

and Multiculturalism, 3, 1, pp. 1 - 22.

_____. (1999). L1 Japanese Attrition and Regaining: The Age and Pre-Attrition Proficiency Variables. In P. Robinson (Ed.), *Representation and Process: Proceedings of the 3rd Pacific Second Language Research Forum Vol. 1*, pp. 321 - 335. Tokyo: PacSLRF.

Returnees' Writing Skills: A Comparison with EFL Learners

Taura Hideyuki

Senri International School, Osaka, Japan

e-mail:PDF02662@nifty.ne.jp

and **Amanda Taura**

Kyoto Girls' High School, Japan

e-mail:t-amanda@rd5.so-net.ne.jp

This study examines whether the residual effects of long-term residence (three years or more) in countries where English is spoken give an advantage to young returnees in one aspect of their later English education: the development of English writing proficiency. A cross-sectional framework was used to investigate the English writing skills of returnees and ordinary Japanese EFL learners from grades 7 through 12 (ages 12 - 18). The Test of Written Language (3rd edition) by Hammill & Larsen (1996) was used to examine various aspects of writing while maintaining reliability and validity. Contextual Conventions (CC), Contextual Language (CL), Story Construction (StC), overall Quotient and the total number of words of the writing samples were evaluated.

The results suggest that extensive exposure to English gives young returnees a solid foundation upon which they can build when they start formal English lessons in junior high school. In particular, the returnees' StC and Quotient figures were significantly higher than those of their peers at the beginning of junior high school, suggesting that their overseas experience gives them greater overall writing competence and greater ability to express themselves in a creative and mature way. The returnees' CC, CL, StC and Quotient scores also improved at a faster rate than those of their peers throughout junior high school, although they tended to drop off in the returnees' final year or two of senior high school. By that time, the ordinary EFL learners appeared to be able to develop their proficiency to levels close to those of the returnees. Thus this study suggests that returnees have a head start in the development of English writing proficiency, but that by the end of high school, this advantage is almost eliminated.

<日本人帰国生徒と日本人EFL学習者にみる英語ライティング力の発達段階の比較>

本研究は、英語圏に長期間滞在した日本人帰国生徒が帰国後英語教育を受ける際に、その体験が生かせるかどうかを、特に英語ライティング力に絞って見たものである。帰国生徒と一般の日本人EFL学習者について、中学1年から高校3年まで(12~18才)の期間について横断的研究がなされた。ツールとしては信頼性・妥当性のある Test of Written Language (Hammill & Larsen, 1996)が用いられ、ライティングの諸側面に光が当てられた。具体的には、英語を書く上での決まり事(CC)、文法(CL)、物語の構成(StC)、この3項目の総合点(Quotient)、及び総単語数が研究対象とされた。その結果、中学校で英語教育を始める際に、帰国生徒は、英語圏滞在を通して確固とした英語の基盤をすでに身につけていることが判明した。特に中学初期の段階では、StCとQuotientに関して、帰国生徒の方が遙かに優れており、これより英語ライティングに於いて、創造的で大人びた文章を綴れるだけの英語力が英語圏滞在中で体得しえたことがわかる。帰国生徒のCC, CL, StC及びQuotientの得点から見られるライティング力に関し、学年を追って一般のEFL学習者よりも早いペースで向上しているが、

高校2,3年になるとこの傾向が見られなくなる。理由としては、この時期までにEFL学習者が帰国生徒に追いつくからであると考えられる。即ち、帰国生徒の英語

ライティング力は、日本人EFL学習者に比べ、中学低学年では明らかに優勢を保っているが、高校の最終学年近くになるとこの優位はあまり顕著ではなくなってしまうようである。

INTRODUCTION

It is often assumed in Japan that Japanese children who accompany their parents during prolonged overseas stays in countries where English is spoken become "bilingual" as a matter of course. Research to date offers some support for this assumption: Ono (1994), Minoura (1991), and Harley et al. (1990) all reported that between three and five years after their arrival in the U.S., ESL students generally attain grade norms and English academic proficiency almost equal to their monolingual counterparts.

However, little is known about how long such proficiency is maintained after these children return to Japan or whether it is really an advantage in their formal study of English in secondary schools in this country. Most returnees who come back to Japan during their primary school years attend local Japanese elementary schools where no English maintenance programs are offered, and may not be exposed to formal English instruction until they begin junior high school--anywhere from a few months to several years later. Thus, attrition of their English language skills may be well underway.

This study explores whether the residual effects of long-term overseas residence in countries where English is spoken benefit returnees in one aspect of their later English education: the development of English writing proficiency. The writing skills of junior and senior high school students who lived in English-speaking countries for three years or more are compared to those of their Japanese peers who had no overseas living experience to see whether the returnees' long-term exposure to English gives them a lasting advantage.

Issues in L2 Writing

Target Groups and Study Design

To date, the majority of research on second language (L2) learners' writing skills (particularly Polio, 1997; Wolfe-Quintero, Inagaki, and Kim, 1998), including both "process-oriented" (organization) and "product-oriented" (linguistic accuracy) analyses, has targeted university students. A major reason for this may be the convenience to researchers of using their own university classes as a part of their investigation. Another possible reason might be that the researchers surmised that their subjects' L2 skills were relatively more stabilized than those of beginning EFL learners and could therefore be measured more reliably.

In contrast, the present study attempts to observe students' L2 writing skills at all stages of their English education at Japanese junior and senior high schools. This is based on the assumption that the natural, on-going maturation of cognitive skills together with school training in academic thinking/writing would reveal improvements in writing skills from year to year in a more noticeable way than with university students. Research on the development of writing skills to date includes longitudinal case studies following a single L2 learner's writing skills (eg. Elliot, 1986), but the present study targets many more subjects cross-sectionally. The research aim is to compare returnees and ordinary Japanese EFL learners in terms of their development of English writing proficiency by examining writing by students of both types in all grades in junior and senior high school.

Aspects of L2 Writing

Research to date has covered many aspects of L2 composition. Jacobes et al. (1981) examined content, organization, vocabulary, language and mechanics. Mullen (1977) looked at control over English structure, organization of material, appropriateness of vocabulary, quantity of writing, and overall writing proficiency (a composite rating derived by adding the scores for the first four categories). Prichard (cited in Brown & Bailey, 1984, p. 26) studied organization, length, content, grammar, vocabulary, and spelling, while Hedgcock & Lefkowitz (1992) considered grammar, vocabulary, and mechanics.

Although most of these studies have dealt with both the "product" and "process" of L2 (ESL or EFL) writing, Polio (1997, p.103) points out that in current writing pedagogy there is an overemphasis on the writing process, or idea generation, rather than simply getting students to write error-free sentences. In the same vein, Silva (1993) calls for further research to fill the gaps of previous studies with a more even distribution of strategic, rhetorical and linguistic concerns. On the other hand, Uzawa (1996), who examined the characteristics of skilled and unskilled writers, contends that unskilled writers are overly concerned with mechanics like spelling, punctuation, and grammar. She further argues that unskilled writers tend to cut short the planning phase before writing and neglect organization and the audience. The present study therefore seeks to explore both the process and product of L2 writing by examining the organization and content as well as the grammar and mechanics, including the punctuation, spelling and vocabulary of writing samples.

Scoring Measurements

The question of global ratings versus more objective scoring methods has been debated in the field of L2 writing. Brown and Bailey (1984) modified Stiggins' (1982) approaches to scoring L2 writing by

adding one more recent quantifying approach, that of holistic scoring in which "each composition is assigned a unitary rating, often on a four- to eight-point scale" (p. 22). An example of this type of holistic scoring can be seen in Hamp-Lyons & Henning (1991). They measure linguistic accuracy on a nine-point scale. A score of 9 is given when the reader sees no errors in vocabulary, spelling, punctuation, or grammar, while a 5 is given when errors frequently intrude.

Analytic scoring is a second type of global rating which "involves isolating one or more characteristics of writing and scoring them individually. Such characteristics may include style, grammar, mechanics, etc." (Brown and Bailey, 1984, p. 22).

A third rating guideline employs objective (quantifying) scoring by using such tools as "T-units or a ratio of cohesive devices to composition length" (Brown and Bailey, 1984, p. 22). The notion of a T-unit (terminable unit; the smallest unit within a sentence that can stand alone), which was first introduced by Hunt (1965), has been employed in many other L2 writing studies such as Flahive & Snow (1980), Kameen (1979), and Perkins (1980). Hunt (1965) found that the ratio of clauses per T-unit and the length of T-units increase as writing skills develop. Flahive & Snow (1980) found strong correlations between holistic scores and T-unit measures with higher-level ESL learners.

In an empirical study comparing major measures used in L2 writing research, Polio (1997) categorized the most frequently used measurements into three types: holistic (including analytic scoring), error-free T-units, and finally an error classification system. In their quest to find which methods, either direct (holistic and analytical) or indirect (multiple choice and objective tests) are more reliable, Perkins, Pohlmann, and Brutten (1988) found the holistic/analytical assessment to be more reliable than objective assessment.

The present researchers considered a number of these different scoring measures used in previous studies (and sometimes labelled under different names) in order to find more reliable measures to use in this study. It was decided to adopt holistic/analytical measures for two main reasons: first, because of the finding by Perkins et al. (1988) mentioned above, and second, because of Ishikawa's (1995) research showing that holistic/analytical measures are aimed at placement and are more suitable when dealing with a wide range of proficiencies (as in the inter-grade differentiation in this study) than when evaluating homogeneous populations at a single proficiency level (intra-grade differentiation). Both a sub-categorical items analysis (analytical) and overall assessment (holistic) were required to look at multi-dimensional aspects of writing.

T-unit scoring was not used in the present study despite Hunt's findings (1965) because researchers such as Ishikawa (1995), Hunt (1966), Homburg (1984), and Gaies (1980) have said that T-unit scoring may only be suitable for higher skilled L2 learners, not students such as the lower proficiency Japanese

EFL junior high school students in this study. An additional reason for not using T-unit scoring was that Flahive & Snow (1980) found that holistic scoring highly correlated with the T-unit measurement. Also, Perkins et al. (1988) showed that the reliability of combined holistic and objective assessments was similar to holistic assessment alone. Holistic scoring therefore seemed to be sufficient.

Error-free T-units and an error classification system were not adopted on the grounds that defining "error" and "error-free" may be problematic due to a lack of definition guidelines in most studies, as noted by Ishikawa (1995) and Polio (1997).

Finally, although the notion of T-units is not used in this study, a different objective measure, counting the total number of words in timed L2 writing, was selected because it has been found to be the best predictor of L2 learners' writing proficiency (Brown & Bailey, 1984; Arthur, 1979; Silva, 1993). For all of these reasons, then, this study adopts holistic/analytic scoring methods plus the objective measure of counting the total number of words in the L2 learners' writing samples.

Method

Subjects

This study analyzed 291 writing samples collected at Senri International School, Osaka, Japan. They were written by two groups of students: 116 returnees and 175 ordinary Japanese EFL learners. Based on the results of research by Ono (1994), Minoura (1991), and Harley et al. (1990), who all reported that it takes three to five years for ESL students newly arrived in America to attain grade norms and English academic proficiency almost equal to their monolingual counterparts, returnees who had spent less than three years in English-speaking countries were not included in this study. The subjects in the returnee group (total number $n = 116$; 14 in Grade 7, 14 in Grade 8, 22 in Grade 9, 12 in Grade 10, 33 in Grade 11 and 21 in Grade 12) had all studied at elementary or junior or senior high schools in English-speaking countries for more than three years. However, the period of time elapsed since they returned to Japan varies, depending upon the individual returnee. There was no screening process involved in selecting the subjects for this study except those for the returnees. The EFL learners (total number $n = 175$; 37 in Grade 7, 25 in Grade 8, 27 in Grade 9, 28 in Grade 10, 26 in Grade 11, and 32 in Grade 12) were students who had not visited or lived in English-speaking countries for more than three months.

Approach

As discussed above, this study examines L2 writing samples for organization and content, as well as grammar and mechanics, and punctuation, spelling and vocabulary. The present researchers looked for a tool which best met all these needs and decided to use the 3rd Edition of the Test of Written Language (TOWL-3) created by Hammill and Larsen (1996). TOWL-3 looks at three abilities in writing. First, it examines the ability to write in compliance with accepted standards, especially those governing punctuation, capitalization, and spelling. These skills are tested within the subset known as Contextual Conventions (CC). Secondly, the Contextual Language (CL) rating describes the subject's ability to use the syntactic, morphological, and semantic elements of English (suitable wording, tense, plurals, noun-verb agreement). Lastly, the Story Construction subset (StC) analyzes the subject's ability to express ideas, opinions and thoughts in a creative and mature way (i.e., logical and coherent story development and reader impact). Thus, TOWL-3 seems to satisfactorily cover all the aspects of L2 writing this study undertakes to examine. Moreover, TOWL-3 observes other relevant factors such as introductory phrases, interrogative sentences, coordinate and subordinate coordination, and articles. Silva (1993) mentions the significance of introductory phrases, which tend to be few in number and small in range in L2 learners' writing samples. Dunkelblau (1990) notes that interrogative sentences appear less frequently in L2 writing. Silva (1993) and Elliot (1986) found more subordinate coordination and Oi (1984) discovered fewer articles in L2 writing.

With reference to scoring measures, this study adopts holistic, analytic methods (with the total word count being the only objective measure) as mentioned above. TOWL-3 uses a two- to four-point scoring scale for all items across the three subsets (CC, CL and StC). Each subset employs an analytical approach for discrete items, such as "subject-verb agreement" in the CL subset (0 points for more than 1 error, 1 point for 1 error, or 2 points for no errors), as well as a holistic approach, evaluating prose in the StC subset, for example, as immature (0 points), ordinary (1 point), or stylish (2 points). There are 11 to 14 items in each subset; the CC, CL, and StC raw scores are individually added up and then transformed into the standard scores provided by TOWL-3 for the respective subsets. (See Table 1.) For every one of the subtests at each age level, the mean score is 10 and the standard deviation (SD) 3. Normative data in TOWL-3 are provided based on 2,217 North American native English-speaking children, aged 7 -18, to give useful guidelines for grading each subject's writing level. TOWL-3 also calculates the total sum derived by adding the three subset standard scores and converting the sum to a quotient using the table reproduced at the end of this paper in Appendix 2B. This "Composite Quotient" estimates the subject's overall writing competence, with a mean score of (M) 100 and the standard deviation (SD) fixed at 15. Quotients of 131 - 165, for example, are described as "very superior" (top 2.34%), 90 - 110 as "average" (49.51%), and 80 - 89 as "below average" (16.12%).

TABLE 1: Guidelines for Interpreting TOWL-3 Standard Scores in CC, CL, & StC

Standard			
S c o r e			
Description			
Percentile			

17 to 20	Very Superior	2.34%
15 to 16	Superior	6.87%
13 to 14	Above Average	16.12%
8 to 12	Average	49.51%
6 to 7	Below Average	16.12%
4 to 5	Poor	6.87%
1 to 3	Very Poor	2.34%

* Full score=20, Mean score=10, SD=3

Both contrived and spontaneous formats are used in TOWL-3. The contrived format taps the discrete elements or the clearly defined boundaries of the language used and evaluates them in isolation; that is, it tests subjects' ability to use specific constructions to write individual sentences grammatically.

The spontaneous format, on the other hand, tests the ability to write freely about any topic pertinent to the reader without boundaries. The studies mentioned in the beginning of this paper (e.g. Polio, 1997; Wolfe-Quintero, Inagaki and Kim, 1998; Hedgcock and Lefkowitz, 1992; Uzawa, 1996) did not use the contrived format due to its unnatural way of examining the written language. This study therefore also adopts the spontaneous format alone, with the aim of employing a test format that is closer to a classroom situation where students are asked to write on a specified topic rather than being tested on a limited range of grammatical questions. Another reason for using only the TOWL-3 spontaneous format is that it requires examinees to write an expository essay. Silva (1993), in reviewing the existing literature, found an overall preference for researchers to use expository essays over argumentative and narrative types. Thus the spontaneous format allows an easier comparison of this study's results with others'.

To supplement the TOWL-3 method, this study also calculates the total number of words in the writing samples for a statistical analysis using objective measurement.

Procedures

Students' spontaneous written stories were elicited by their English instructors during their English classes (10 classes in total) in the following manner: A piece of paper with a picture of some "prehistoric" creatures on it (Appendix 1) was handed to each student. The following instructions were then given by all the examiners, as directed in the TOWL-3 examiner's manual.

This exercise is designed to see how well you can write a story. Look at the picture before you. You are to write a story about that picture. Before you begin writing, take time to plan your story. A well-written story usually has a beginning, middle, and end. It also has characters that have names and perform certain actions. Use paragraphs to help organize your story. Correct punctuation and capitalization will make your story easier to read. After you have made a plan for your story, begin writing. Try to write as long a story as you can. If you need more paper, just let me know. You will have only 15 minutes to think about your story and to write it. Write the best story you can. Ready? Begin.

Because the EFL learners in Grades 7 and 8 had only beginner comprehension skills, the students in these classes were also given a Japanese translation of the instructions.

The 15-minute time limit for planning and writing of the story is set down in the guidelines of the TOWL-3 spontaneous format. When 15 minutes had elapsed, all the essays were collected for scoring as stipulated by the TOWL-3 examiner's manual.

Scoring

Four ESL teachers, one of whom is a native speaker of English, studied the TOWL-3 examiner's manual and practiced scoring the sample stories in the appendix before rating each subject's writing

sample. The four raters evaluated the same writing sample and their scores for the CC, CL and StC subsets are shown in Table 2.

TABLE 2: Raters' Scores for CC, CL and StC Subsets on Identical Writing Sample

Score for				Rater's Mean	
	CC	CL	StC	CC	CL
Rater 1*	18	11	9		
Rater 2	7	16	11		
Rater 3	9	18	11		
Rater 4	11	14	13		

* Rater 1 is the only native English speaker.

The Pearson's correlation coefficients for the four raters are shown in Table 3.

TABLE 3: Pearson's Correlation for Raters

Rater	Rater 1	Rater 2	Rater 3	Rater 4
Rater 1				
Rater 2				
Rater 3				
Rater 4				

Rater 4

Rater 1	1.00	0.97	1.00	0.87
Rater 2		1.00	0.96	0.96
Rater 3			1.00	0.87
Rater 4				1.00

The Pearson interrater reliability with Z transformation was .83.

Results

SPSS Version 6.1.1¹ was used to statistically analyze the five scores (CC, CL, StC, Quotient, and total number of words in the essay) for the 291 subjects, using t-tests for intra-grade comparison of the returnees and EFL learners in each grade and then ANOVAs (analysis of variances) for inter-grade comparison of the five variables across the six grades. The intra-grade comparisons of the two groups are presented in Table 4, with the T-test results shown in Appendix 2A, while the inter-grade comparisons are provided in Table 5, with the ANOVA and Scheffe test results in Appendix 2B.

As shown in Table 4, significant intra-grade differences (adjusted by Bonferroni tests at an acceptable level of .0083) were observed in the StC and Quotient figures of the students in Grade 7. In other words, the returnees' ability to express ideas, opinions and thoughts in a creative and mature way (StC) and their overall writing competence (Quotient) were significantly superior to those of their peers (ordinary EFL learners) in seventh grade--the starting point of formal English education in the Japanese education system. In contrast, no significant differences were found between the two groups in any grade in terms of the Contextual Conventions (CC) ratings, Contextual Language (ability to use suitable wording, tense, plural and noun-verb agreement) scores, or the total number of words written.

TABLE 4: Intra-Grade Comparisons of Returnees' and Ordinary EFL Learners' Essay Ratings

								Grade
								<i>n</i>
								Group
								CC
								CL
								StC
								Quotie
								nt #
								Words
								7
								14

Returnees	7.90	7.00	*5.9	**80.4	160.10			
	37	EFL	7.20	2.60	*3.1	**64.4	43.80	
8	14	Returnees	9.90	8.90	8.40	94.00	154.70	25
EFL	6.60	3.00	3.60	63.90	57.20			
9	22	Returnees	9.90	11.80	10.60	104.90	164.00	
	27	EFL	7.80	4.00	4.50	67.90	120.90	

10	12	Returnees	10.80	10.80	14.00	112.10	158.60
	28	EFL	8.10	4.20	5.00	70.70	90.50
11	33	Returnees	11.90	13.40	12.50	116.00	193.60
	26	EFL	7.30	8.80	9.00	91.30	90.70
12	21	Returnees	9.20	10.10	10.80	99.90	155.00
	32	EFL	6.90	10.80	10.20	99.90	103.60

(1) * $p = .002$ (Intra-grade differences between G7 bilingual & EFL in StC scores)

(2) ** $p = .007$ (Intra-grade differences between G7 bilingual & EFL in Quotient scores)

As shown in Table 5, the inter-grade comparison using one-way ANOVAs showed significant differences (on post-hoc Scheffe tests with the alpha level adjusted to $p < .05$) across the six grades in all five variables (CC, CL, StC, Quotient and the total number of words), although not always in both groups. In terms of Contextual Conventions such as capitalization, punctuation and spelling, EFL learners from Grade 7 through 12 scored within a similar range, yielding no statistical difference, but the returnees in Grade 11 scored significantly better than the returnees in Grades 7 and 12. In contrast, there appeared to be a gradual improvement from Grade 7 to 12 in the Contextual Language evaluations (suitable wording, tense, plural and noun-verb agreement) for the ordinary EFL learners, with significant differences between the ratings for the students in Grades 7 - 10 as compared to those in Grades 11 - 12. A similar difference was apparent for the returnees, with the evaluations of those in Grades 9 - 12 significantly higher than for those in Grades 7 - 8. Likewise, the ratings for Story Construction (expressing ideas, opinions and thoughts in a creative and mature way) rose with the subjects' grade in school, with statistical differences between the EFL learners in Grades 7 - 10 and those in Grades 11 - 12, and between the returnees in Grades 7 - 8 and those in Grades 9 - 12. The Quotient (indicating overall writing ability) also shows a gradual improvement all the way through, with the ratings for the Grade 7 - 10 group of EFL learners differing significantly from the Grade 11 - 12 group, and those for the Grades 7 and 8 returnee group significantly different from the Grade 9 - 12 returnee group. Finally, the total number of words written by the EFL learners in the Grade 9 - 12 group was significantly higher than the number written by the EFL learners in the seventh and eighth grade, although there were no differences between the different returnee grade groups.

TABLE 5: Inter-Grade Comparisons of CC, CL, StC, Quotient and Number of Words

EFL Learners	Returnees	Returnees
CC	No significant difference observed	G7<G11, G12<G11
CL	G7<G11/12, G8<G11/12, G9<G11/12, G10<G11/12	G7<G9/10/11, G8<G11, G12<G11
StC	G7<G11/12, G8<G11/12, G9<G11/12, G10<G11/12	G7<G9/10/11, G8<G10/11
Quotient	G7<G11/12, G8<G11/12, G9<G11/12, G10<G11/12	G7< G9/10/11/12, G8<G11, G12<G11
# Words	G7<G9/10/11/12, G8<G9/10/11/12	No significant difference observed

Discussion

In this section, the results presented above are discussed in the order the evaluation subsets were originally described: first CC, then CL, StC, Quotient, and finally the total number of words. Afterward, decreases in G12 returnees' scores are highlighted and possible interpretations of these results are considered.

Development in Contextual Conventions The ANOVA results of the contextual conventions ratings for the essays by the returnees and EFL learners in each grade are presented below in Figure 1. English native norms for TOWL-3 are set in the range of 8 - 12. In Grade 7, the first year of their English

education, the EFL learners appear to have almost caught up with their returnee peers in terms of their mastery of contextual conventions, and are close to the lower end of the English native speaker norm range. At the time these writing samples were taken, the G7 EFL learners had studied English for only eight months. In other words, it may take only eight months or less for new English learners to pick up basic English writing conventions such as capitalization and punctuation.

FIGURE 1: Inter-Grade Comparison of Contextual Conventions Scores

Yet despite this rapid acquisition of writing conventions in the first eight months, EFL learners remain in the "below average" range throughout their six years of secondary school, with the only exception being in Grade 10. This contrasts to the gradual improvement in the returnees, who score in the "average" (English native norm) range from G7 through G12, and score higher than the EFL learners in all six grades.

One interpretation of this could be derived from Flavell's (1985) argument on the necessity of extensive exposure to the target language prior to the gradual growth of metacognitive knowledge. The returnee subjects have been in English speaking countries for at least three years, whereas their EFL counterparts have had no experience of this kind. The accumulation of exposure to the target language may be a necessary factor to boost CC scores as seen in the returnees, although this interpretation can be only verified by observing the gradual progress in CC of older EFL subjects in further studies. The CC scores of the returnee students improve as they study English at school until they reach G12. The

decrease in this skill in G12 returnees will be discussed later.

Development in Contextual Language and Story Construction

Figures 2 and 3 show the ANOVA results of the Contextual Language and Story Construction ratings for the essays by the returnees and EFL learners in each grade. The CL and StC scores of both the returnee and EFL subjects fall in the "below average" range in Grade 7, but from Grade 8 onwards the returnees score either "average" or "above average", while the EFL learners' scores do not reach the "average" range until Grade 11.

FIGURE 2: Inter-Grade Comparison of Contextual Language Scores

One needs to consider the distinctive features of the returnee subjects to explain this. They have all spent three years or longer in English speaking countries. On returning to Japan, however, most attend local Japanese elementary schools, where no English maintenance programs are offered, for periods ranging from a few months to several years. Thus, at the beginning of their English language education in Japan, their English language skills may have atrophied. However, because of the basic knowledge of English that they have already acquired, the length of time required by returnees to reach a more advanced language level (native norms, or the "average" TOWL-3 score range of 8 - 12) is shorter in

comparison with EFL learners, who seem to need about four years to reach the "average" range.

FIGURE 3: Inter-Grade Comparison of Story Construction Scores

The four-year time frame for the EFL students to reach English native norms will be briefly considered here. EFL learners' CL and StC scores share almost identical progress patterns. Although they start off in the "very poor" range of 0 - 2 points in Grade 7, the EFL learners maintain steady progress all the way up to Grade 12, recording a statistically significant leap (post-hoc Scheffe tests with a significant level of .05 for both CL and StC) from Grade 10 to Grade 11, when they reach the average range of 8 - 12 points, or the TOWL-3 native norms.

A possible interpretation of this could be related to the beginning of formal grammar lessons at school. Systematic grammar instruction is generally given at the senior high school level (Grades 10 - 12) in Japan, and it supplements the students' initial exposure to basic English structures and vocabulary in junior high school. One of the reasons why CL and StC ratings progress in a similar way could be that a more solid grammar base may allow EFL learners to concentrate on story construction rather than grammar. Another possible explanation may be that the complicated Japanese rhetoric studied at the senior high school level could be transferred over to English story-construction skills. More clearly defined interpretations still remain to be derived from further studies; however this research shows that EFL learners can reach native norms in contextual language and story construction skills after four years of study.

Turning to differences, intra-grade variability between the returnee and EFL subjects occurs only in

G7 Story Construction. One explanation of this difference in StC skills might be that story construction requires the integration not only of grammar and vocabulary but also other components of English writing such as rhetoric. Thus eight months may be too short a period for EFL learners to combine all these components and produce English prose equivalent to or not significantly lower than that of the returnees. The development of rhetoric and general cognitive skills is certainly a worthwhile issue to explore, but it goes beyond the scope of this study.

The deviation of the G10 returnees' CL scores from the near-linear progression seen from G7 through G11 is very hard to interpret. However the fact that their scores are not significantly different from G9 and G11 returnees and that they still fall in the categories of "average" or "above average" seem to indicate that this slight deviation does not contradict the overall appearance of progress. However it does contradict the explanation given earlier that StC skills improve in accordance with grammar proficiency (CL). This issue will be referred to again later when the G12 returnees' written language deficiency is discussed.

Overall Development

The Quotient scores (Figure 4) are discussed here as a summary of all the subsets, since, as the "total sum of its three subsets and most reliable scores" (Hammill and Larsen, 1996), they provide an overview.

Intra-grade analysis yielded statistical differences between EFL and returnee subjects only in G7, which could be explained in the same way as the difference in Story Construction scores, that is, the amount of time needed to integrate knowledge of the different language components. EFL learners showed a linear upward improvement in Quotient almost identical to their improvement in CL and StC scores, with the same significant dividing line between G10 and 11. Although G7 EFL learners started off in the "very poor" range in CL and StC scores, they were "below average" in Quotient. This is partly due to the calculation of the Quotient as a combination of CL, StC and CC scores, since the EFL learners' CC scores were relatively high. Nonetheless, the EFL learners reached the "average" level in these three subsets in G11 and improved even further in G12. This could indicate that four years of vocabulary building, grammar learning, and proficiency in various writing techniques, including rhetoric, help them to catch up with native norms.

FIGURE 4: Inter-Grade Comparison of Quotients

The returnees' Quotients show an upward progress from G7 to G11 within the range of "average". Although their Grade 10 scores were lower than their Grade 9 scores in CL and higher than their Grade 11 scores in StC, these differences were not statistically significant. Despite such deviations in CL and StC, it is possible to conclude that in terms of their overall writing proficiency as indicated by the Quotient, the returnees show a linear improvement in writing skills despite their limited exposure to English (five one-hour English lessons a week) in junior and senior high school.

Development of Story Length

The total number of words is discussed only briefly here, since there is no normative data to compare the results against, as the researchers were able to do with the TOWL-3 scores.

As Table 4 and Figure 5 show, there are no inter-grade differences in the length of the essays written by the returnees (a total of 150 words or more were produced in all grades), but the G7 - 8 EFL learners scored significantly lower than the G9 - 12 group. This seems logical when we consider that G9 is the grade when writing is heavily emphasized in the curriculum for the English lessons which the present EFL subjects take at their school.

The question of why the EFL learners were in the main part below the level of their returnee peers throughout the six years could partly be answered by their lack of extensive exposure to English; however only further (possibly longitudinal) research could lead us to a more conclusive answer.

FIGURE 5: Inter-Grade Comparison of Total Number of Words

Decline in G12 Returnees' Scores The total number of words in the essays by G12 returnees, like their CC, CL, and StC scores and their Quotients, reflect a drop from G11. This decline in the general upward tendency is difficult to interpret, but two possible reasons are considered here.

One interpretation is that Grade 12 is a period when both returnee and EFL students are busy preparing for university entrance exams. The most important subject for most university applicants in Japan is English. It receives heavy attention by EFL learners, leading to continued improvement and higher scores in G12 than in G11. Meanwhile, the English proficiency level of returnees is high enough for them to put more emphasis on other subjects, and as a result, their writing scores fall.

Another possible explanation is that the group of students in Grade 12 in this particular study happened to have a low overall English writing proficiency. This study could be administered once again in a year's time and the data compared. If similar results were to occur, then the first interpretation would be supported. If the results showed continuous progress in G12, however, the second interpretation would be possible. Data from additional studies are necessary for follow-up on this issue. The results from another administration of this study would also be very useful in order to interpret the deviation in G10 returnees' CL and StC scores.

Conclusion

Extensive exposure to the target language in English-speaking countries for a lengthy period of time

seems to set a solid foundation upon which returnees can build more sophisticated writing skills when they start formal English lessons at school. This includes the ability to write in compliance with certain standards (CC), which G7 EFL learners also learn fairly quickly. EFL learners show a slower pace in reaching the English native speaker "average" range during their six years of secondary school, while the returnees show an almost linear upward progress within the "average" range.

The same difference between returnee and EFL learners is true for their CL, StC, and Quotient scores. In contextual language (CL) and story construction (StC), returnees may possibly be affected by their need to refamiliarize themselves with English for a year or so, since they may well not have used it for a while prior to the commencement of formal English lessons in junior high school. However once they reach the average range, their skills continue to improve and may even reach the above-average range for native speakers.

Formal grammar instruction (which begins in G10 in Japanese high schools) seems to work very well for both EFL learners and returnees in terms of writing skill improvement, as seen in their G11 CL scores. The Quotients show that five hours a week of English lessons are often enough for returnees to improve their writing skills. Returnees' Quotient scores go up as their grammar (CL) improves, suggesting that L2 grammar proficiency may be necessary for overall writing proficiency. Further studies could include a separate independent grammar test and compare the results with the Quotient for a more accurate correlation between grammar and overall writing skills. In examining the relationship between grammar and overall writing proficiency, StC scores also proved to have a high correlation with the Quotient.

A finding which has practical application for EFL learners within the secondary school framework is that approximately four years is needed before they reach grade norms as indicated by the CL and StC scores and the Quotient. There is a tendency for EFL learners to boost their sub-skills in a rapid manner and gradually bring their proficiency levels close to those of the returnees by the end of Grade 12.

In summary, the results suggest that extensive exposure to English gives young returnees a head start in the development of English writing proficiency, but that by the end of high school, this advantage is almost eliminated, perhaps in part due to the different perspectives taken by the two groups towards the English component of college entrance exams. To ascertain whether the G12 returnees may actually show a greater improvement at this level than the EFL learners, more evidence needs to be drawn from further studies which test a greater number of G12 returnees.

NOTES

1. *SPSS for Macintosh Release, 6.11 Standard Version*. SPSS, Inc. (1995).

ACKNOWLEDGEMENTS

We would like to thank the Modern Languages Department of Senri International School for their assessment of the writing samples and Uematsu Shigeo of Setsunan University for his advice on statistical analysis. We would also like to express our thanks to Mary G. Noguchi and the three anonymous readers for their helpful assistance in preparing this article for publication.

REFERENCES

- Ablrechtsen, D. (1997). One Writer, Two Languages: A Case Study of a 15-Year-Old Student's Writing Process in Danish and English. *International Journal of Applied Linguistics*, 7, 2, pp. 223 - 250.
- Arthur, B. (1979). Short Term Changes in EFL Composition Skill. In Yorio, C.A., Perkins, K., and Schachter, J. (Eds.), *On TESOL '79; The Learner in Focus*, pp. 330 - 342. Washington, D.C.: TESOL.
- Brown, J.D. and Beiley, K.M. (1984). A Categorical Instrument for Scoring Second Language Writing Skills. *Language Learning*, 34, 4, pp. 21 - 42.
- Crookes, G. (1989). Planning and Interlanguage Variation. *Studies in Second Language Acquisition*, 11, pp. 367 - 383.
- Dunkelblau, H. (1990). A Contrastive Study of the Organizational Structure and Stylistic Elements of Chinese and English Expository Writing by Chinese High School Students. *Dissertation Abstracts International*, 51, 4, 1143A.
- Elliott, M. (1986). Nasr's Development as a Writer in His Second Language: The First Six Months. *ARAL*, 9, 2, pp. 120 - 153.
- Ellis, R. (1994). *The Study of Second Language Acquisition*. Oxford: Oxford University Press.
- Flahive, D.E. and Snow, B.G. (1980). Measures of Syntactic Complexity and Evaluating ESL Composition. In Oller, J.W. and Perkins, K. (Eds), *Research in Language Testing*, pp. 171 - 176. Rowley, Mass.: Newbury House.
- Flavell, J.H. (1985). *Cognitive Development (2nd Ed.)*. Englewood Cliffs, NJ: Prentice-Hall.
- Gaies, S.J. (1980). T-unit Analysis in Second Language Research: Application, Problems, and Limitations. *TESOL Quarterly*, 14, pp. 53 - 60.
- Hamp-Lyons, L. and Henning, G. (1991). Communicative Writing Profiles: An Investigation of the Transferability of a Multiple-Trait Scoring Instrument Across ESL Writing Assessment Contexts. *Language Learning*, 41, pp. 337 - 373.
- Hammill, D.D., and Larsen, S.C. (1996). *Test of Written Language (3rd Edition)*. Austin, Texas: pro-ed.
- Harley, B., Allen, P., Cummins, J., and Swain, M. (Eds) 1990. *The Development of Second Language Proficiency*. Cambridge: Cambridge University Press.
- Hedgcock, J., and Lefkowitz, N. (1992). Collaborative Oral/Aural Revision in Foreign Language Writing Instruction. *Journal of Second Language Writing*, 3, pp. 255 - 276.
- Homburg, T.J. (1984). Holistic Evaluation of ESL Compositions: Can it be Validated Objectively? *TESOL Quarterly*, 18, pp. 87 - 107.
- Hunt, K.W. (1965). *Grammatical Structures Written at Three Grade Levels*. Champaign, Ill.: National Council of Teachers of English.
- _____. (1966). *Sentence Structures Used by Superior Students in Grades Four and Twelve, and by Adults*. Tallahassee: Florida State University. (ERIC Document Reproduction Service No. ED 010 047).
- Ishikawa S. (1995). Objective Measurement of Low-Proficiency EFL Narrative Writing. *Journal of Second Language Writing*, 4, pp. 51 - 70.
- Jacob, H.L., Zinkgraf, S.A., Wormuth, D.R., Hartfiel, V.F., and Hughey, J.B. (1981). *Testing ESL Composition: A Practical Approach*. Rowley, Mass.: Newbury House.
- Kameen, P.T. (1979). Syntactic Skill and ESL Writing Quality. In C.A. Yorio, K. Perkins, and J. Schachter (Eds.), *On TESOL '79; The Learner in Focus*, pp. 343 - 350. Washington, D.C.: TESOL.
- Kobayashi T. (1992). Native and Nonnative Reactions to ESL Compositions. *TESOL Quarterly*, 26, 1, pp.

81 - 112.

- Kroll, B. (Ed.). (1990). *Second Language Writing: Research Insights for the Classroom*. Cambridge: Cambridge University Press.
- Minoura Y. (1991). *Kodomono ibunka taiken* [Children's Cross-Cultural Experiences]. Tokyo: Shisakusha.
- Mullen, K.A. (1977). Using Rater Judgements in the Evaluation of Writing Proficiency for Non-Native Speakers of English. In Brown, H.D., Yorio, C.A., and Crymes, R.H. (Eds.), *On TESOL '77: Teaching and Learning ESL - Trends to Research and Practice*, pp. 309 - 320. Washington, D.C.: TESOL.
- Oi K. (1984). Cross-Cultural Differences in Rhetorical Patterning: A Study of Japanese and English. *Dissertation Abstracts International*, 45, 8, 2511A.
- O'Loughlin, K. (1994). The Assessment of Writing by English and ESL Teachers. *ARAL*, 17, 1, pp. 23 - 44.
- Ono H. 1994. *Bairingaruru no kagaku* [Science of Bilingualism]. Tokyo: Kodansha.
- Perkins, K. (1980). *Using Objective Measures of Attained Writing Tasks*. Denver: National Assessment of Educational Progress.
- Perkins, K., Pohlmann, J.T., and Brutton, S.R. (1988). A Factor Analysis of Direct and Indirect Measures of English as a Second Language Writing Ability. *Educational and Psychological Measurement*, 48, pp. 1111 - 1121.
- Polio, C.G. (1997). Measures of Linguistic Accuracy in Second Language Writing Research. *Language Learning*, 47, 1, pp. 101 - 143.
- Scarcella, R. (1984). Cohesion in the Writing Development of Native and Non-Native English Speakers. *Dissertation Abstracts International*, 45, 5, 86A.
- Stiggins, R. (1982). A Status Report on Writing Assessment. In Cronnell, B. and Michael J. (Eds.), *Writing: Policies, Problems, and Possibilities*, pp. 137 - 158. Los Alamitos, Calif: SWRL Educational Research and Development.
- Uzawa K. (1996). Second Language Learners' Processes of L1 Writing, L2 Writing, and Translation from L1 into L2. *Journal of Second Language Writing*, 5, 3, pp. 271 - 294.
- Wolfe-Quintero, K., Inagaki, S., and Kim, H. (1998). *Second Language Development in Writing: Measures of Fluency, Accuracy & Complexity*. Honolulu, Hawaii: University of Hawaii Press

APPENDIX 1

Picture of "Prehistoric Creatures" Used to Elicit Essays

APPENDIX 2
A: T-test Results

T-tests for Independent Samples of Grade 7 (StC)

					Variable
					n
Mean					Standard

SE of Mean					
	Returnees	14	5.8571	3.880	1.037 EFL Learners
37	3.7838	2.428	.399		

Mean Difference = 2.0734
Levene's Test for Equality of Variances: F=10.180 P=.002

T-tests for Independent Samples of Grade 7 (Quotient)

Variable

n

Me

an S D

SE of Mean

Returnees	14	80.3571	16.194	4.328	EFL Learners
37	65.8649	10.881	1.79		

Mean Difference = 14.4923

Levene's Test for Equality of Variances: F=7.785 P=.007

B: ANOVA Results

Returnees' G7 - G12 CC Scores

Sou

rce

D.F

S u

Sum of Squares	Mean Squares	F Ratio	F Prob.
Between Groups	5	203.15171	40.6303
Within Groups	109	852.0135	7.8166
Total	114	1055.1652	

Multiple Range Tests: Scheffe test with significance level .05

The difference between two means is significant if

MEAN (J) - MEAN (I) $\geq 1.9769 \cdot \text{RANGE} \cdot \text{SORT} (1/N(1)+1/N(J))$

with the following value(s) for RANGE: 4.79

EFL Learners' G7 - G12 CL Scores

						Sou
rce						.
D.F						S u

Sum of Squares	Mean Squares	F Ratio	F Prob.		
Between Groups	5	1143.2242	228.6449	28.4786	.0000
Within Groups	109	875.1235	8.0287		
Total	114	2018.3478			

Multiple Range Tests: Scheffe test with significance level .05
 The difference between two means is significant if
 $MEAN(J) - MEAN(I) \geq 2.0036 * RANGE * \sqrt{1/N(1) + 1/N(J)}$
 with the following value(s) for RANGE: 4.79

Returnees' G7 - G12 CL Scores

						Sou
rce						.
D.F						S u

Sum of Squares	Mean Squares	F Ratio	F Prob.		
Between Groups	5	498.7607	99.7521	13.2599	.0000
Within Groups	109	819.9871	7.5228		
Total	114	1318.7478			

Multiple Range Tests: Scheffe test with significance level .05

The difference between two means is significant if
 $MEAN (J) - MEAN (I) \geq 1.9394 * RANGE * \sqrt{SORT (1/N(1)+1/N(J))}$
 with the following value(s) for RANGE: 4.79

EFL Learners' G7 - G12 StC Scores

						Sou
rce						.
D.F						S u

Sum of Squares	Mean Squares	F Ratio	F Prob.
Between Groups	5 899.3361	179.8672	23.8242 .0000
Within Groups	109 822.9248	7.5498	
Total	114 1722.2609		

Multiple Range Tests: Scheffe test with significance level .05
 The difference between two means is significant if
 $MEAN (J) - MEAN (I) \geq 1.9429 * RANGE * \sqrt{SORT (1/N(1)+1/N(J))}$
 with the following value(s) for RANGE: 4.79

Returnees' G7 - G12 StC Scores

						Sou
rce						.
D.F						S u

Sum of Squares	Mean Squares	F Ratio	F Prob.
Between Groups	5 636.0619	127.2124	15.4407 .0000
Within Groups	109 898.0250	8.2388	

