaluating. Descriptive analysis, one-way ANOVA, one-way repeated-measures ANOVA, and post hoc tests were performed to compare means and to detect the within- and cross-group differences. The qualitative data originated from student diaries, their written texts, the researcher's field notes of classroom observations, and the interviews with students and teachers. The interviews were transcribed, coded and analysed. The qualitative information was fully exploited to interpret or to back up the findings. For example, it was found that the experimental group outperformed the two control groups and maintained improvements in written English after they had received the metacognitive training in planning, monitoring, and evaluating (Nguyen & Gu, forthcoming). This finding was backed up by group interviews in which volunteer students said in the future they would continue to use the metacognitive skills of planning, monitoring, and evaluating. Another instance is that the experimental group was shown by the questionnaire to have practised self-regulatory skills a little more than control classes (Nguyen & Gu, forthcoming). The group interview reinforced the

findings about students' improved self-regulation. They said the way they approached a writing task had changed since they embarked on the training. They tried to organise their essays and to think about vocabulary to be used. The group interview not only supported the results of the quantitative data from questionnaires but also provided insights into the aspects of each metacognitive skill that students exercised.

Limitations

Some limitations of this study include the self-report nature of the questionnaires and the number of times the participants completed the questionnaires. Firstly, the quantitative data came from questionnaires where learners self-reported their learning activities. It was not certain whether they actually performed self-initiated and self-regulated learning activities as they claimed in the questionnaires. Future research could use close observation and if possible video-tape learners' learning behaviours exhibited both inside and outside the classroom. Secondly, the longitudinal characteristic of the study could have triggered improved learner autonomy among the learners in the experimental group. They could have become more autonomous because they had answered the questionnaires twice and had realised what would be good for them, not because they benefited from the metacognitive training.

Conclusion

The study indicated that learner autonomy could be explored thoroughly and measured rigorously and reliably by carefully following three main principles. These principles entail having an operationalised definition of learner autonomy, looking at it from different points of view using both qualitative and quantitative research approaches to collecting data, and carefully developing and validating data collection instruments. Firstly, the operationalised definition made measuring learner autonomy possible. Self-initiation and self-regulation, the two main elements of learner autonomy, were assessed separately. Secondly, both quantitative and qualitative methods contributed to a comprehensive analysis of the issues explored in the study. Quantitative research methods were employed to investigate the relationship between learner autonomy and language learning results as well as between the metacognitive training and learner autonomy. The qualitative process entailed collecting opinions and feedback from learners and teachers about the metacognitive training through interviews, learners' diary entries, and classroom observations. Additionally, learners' learning logs and interviews were also employed in the trialling phase to develop the questionnaire which served as the main instrument for this research project. In any report of a study, it is important to mention what was done in the pilot studies because they affect the validity of the investigative tools. Of great importance is the fact that the qualitative data played a significant role in interpreting and reinforcing the

findings from quantitative data. Thirdly, questionnaire validation played a prominent part in yielding reliable data and the validation could be claimed to be the strength of the study. In summary, the application of the three principles discussed in this article seemed to make it possible to measure learner autonomy rigorously.

Notes

A detailed report of Phase one was published in Nguyen (2008).

A detailed report of Phase two will be published in Nguyen and Gu (forthcoming).

This article reports on the methodology used in the author's PhD study.

References

Benson, P. (2001). *Teaching and researching autonomy in language learning*. Harlow, England: Longman.

Brown, J. D. (2001). *Using surveys in language programs*. New York: Cambridge University Press.

Champagne, M.-F., Clayton, T., Dimmitt, N., Laszewski, M., Savage, W., Shaw, J., et al. (2001). The assessment of learner autonomy and language learning. *AILA Review, 15*, 45-54.

Chan, V. (2001). Readiness for learner autonomy: What do our learners tell us? *Teaching in Higher Education, 6*, 505-518.

Chan, V. (2003). Autonomous language learning: The teachers' perspectives. *Teaching in Higher Education, 8*, 33-54.

Cunningham, J. & Carlton, W. (2003). Collaborative newsletters. In A. Barfield & M. Nix (Eds.), *Autonomy you ask!* (pp. 113-128). Tokyo: Learner Development Special Interest Group of the Japan Association for Language Teaching.

Dam, L. (1995). *Learner autonomy: From theory to classroom practice* (Vol. 3). Dublin: Authentik Language Learning Resources.

Dam, L., & Legenhausen, L. (1996). The acquisition of vocabulary in an autonomous learning environment: The first months of beginning English. In R. Pemberton, E. S. L. Li, W. W. F. Or, & H. D. Pierson (Eds.), *Taking control: Autonomy in language learning* (pp. 265-280). Hong Kong: Hong Kong University Press.

Field, A. (2005). *Discovering statistics using SPSS*. London: SAGE.

Gao, Y., Zhao, Y., Cheng, Y., & Y. Zhou (2004). Motivation types of Chinese university undergraduates. *Asian Journal of English Language Teaching, 14*, 45-64.

Gao, Y., Zhao, Y., Cheng, Y., & Y. Zhou (2007). Relationship between English learning motivation types and self-identity changes among Chinese students. *TESOL Quarterly, 41*, 133-155.

Holec, H. (1981). *Autonomy and foreign language learning*. Oxford: Pergamon Press.

Kohonen, V. (2000). Exploring the educational possibilities of the "dossier": Some suggestions for developing the pedagogic function of the European language portfolio. In D. Nunan (Ed.), *Collaborative language learning and teaching* (pp.14-39). Cambridge: Cambridge University Press.

Kohonen, V. (2001). Developing the European language portfolio as a pedagogical instrument for advancing student autonomy. In L. Karlsson, F. Kjisik, & J. Nordlund (Eds.), *All together now* (pp. 20-44). Helsinki: University of Helsinki Language Centre.

Mizuki, P. (2003). Metacognitive strategies, reflection, and autonomy in the classroom. In A. Barfield & M. Nix (Eds.), *Autonomy you ask!* (pp. 143-156). Tokyo: Learner Development Special Interest Group of the Japan Association for Language Teaching.

Mynard, J. (2006). Measuring learner autonomy: Can it be done? *Independence – Newsletter of the IATEFL Learner Autonomy Special Interest Group, 37*, 3-6.

Nachi, H. E. (2003). Self-assessment and learner strategy training in a coordinated program: Using student and teacher feedback to inform curriculum design. In A. Barfield (Ed.), *Autonomy you ask!* (pp. 157-174). Tokyo: Learner Development Special Interest Group of the Japan Association for Language Teaching.

Natri, T. (2007). Active learnership in continuous self- and peer-evaluation. In A. Barfield & S. H. Brown (Eds.), *Reconstructing autonomy in language education: Inquiry and innovation* (pp. 108-120). New York: Palgrave Macmillan.

Nicoll, H. (2007). Seeking autonomy in a lecture course. In A. Barfield & S. H. Brown (Eds.), *Reconstructing learner autonomy in language education: Inquiry and innovation* (pp. 120-130). New York: Palgrave Macmillan.

Nguyen, T. C. L. (2008). Learner autonomy and EFL proficiency: A Vietnamese perspective. *Asian Journal of English Language Teaching, 18*, 67-87.

Nguyen, T.C.L & Gu, P. Y. (forthcoming). Strategy-based instruction: A learner-focused approach to developing learner autonomy. *Language Teaching Research*.

Nunes, A. (2004). Portfolios in the EFL classroom: Disclosing and informed practice. *ELT Journal, 58*, 327-335.

Pearson, N. (2004). The idiosyncrasies of out-of-class language learning: A study of mainland Chinese students studying English at tertiary level in New Zealand. *Proceedings of the Independent Learning Conference 2003*. Retrieved from at http://inle.org/ILA/ila03/ila03-pearson.pdf

Pickard, N. (1995). Out-of class language learning strategies: Three case studies. *Language Learning Journal, 12*, 35-37.

Pickard, N. (1996). Out-of-class language learning strategies. *ELT Journal, 50*, 150-159.

Rao, Z. (2005). Developing learner autonomy through portfolios. *Teachers' Edition, 19*, 20-24.

Sert, N. (2006). EFL student teachers' learning autonomy. *The Asian EFL Journal Quarterly, 8*(2), 180-201.

Shimo, E. (2003). Learners' perceptions of portfolio assessment and autonomous learning. In A. Barfield & M. Nix (Eds.), *Autonomy you ask!* (pp. 175-188). Tokyo: The Learner Development Special Interest Group of the Japan Association for Language Teaching.

Sinclair, B. (1999). Wrestling with a jelly: The evaluation of learner autonomy. In B. J. Morrison (Ed.), *Experiments: Evaluation in self-access language learning* (pp. 95-109). Hong Kong: Hong Kong Association for Self-Access Learning and Development.

Smith, R. C. (2001). Group work for autonomy in Asia: Insights from teacher-research. *AILA Review, 15*, 70-81.

Spratt, M., Humphreys, G., & Chan, V. (2002). Autonomy and motivation: Which comes first? *Language Teaching Research, 6*, 245-266.

Stephenson, J., & Kohyama, M. (2003). Turning freshmen into autonomy through student-directed language learning projects. In A. Barfield & M. Nix (Eds.), *Autonomy you ask!* (pp. 101-112). Tokyo: Learner Development Special Interest Group of the Japan Association for Language Teaching.

Tagaki, A. (2003). Learner autonomy and motivation in a cooperative learning class. In A. Barfield & M. Nix (Eds.), *Autonomy you ask!* (pp. 129-142). Tokyo: Learner Development Special Interest Group of the Japan Association for Language Teaching.

Vickers, C. H., & Ene, E. (2006). Grammatical accuracy and learner autonomy in advanced writing. *ELT Journal, 60*, 109-116.

BOOK REVIEWS

Hall, C. J., Smith, P. H., Wicaksono, R. (2011). *Mapping applied linguistics.* London: Routledge. ISBN 978-0-415-55913 (pbk). 411 pp. $67.00.

Descriptions of our field have come a long way in the quarter century since the *Longman dictionary of applied linguistics* (Richards, Platt, & Weber 1985). In fourteen chapters, divided into three parts, the metaphor of mapping binds together topics as wide-ranging as language loss, language policy and forensic linguistics. The authors, who describe themselves as coming from the north of England (Hall and Wicaksono) and the U.S.A. (Smith) promise to be "innovative and wide-ranging" (p. i).

A question starts the 20-page introduction: "Why do we use different languages?" The reader is drawn in through 10 intriguing "ways we're led astray in language and applied linguistics" (p. 4), each labelled a Dead End. As an example, Dead End 6 refutes the belief that "the way groups use their language reflects their intelligence" (p. 12). Less colloquially expressed, the authors assure readers that they do not believe in a discipline that "fosters unbridgeable polarities and intellectual territoriality" (pp 19-20), a point they illustrate with a Wikipedia map of the world from an unusual angle. Following this introduction, the four chapters of Part A concern language in everyday use. Here the authors discuss language variation, key populations, discourse analysis and language planning. A section about naming populations caught my attention, following a recent discussion I was part of in which people compared the words used to describe 'ethnicity' on official forms in different English-speaking countries. Here the topic was taken further, including viewpoints about the naming preferences of people with disabilities such as deafness and blindness. This appears to be a minefield for anyone who aims to respect the wishes of specific groups.

Part B, which is about learning and education, opens with Chapter 6 on literacy. Here the examples stretch back 6,000 years and reach to such recent developments as emoticons. New Zealand readers will be interested to find more than one reference to Māori language issues. In Chapter 7 these are in relation to language and education, and in Chapter 11 (of Part C) the work of the Māori Language Commission is described in relation to lexicography.

Part C moves to Expert Uses, with a chapter each on translation, lexicography, forensic linguistics and language pathology. The list of linguistic and applied linguistic knowledge required by a translator underlines the highly sophisticated nature of that profession. As one illustration from the list of thirteen, the professional needs to know the customary pragmatic routines through which "communicative intentions and effects" are mapped onto linguistic expressions (p. 232).

Finally, Chapter 14 looks ahead. A section on the profession in practice includes links between professionals and those they must liaise with. Comments are included

from applied linguists (former students of the authors?) practising in diverse parts of the world, including one researcher in Gaza, and another in China.

Occasionally, in the end of chapter references, there was a sense that Routledge publications took precedence over others that might seem equally relevant, such as Thornbury's introduction to discourse analysis from Macmillan (not mentioned) published the same year as Gee's from Routledge which is listed. Also, although the historic but still popular *Longman dictionary* mentioned above is not listed, many of its entries are included. Of course authors have to draw the line somewhere. That quibble aside, what features might encourage readers to add this book to their collection or, even to list it as a prescribed text for classes? A strength is its readability, notwithstanding the warning in the introduction that there is to be no "dumb[ing] down" (p. xxi). Accessibility to the topics is achieved in a number of ways, including many diagrammatic representations of the points being made. As an example, Figure 14.4 illustrates "Flows of responsibility for standards and codes of practice in applied linguistics" (p.337).

Finding one's way around the book is helped by several devices, including the clearly signalled chapter headings and sub-headings and a cross-referenced glossary. In addition, definitions appear in the margins from time to time. Then there are the three indices: a general one of 16 pages, another listing languages and a third for countries mentioned. The references are in two parts: one list follows each chapter and the other is in a final bibliography; however, it wasn't altogether clear to me why some titles appear in both, while others are in one place only.

Cartoons also make some points quite pithily. One of my favourites showed Ms Potter, a first grade teacher, explaining to the children that "writing is just like texting except that you have to use all of the letters in each word" (p. 325). All in all, the book fulfils its "wide-ranging" promise.

My impression is that the book will appeal to a more groups than the three suggested by the authors: students of applied linguistics, practising professionals (speech therapists, lexicographers and others) and general readers. The world of teacher educators will also find plenty here to feed their lectures and workshops. The activities section that concludes each chapter is aimed at the student readership, but the range of print and electronic sources shows how applied linguistics is of interest to wider groups. There is a companion website.

To do the book justice, this review would need to exceed the word limit. It comes with a recommendation both for the obedient reader who follows the authors' suggestion that it should "be read most profitably in sequence" (p. xxiii), and to those who are interested in dipping into topics beyond their own specialist areas for some random, but fascinating details. Even for those whose shelves already have many

titles promising an overview, handbook, summary or encyclopaedia of our field, this latest volume is still a worthwhile addition.

References

Richards, J. C., Platt, J., and Weber, H. (1985). *Longman dictionary of applied linguistics.* Harlow, Essex: Longman.

Thornbury, Scott. (2005). *Beyond the sentence: Introducing discourse analysis.* Oxford: Macmillan.

MARILYN LEWIS, Honorary Research Fellow, The University of Auckland

Rom, M., & Orly B. 2011. *Feminism, family, and identity in Israel: Women's marital names*. New York: Palgrave Macmillan. ISBN-13:978-0-230-10015-2 (hbk). 250pp .

Until comparatively recently (1996), there were very few options for Israeli women upon marriage. The 1956 Law of Names specified very clear standards for the correspondence between a woman's name and her marital status. Laws do not change in and of themselves, of course; social and attitudinal change in Israel laid the ground for a liberalisation of the options available to women, and following the passing of the revised 1996 law, women and men in Israeli society were able to exercise more creativity in choosing a name to bear. This book uses interviews with Israeli women to document the choices made by women since then, looking not only at the various options for naming practices, but also exploring through discourse analysis the kinds of social and psychological factors that these choices are embedded in. The analysis is broadly social constructionist, drawing on the notion of communities of practice to describe different family and religious community norms, and poststructuralist, attempting to move beyond binary oppositions.

The authors conducted interviews lasting around one and a half hours with 42 Israeli Jewish women who had responded to a call for participants interested in talking about their naming choices. They were swamped with responses almost immediately – clearly, naming is something that is of high salience in Israel at this point in time. The options that women in their survey chose were (i) to retain their birth family name, (ii) to switch to their partner's family name after marriage, (iii) to create a new name with their partner and both switch; (iv) to hyphenate their own and their partner's name and switch to that, and (v) to hyphenate and both switch. (Because marriage remains under the aegis of the Orthodox branch of the church, not all women interviewed had undertaken the full marriage ceremony, hence the use of the term 'partner' here.)

The process of changing one's name, even after the 1996 change in law, is arduous and highly bureaucratic in Israel. The women and men who see it through are therefore highly motivated and have discussed the matter and its practical and ideological implications at length with their partner, their families, and friends. Some have been abused, threatened and discouraged in all manner of ways. Some have received tremendous and unconditional support from the partners. We hear their stories in this volume.

The book begins with a substantial Preface, and even more substantial Introduction that outlines issues and previous research on naming that frames the current volume. Chapter 1 'Local context of identity formation' establishes the sociopolitical context of naming in Israel and briefly outlines the research methods. The data collection and data handling is the subject of a much longer appendix (Appendix 2) which explains how they chose pseudonyms for all the women interviewed. Since names are the

subject of the research, these had to be chosen with care to not only anonymise the interviewees, but to retain the social, ethnic, and religious associations of the interviewee's true name. A panel of six judges were given the women's and men's names before and after marriage and were asked to create names which were similar. Lists of proposed names were swapped between the judges for them to evaluate until there was convergence on the best pseudonyms (the discussion of methods in the appendix is very helpful for anyone who might embark on a similar project elsewhere in the future).

Chapter 2 'Naming identities: Politics of identity' returns in more depth to the theoretical issues and introduces notions of agency, performativity, social positioning and communities of practice. The authors outline their commitment to a pluralistic notion of identity. Chapter 3 'Israeli ambivalence and gender relations' covers gender relations and feminism in Israel and the significance of the family within national and religious discourses in Israel ('familism'). Chapter 4 'The appropriate name' begins to explore in detail the choices being made by Israelis when they focus on names. Ethnic distinctions are important – whether a name sounds Ashkenazi (European) or Mizrah'i (Middle Eastern), whether it sounds Israeli or diasporic. Changes to a name, especially where both partners change their name, tend to favour creation of what are considered pleasant or auspicious Hebraicised names. As this indicates, aesthetic considerations are also part of the discourses surrounding naming. Some names are deemed to simply sound 'nicer', or 'better' – a metaphor invoked repeatedly was 'the name of a pilot'. In May 1948, army officers were all ordered to Hebraicise their names, and as Rom and Benjamins explain, this created a potent layering of significations for Hebraic names. Not only were the army officers celebrated as heroes in the forging of the nation-state, they were recruited from the societal elite, thereby creating a nationalist and class-based ideal for white men that is tightly bound to Hebraicised names, even ones such as *Yogev*, meaning 'peasant; agricultural worker', which might be expected to have relatively little social caché on denotational grounds.

Chapter 5 'Getting more out of it: Identity positioning through the name' explores ideals of femininity and domestic roles. Chapter 6 'Time and space dimensions of self-naming' considers social resistance and obstacles to naming practices. Chapter 7 'Name in relations' looks in more detail at the repercussions and considerations surrounding naming choices in intimate relationships. Chapter 8 'Discussion' is really a conclusion, drawing together theory and practice from the previous chapters.

The book is marred by numerous typos, some of which are things the spell-checker of the Palgrave copy-editor should have picked up, but also include misattribution of quotes, and misspelling of names in the text (or worse still in the text and references – Etienne Wenger's work on communities of practice is attributed to E. *Wagner*). As a whole, the book provides a fascinating perspective on the response of Jewish Israeli

women (and some men) to a loosening up of the social conventions surrounding naming, and the editors have chosen illustrative extracts with skill and care.

MIRIAM MEYERHOFF, The University of Auckland

Archer, D. & Grundy, P. (2011). *The pragmatics reader.* Abingdon, UK: Routledge. ISBN 978-0-415-546607. 529pp.

There is growing interest in recent years in the study of pragmatics, and a number of new branches of the field have emerged over the past 20 years. Pragmatics now ranges from applied to purely theoretical interests, and there is a tendency for students and even scholars to work in their own branch of this field without much reference to or awareness of the other branches. There is therefore a need for literature that informs newcomers to the field and advanced students at undergraduate and post graduate level about its breadth and history.

Archer and Grundy's comprehensive collection of articles sets out to do just this. The stated aim is to showcase current ideas and exemplary research, and also to reflect the international nature of this field. The book is divided into 10 sections, each representing one branch of pragmatics: linguistic pragmatics, post-Gricean pragmatics, indexicality, historical pragmatics, politeness, face and impoliteness, cross-cultural and intercultural pragmatics, pragmatics and conversation development and impairment, and pragmaticians on pragmatics. In each section are up to six articles outlining the characteristic thinking of key scholars in each of the branches. Names well known in the field are well represented including Grice, Brown and Levinson, House, Kasper, Sperber and Wilson and Verschueren. Each article is accompanied by a variety of well thought out pre-reading and post-reading questions to challenge the most able of students, together with a bibliography to inform further reading.

As a practitioner who has read a number of articles in the areas of cross cultural pragmatics, linguistic pragmatics and politeness, face and impoliteness and carried out action research in some of these areas, this book gave me a new perspective on the field of pragmatics. It was useful and informative to find out where familiar branches fitted in thematically and historically with others, and to learn about new fields of relevance to my practice of which I had been previously unaware, such as intercultural pragmatics. The introductions were accessible to intelligent novices in each branch, and the questions helped me to challenge myself in the more controversial areas. It was also good to see the original seminal articles in important areas like Gricean pragmatics and politeness theory, where I have read a variety of secondary literature.

There are some reservations. Some of the articles were very challenging for someone completely new to the field or the branch, and required some previous study of philosophical concepts or the specialist terms used. However the book doesn't have to be read cover to cover. The introductions to a range of sections could be read for an overview, and only some branches studied in depth. The book provides excellent and challenging questions to help the reader gain a notion of where these branches fit

into the field. With the above reservations I would recommend this book to the dedicated student of any branch of pragmatics who wishes to read seminal articles by original thinkers in an accessible collection, or wants to know more about the range of pragmatic study, as well as to anyone who wishes to get an sense of the field and is prepared to put in some concentrated reading and thinking.

HEATHER DENNY, AUT University

Wei, L. (Ed.). (2011) *The Routledge applied linguistics reader*. London: Routledge. ISBN 978 0 415 56620 9. 530pp.

What is understood by the term *applied linguistics* and what the term encompasses has changed considerably over the years, and recently several state-of-the-art handbooks focusing on different aspects of the discipline have emerged (e.g. Davies & Elder, 2004). The editor of this book states that by the end of the twentieth century applied linguistics had developed into a diverse, interdisciplinary field, transcending known territory of language classrooms and structural linguistics and entering new areas such as multilingual and professional contexts, new media, and the effects of globalization. Applied linguists also try to reconceptualise the field as part of social science (e.g. Sealey & Carter, 2004), raising awareness that failure or success in language competence is not only a linguistic or psychological issue, but one that has to be viewed in a broader context of, for example, socio-political or socio-cultural settings.

For applied linguistics to have a bigger impact, the editor asks for a critical perspective in applied linguistics research to establish critical enquiry with a focus on language in the real world. This aim is also at the centre of this Reader.
The volume is an interesting collection of chapters addressing issues and topics relevant for applied linguistics and language learning. The 26 individual chapters are all reprinted articles selected from journals and edited volumes, and range from 1995 to 2009 publications. The chapters are streamed into four themes, each briefly introduced by the editor. At the end of each segment there are “Notes for students and instructors". They are divided into study questions, study activities and quite an extensive list of further readings and publications of possible interest to the reader. This part is very informative for the reader, and particularly the study questions and activities provide excellent scaffolding for deeper insights and understanding, as all the questions relate back to the individual chapters of the section and encourage the reader to revisit and critically examine each one. The activities ask the reader to apply some of the issues addressed in the preceding chapters, and are quite practical. For example, one activity suggests recording a conversation and listening to several of its linguistic features, or interviewing language learners about their idea of identity, or observing multilingual learners and their use of language in different contexts. The notes for learners and instructors are particularly helpful for readers, as they help us to understand chapters with a high density of information, theory and concepts. I imagine the notes would be especially beneficial to students to facilitate alleviate comprehension.

Part I of the volume contains five chapters, and is related to various concepts of native speakers and language learners. It starts off with a critical survey of interpretations of who or what a native speaker is (Alan Davies), followed by a chapter questioning the relevance of native speakers in English as a Second Language teaching and proposing language expertise, language inheritance and affiliation as

more adequate concepts (Constant Leung, Roxy Harris, and Ben Rampton). The next chapter attempts to present "ownership of English" as a concept to overcome the dichotomy of native speaker-non-native speaker (Christina Higgins). Non-native speaker teachers are discussed next (Enric Llurda) and this part concludes with exploring the idea of what the nature of the L2 user is (Vivian Cook).

Part II attends to issues of reconceptualising language in language learning and practice, and contains six chapters dealing with how English is seen and used in different socio-cultural contexts of institutional practices (Angel Lin, Wendy Wang, Nobuhiko Akamatsu, and A. Mehdi Riazi), authenticity from the perspective of hip-hop (Alastair Pennycook), issues of English as a lingua franca (Barbara Seidlhofer) and in multilingual communities (Suresh Canagarajah), the notion of identity in L2 writing (Ken Hyland), the usefulness of a corpus-based approach to applied linguistics (Douglas Biber, Susan Conrad and Randi Reppen), and how to apply the corpus-based approach regarding creative language use in a range of discourses (Ronald Carter and Michael McCarthy).

Part III presents critical issues in applied linguistics, and has six chapters ranging from problems in developing a theory of social identity which integrates the language learner and the language learning context (Bonny Norton Peirce) to the presentation of a post-structuralist theory of identity (David Block) to the components of gender, class and race within the concept of identity (Ryuko Kubota) to the concept of communicative competence. The latter is addressed in the last three chapters of this section that focus on the positive impact communicative competence has on shifting approaches in language teaching to communicative ones (Constant Leung), on the communicative competence of multilingual learners in a multicultural context (Claire Kramsch and Anne Whiteside), and on examining discourses, such as genre and style of speaking, on communication skills (Deborah Cameron).

Part IV examines applied linguistics in a changing world comprising eight individual chapters. The first one presents the difficulties of an ESL researcher publishing articles internationally, raising issues of discourse community and learning as participation (John Flowerdew), followed by a challenging view to rethink priorities in language testing research and addressing language assessment as social practice (Tim McNamara). A qualitative study of ESL programs for immigrants and the linguistic and social processes involved is next (Patricia Duff, Ping Wong and Margaret Early). Subsequent chapters discuss language policies in multilingual societies (Nancy Hornberger), how Translation Studies and Political Discourse Analysis can benefit from each other (Christina Schäffner), how everyday creativity works in language and its influence within applied linguistics (Janet Maybin and Joan Swann), the high linguistic and conceptual complexity of judicial texts and police cautions challenging ESL speakers with high proficiency (Aneta Pavlenko), and the discourse of organic food promotions in relation to commercial and political discourses (Guy Cook, Matt Reed and Alison Twiner).

The final section of the book, entitled ‘Doing applied linguistics’ is authored by the editor Li Wei and Zhu Hua. This section stands apart from the previous ones, adopting a more generic perspective in looking at the topic of method and methodology in applied linguistics, the processes involved in doing research and ethical considerations. This section is very clear, and gives the reader a very well structured overview of how research should and could be done in the field. I would imagine it to be useful for both students and for researchers. Like the previous four sections, it is followed by study questions and activities.

All in all, this is a high quality collection by leading scholars on an array of topics. All chapters are clearly written, although information and concepts are very dense in some chapters. The book would be useful for both undergraduate and postgraduate students and, as stated intention in the book’s preface preface, it can easily be used as a teaching text. The Reader will also prove useful and very informative for researchers and practitioners in language learning and teaching.

References

Davies, A. & Elder, C. (Eds.). (2004). *Handbook of applied linguistics*. Oxford: Blackwell.
Sealey, A. & Carter, B. (2004). *Applied linguistics as social science*. London: Continuum.

CHRISTINE BIEBRICHER, The University of Auckland

Turner, J. 2011. *Language in the academy: Cultural reflexivity and intercultural dynamics.* Bristol, Buffalo and Toronto: Multilingual Matters. IBSN-13:978-1-84769-321-1. 220 pp.

Given the rapid growth in numbers of South-East Asian students in universities in New Zealand, the examination of assumptions about university language which Turner's book offers is very valuable. It is likely to be of interest to writing teachers and EAP practitioners as well as tertiary teachers of international students more generally.

The book provides a fascinating discussion of the nature of modern academic language. Turner discusses the roots of written language in the academy in early scientific writing, and the origins of both written, and in particular of spoken interaction, in. for example, tutorials, in Socratic dialogue. Turner argues that language, partly because of this origin in early science writing has been marginalised and made invisible in the academy. She notes the importance of clarity in science writing, and the way that, for early science writers such as Newton, it was essential to make clear to readers the natural phenomenon that the scientist had investigated experimentally. Such writing was highly persuasive, and as Bazerman (1988) has discussed, stood in the place of the reader observing the experiment themselves.

Other key highly prized features of early and modern science writing were conciseness and brevity, without literary flourishes. The implication was that such writing was more likely to provide a true reflection of nature. Writing needed to be transparent, so as not to stand in the way of its meaning. The writer's ideas are all-important, and language must express them clearly without obscuring them. Most interesting in Turner's argument is the way that the need for writing to be transparent led to it becoming invisible. It is only when problems in communication occur that language becomes visible. Turner discusses how often the language of undergraduate students, both domestic and international, is pathologised, and seen in terms of deficit. Students' language becomes visible only when they struggle to use language in the ways the academy views as natural, but which is in fact highly constrained by convention.

However, although Academic literacy/EAP is used as a solution to this perceived deficit, this discipline is marginalised, because in the academy content is valorised above language. Turner uses Swales' notion of the "the Ivory Ghetto": the fact that the bygone elite standards of the Ivory Tower need to be propped up by a discipline that ought, by the standards of the academy, not to be necessary.

The book examines the nature of "western" academic language by considering it in the light of the very different tradition of teaching and learning that arose from Confucianism. This tradition prizes harmony (and thus remaining silent when you disagree with the majority). By contrast, the Socratic dialogic tradition prizes

students' ability to express opinions and critique; tertiary teachers thus view their role as encouraging this through the use of questioning. Turner provides a very interesting analysis of the mismatch for South East Asian students between their own and their western teachers' expectations of roles for learners and teachers.

Turner relies on data from one-on-one tutorials between Japanese Fine Art students in the UK and their tutors. Tutors use inductive questioning to try to elicit from students a critique of their artworks, and an analysis of the development of their work. This expectation is frustrated by students' expectations of the tutors as guides who will evaluate their work rather than expecting them to evaluate it themselves.

Turner characterises the teacher's role as a "midwife" in the Socratic tradition in which they "deliver" students' understanding and logic through talk. By comparison, in the Confucian tradition the teacher is a model and guide. Interestingly, the resulting change in teaching and talk is not all on the side of the students. Instead, she demonstrates the development of intercultural interaction as a result of "reverse midwifery" by the Japanese students. The students' expectations and refusal or inability to meet tutor expectations of critical analysis in the tutorials induce the tutors to explain their expectations explicitly, and to provide some elements of the modelling that their Japanese students expect.

This book addresses some very important issues; it draws on real data from tutorial interaction, but places this within a broad survey of relevant literature as well as within a thoughtful discussion and analysis of the Socratic and Confucian traditions. The importance of the book for me as a teacher of writing to international students is in the balanced insight it provides into the perspective and expectations of South East Asian students. It advances evidence against too-easy interpretations of the classroom behaviour of such students as passive, as reliant on rote learning, or as lacking in critical thinking ability. Crucially it suggests a need for a movement towards an intercultural "third space" in interactions between international students and their tutors and lecturers in western countries.

References

Bazerman, C. (1988). *Shaping written knowledge*. Madison: University of Wisconsin Press.

Swales, J. (1990). *Genre analysis English in academic and research settings*. Cambridge: Cambridge University Press.

JEAN PARKINSON, Victoria University of Wellington

O'Sullivan, B. (Ed.) (2011). *Language testing: Theories and practices*. Basingstoke, UK: Palgrave Macmillan. ISBN 9780230230637(pbk.). 310 pp.

This volume is a collection of original articles on current work in language testing in several countries around the world, including the UK, Mexico, Turkey, the United Arab Emirates and Japan. It is part of a Palgrave series entitled Advances in Linguistics, which covers various topic areas in applied linguistics rather than purely linguistic analysis, to judge by the volumes that have appeared so far.

Even after reading the series editor's preface and the editor's introduction, it is not entirely clear what the purpose of the book is. The target audience is identified as postgraduate students, as well as tutors or researchers in cognate fields seeking to learn about current issues and future directions in language testing. This would suggest the need for quite a substantial introductory section to the book, to orient the intended readers to the field. In fact, the introduction is relatively short and is primarily concerned with outlining two bases on which the volume is supposedly organised. The first is in terms of three themes: validation, professionalisation and localisation. Validation is certainly prominent in a number of the chapters but the other two themes are rather less so. The second organising principle is a division of the 14 chapters into four sections, identified by such vague descriptions as "the application of theory to practice". However, the section divisions are not included in the Contents or signalled in any way through the book.

The editor, Barry O'Sullivan, is a prominent academic in the British language testing community, who has completed numerous consultancies and test development projects internationally. Thus, the book is probably best seen as a set of articles that reflect his range of academic and professional interests in the field, particularly since quite a few of the contributors have been associated with him as colleagues, research students or project participants. There is a great deal of interesting and innovative work on testing and assessment reported in the various chapters, but the diversity of the content makes it difficult to summarise in a short review like this.

Some of the chapters deal with general issues in the field. Barry O'Sullivan and Cyril Weir consider the limitations of current frameworks for test validation and argue that Weir's socio-cognitive approach is more useful for practitioners to use. Drawing on his experience with the Common European Framework of Reference, Brian North discusses a number of the challenges in describing levels of language ability. A related chapter by Elif Kantarcioglu and Spiros Papageorgiou outlines the process of setting standards and establishing cut scores in language tests.

Three chapters report on substantial research studies conducted in Japan. Paul Joyce explored which of numerous linguistic and psycholinguistic measures were the best predictors of L2 listening ability, using structural equation modelling. On the other hand, Toshihiko Shiotsu used Rasch analysis to investigate the relationship between

learners' self-reported knowledge of high-frequency English vocabulary and the relative frequency of the words in the British National Corpus. In the third study, Barry O'Sullivan and Fumiyo Nakatsuhara drew on the techniques of conversation analysis to find ways of measuring different styles of interaction as students responded to tasks in a group oral test.

The theme of localisation mentioned in the introduction is best exemplified by Adriana Abad Florescana and her 11 co-authors, who describe the development of a low-cost but good quality English proficiency test for students graduating from universities in Vera Cruz, Mexico, for whom the cost of taking an international test like IELTS is prohibitive. However, a chapter by Anthony Green seems to go in the other direction. Green describes a project commissioned by the British Council to produce a placement test to be used in Council testing centres around the world.

Towards the end of the volume, three chapters focus on tests in the context of tertiary admissions. Pauline Rea-Dickins and her colleagues found that admissions tutors at a UK university did not make very well-informed decisions in interpreting the test scores of postgraduate international students, suggesting a need for more assessment literacy. Annie Brown and Paul Jaquith describe the development of an online rater training and marking system to cope with the writing component of an admissions exam for a multi-campus institution in the Emirates. And John O'Dwyer gives an elaborate account of the ongoing efforts of English teaching staff to integrate formative assessment and a formal exam in determining which of their students can be admitted to an English-medium university in Turkey.

The other three chapters defy easy categorisation. Alan Davies contributes a rather curious little essay on applying the Sapir-Whorf linguistic relativity principle to language testing; Abdul Halim Abdul Raof discusses how to rate conference presentations by civil engineers in Malaysia; and Deirdre Burrell and her colleagues at the University of Reading describe how they use diagnostic and formative assessment procedures to develop the language knowledge of their students who are preparing to be secondary teachers of French.

Series editor Chris Candlin describes the volume as "a conspectus of perspectives on key themes", and that might be as good a way as any to sum it up.

JOHN READ, The University of Auckland

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. Avoid using templates and styles that will affect editorial changes and print formatting. Submission of a manuscript of any type implies that it has not been published previously and that it is not under consideration for publication elsewhere.

1.2 A separate title page should include the following:

- the title of the submission
- author's name, and in the case of more than one author, an indication of which author will receive the correspondence
- affiliations of all authors
- full postal address and telephone, e-mail and fax numbers of all authors
- a brief autobiographical sketch of the authors(s) (50-80 words)
- any references removed for the review process

1.3 Copies should be submitted as a Word attachment to one of the co-editors, Dr Gillian Skyrme or Dr Anne Feryok:

g.r.skyrme@massey.ac.nz anne.feryok@otago.ac.nz

1.4 All relevant submissions will be reviewed by members of the Editorial Board or other referees.

2. Presentation of Manuscripts (All Types)

2.1 Sections 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, but should

also be in format suited to editorial changes and print formatting. Figures (e.g. charts and diagrams) and tables should be numbered consecutively in the order to which they are referred. They should not be included within the text, but submitted each on a separate page. All figures and tables should have a number and a caption. Use APA (American Psychological Association) conventions.

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 list of 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).

Where the work of the authors of the article is cited, to avoid identification during the review process the reference within the text should be '(Author, [date])', but there should be no entry in the list of references. Provide these references on the title page.

2.5 Use APA style. 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:

References

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.

Internet sources

Sanders, R. (2006). The imponderable bloom: Reconsidering the role of technology in education. *Innovate Journal of Online Education, 2*(6). Retrieved from http//www.innovateonline.info/index.php?view=article&id=232

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

If articles are not submitted in APA style, they will be returned during the review process for authors to revise.

3. Articles

3.1 Articles should normally be between 3000 and 5000 words in length, exclusive of references, figures and tables, and appendices; please be reasonable. Articles over 6000 words will be returned without review unless prior arrangements have been made with the co-editors.

3.2 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. Include three to five key words on a separate line at the end of the abstract.

4. 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. Short reports should be no longer than 2500 words, exclusive of references, figures and tables, and appendices; please be reasonable. Short reports do not include an abstract or key words. Submissions to this section

follow the submission and presentation guidelines. Those interested in contributing to this section should contact the Co-editors.

5. 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. Submissions to this section follow the submission and presentation guidelines. Those interested in contributing reviews should contact the Reviews Editor, Dr Rosemary Wette, r.wette@auckland.ac.nz.