THE EFFECT OF MODE OF DISCOURSE ON OBJECTIVE MEASURES OF EFL PROFICIENCY IN JAPANESE UNIVERSITY STUDENTS*

Kinue Hirano

Joetsu University of Education

1. INTRODUCTION

Studies attempting to establish an index of second language (L2) development, based on T-unit analysis, have been conducted by a number of L2 researchers (Larsen-Freeman and Strom 1977; Larsen-Freeman 1978, 1983; Kameen 1979; Vann 1979; Harrington 1986). Larsen-Freeman (1983) suggests that this index would be an objective measure as a developmental yardstick by which to assess a learner's global L2 proficiency reliably and objetively (p.287). Such an index would be of great use to program administrators, teachers, and researchers (Larsen-Freeman 1983: 301).

1.1 The effectiveness of objective measures

The T-unit, or minimal terminable unit, is a measure of syntactic complexity/maturity which has been applied widely by first language (L1) researchers (Hunt 1965 and others) and later by L2 researchers.[1] Hunt(1970) defined the T-unit as an independent clause and any subordinate clauses or nonclausal structures attached to or embedded in it.

While some L2 studies (Monroe 1975; Cooper 1976; Kameen 1979) found that the average number of words per T-unit (mean T-unit length) increased with proficiency levels, and it was effective in discriminating among adult L2 learners at different levels of proficiency, other studies in English as a second language (ESL) (Scott and Tucker 1974; Larsen-Freeman and Strom 1977; Larsen-Freeman 1978, 1983) favor as an objective and useful measure of overall L2 proficiency an error-free T-unit (EFT), a T-unit that

16

takes the elimination of errors into account. Larsen-Freeman (1983) states that error-free measures are "more powerful measures (i.e., ones giving us both higher correlations with another measure of proficiency and giving us greater variance or dispersion of scores among proficiency levels)" (p.288).

In their attempt to construct an index of L2 development, Larsen-Freeman and Strom (1977) found that the best discrimination measures were the average number of words per T-unit (mean T-unit length) and the total number of EFTs per composition; the total number of EFTs increased linearly and significantly although the linear increase in mean T-unit length was statistically nonsignificant across the five groups.

In her pursuit of an index of L2 development, Larsen-Freeman (1978, 1983) investigated the sensitivity of T-unit measures to discriminate between proficiency levels. Larsen-Freeman (1978) found that the measures that increased linearly with proficiency and that discriminated among different proficiency levels were: mean T-unit length, the percentage of EFTs (the ratio of EFTs to total T-units), and the average number of words per EFT (mean EFT length); the best discriminators among the five levels of ESL proficiency were found to be the percentage of EFT and mean T-unit length.

Larsen-Freeman (1983) further examined the effectiveness of objective measures on which to base an index of L2 development in three kinds of ESL studies: 1) oral tasks; 2) controlled versus free writing; and 3) the effect of time on writing. Statistical analysis revealed that the different measures were of varied sensitivity in discriminating between proficiency levels; a measure worked well in one case, but not in another case. For the average number of words per EFT, she noted that this measure did not discriminate significantly between all adjacent proficiency levels (e.g., level 1-level 2, level 2-level 3) in her three studies, although it has been reported as a more powerful measure

of L2 proficiency.

In EFL, less work has been done on the adequacy of objective measures of overall English proficiency as an index of EFL development. The results of studies by Kadota (1990) and Tomita (1990) preferred EFTs to mean T-unit length in assessing Japanese learners' overall EFL proficiency. Hirano (1990; (in press)) investigated the argumentative writing of Japanese EFL university students by using a number of objective measures of proficiency that have been claimed to discriminate among various levels. Hirano (in press) found that regardless of audience, the most powerful of the ten objective measures in discriminating between all adjacent proficiency levels of Japanese EFL university students at three levels were mean T-unit length, the total number of words in EFTs, the percentage of EFTs, and the total number of EFTs; however, mean T-unit length did not prove to be as valid and discriminatory a measure as is desired in ESL.

1.2 The effects of task variables on objective measures

The effect of task variables, such as audience, age, mode of discourse, and cognitive style, on syntactic complexity has been examined in L1 studies (Smith and Swan 1978; San Jose 1972; Hunt 1965) and L2 studies (Dvorak 1987; Hirano (in press)).

A number of studies have examined the effect of mode of discourse on syntactic complexity in speech and writing. L1 and L2 studies have found that different modes of discourse result in different levels of syntactic complexity of written language in L1 elementary and high school students (San Jose 1972; Perron 1977; Crowhurst and Piche 1979; Crowhurst 1980) and of spoken and written language in FL university students (Dvorak 1987). All these previous studies indicate that syntactic complexity is greater in the argumentative mode than in other modes of writing. For example, San Jose (1972) and Perron (1977) found that the mean length of T-units was greatest in argumentation, followed by exposition, narration, and description. Crowhurst and Piche (1979)

18

and Crowhurst (1980) concluded that age norms of syntactic development must take into account differences in mode of discourse.

Much less work has been done on the effect of task variables on the effectiveness (i.e., validity and discriminatory power) of objective syntactic measures of ESL/EFL proficiency, in comparison to L1 and L2 studies on the effect of task variables on syntactic complexity. Larsen-Freeman (1983) found that a controlled writing task did not seem, particularly more discriminating among proficiency levels than free composition although any given measure did not prove to be effective in all cases. Dvorak (1987) found that there were no significant interactions between proficiency level and mode of discourse among learners of Spanish as a foreign language. Hirano (in press) investigated the effect of audience on the sensitivity of objective measures of EFL proficiency in Japanese university students. She found that assigning an audience had no significant effect on the validity and discriminatory power of objective measures; however she suggested the need for further research on audience effect.

To my knowledge, no research has been done on investigating whether mode of discourse will affect the sensitivity of objective measures in discriminating between proficiency levels of Japanese EFL university students. Is it possible to have an EFL index independent of the effect of mode of discourse ? If not, does mode of discourse affect some objective measures but not others? What effects does mode of discourse have on objective measures in the assessment of overall EFL proficiency for Japanese university students ? It is obvious that in order to answer these questions, studies are needed to examine whether mode of discourse will affect the usefulness of T-unit measures for use as measures of EFL proficiency in order to see under which writing assignment condition objective measures are more valid and discriminatory indicators of Japanese EFL development.

The purpose of this study is to determine (1) the effect of mode of discourse, i.e., description and argument, on the fluency, complexity and accuracy of compositions written under time pressure by Japanese university students; and (2) whether there are significant differences between the two modes in the efficacy (i.e.,the validity and discriminatory power) of objective measures of overall EFL proficiency in Japanese university students in an attempt to establish an index of EFL development.

2. METHOD

2.1 Subjects

A total of 80 Japanese EFL students at Niigata University, Joetsu University of Education and Niigata Prefectural Women's Junior College served as subjects in this study. They consisted of 44 non-English majors (freshmen) and 36 English majors (33 sophomores and 3 graduates). No subject reported previous living experience in an English-speaking country. The subjects were divided into three ability groups: high, middle, and low. Their assignments to high, middle, and low ability groups were based on their scores on the *Comprehensive English Language Test for Learners of English* (CELT), [2] with a KR21 reliability of .789 on the structure subtest, .826 on vocabulary, and .727 on listening. The CELT scores (converted scores) ranged from 34.00 to 87.00 with a mean of 55.00 (out of a possible 100)--the equated score derived from adding the converted scores on the three subtests: structure, vocabulary, and listening. As indicated in Table 1, the high group (25 % of the total subjects), which consisted of 20 subjects, had a mean score of 69.62, with CELT scores of 61.00 and above. The middle group (50 %), with 40 subjects, had a mean score of 55.00, scoring from 46.67 to 60.67. The low group (25 %), with 20 subjects, had a mean score of 40.37, with CELT scores less than 46.00. One-way analyses of variance (ANOVAs) found significant differences among the three proficiency levels in

20

their average scores on the CELT ($F(2,77)=186.85$, $P< .001$).

2.2 Procedure

Each student wrote in two modes of discourse, i.e., description and argument. These two modes were chosen since description was shown to be least syntactically complex and argument most syntactically complex in previous studies (Perron 1977; Crowhurst and Piche 1979). In order to counterbalance the order of presentation across the subjects, they were randomly assigned to either argumentation-first or description-first tasks. The topics in each mode differed; they were, therefore, not controlled.

The instructions for the description assignment were to describe in written form a series of pictures in twenty minutes (see Appendix).[3] For the argument assignment, students were asked to write in thirty minutes on the topic of whether they would prefer to live in a large city or in a small town. The topic and time limit were equivalent to those in Larsen-Freeman's (1978) study. However, Larsen-Freeman's procedure was used with slight modifications. That is, the essay length was not assigned. Students were given the following instructions: "Begin with the sentence 'I prefer to live in a large city/small town.' Then give reasons why you prefer to live in a large city/small town and develop your main idea." Students were told to write as much as possible. No dictionaries were permitted. No subject reported previous writing on this topic in either English or Japanese.

Whereas, in previous studies, subjects wrote in different modes and were given the same amount of writing time or were allowed to write at their own speed, the subjects in the present study were given different restricted time limits to write in each mode: they were given 20 minutes for description and 30 minutes for argument. The reason the differing time allowances were chosen for subjects writing in the different modes was that the degree of time pressure differs across writing tasks. Time limits were determined on the basis of the results of the questionnaire filled out by

students in a preliminary study. They reported that 20 minutes for description and 30 minutes for argument were not enough to complete the tasks adequately. It was assumed that it would be better to collect the data by differentiating time limits depending on the difficulty of writing tasks in the present study. A restrictive time limit seemed beneficial in conducting this study.

2.3 Data analysis

The written compositions were analyzed using ten measures (see Table 2). Unintelligible strings or fragments(e.g., "The reason... and abound in nature") were not counted as T-units. Exclamations and parenthetical expressions were counted as T-units. The researcher and another Japanese rater, who had had practice with the segmentation system, independently segmented 12 descriptions (15 % of a total of 80) and 12 arguments (15 %) into T-units and dependent clauses and calculated them. Interrater reliability on the number of T-units was .996. All other writings were segmented by the researcher. Following Larsen-Freeman and Strom's (1977) procedure, this study counted as error-free any T-unit that was perfect in all aspects, including spelling and punctuation. Errors in each of the students' compositions were identified independently by two native English speakers and the researcher. When differences of opinion in the identification of errors occurred, another native English teacher was asked for his or her judgment.

For word count, numbers (e.g., 5,000), and proper names (e.g., New York) were counted as one word. Counted as one word were compound nouns written as two words by students (e.g., police man, country side). The following words were counted as two words: contractions (e.g.,I'm, don't), titles plus names (e.g., Mr. Smith), and hyphenated words (e.g., part-time).

To examine the validity of objective measures of language proficiency, Pearson product moment correlations were used. 2 (mode of discourse) by 3 (level of proficiency) ANOVAs were

22

Table 1. Means and SDs of the CELT

Group	N	Mean	SD
low	20	40.37	3.25
middle	40	55.00	3.79
high	20	69.62	7.01
all	80	55.00	11.36

Table 2. Correlations between objective measures and CELT scores

Objective measure	Description	Argument	CR
Total # of words per composition	.602**	.523**	.37
Total # of T-units	.230*	.217	.11
Mean T-unit length	.496**	.585**	-.67
Mean clause length	.412**	.355**	.40
Mean # of clauses per T-unit	.313**	.339**	-.18
% of dependent clauses	.325**	.361**	-.25
Total # of words in EFTs	.647**	.564**	.66
Total # of EFTs	.533**	.510**	.17
Mean EFT length	.356**	.413**	-.42
% of EFTs	.533**	.473**	.45

* p< .05   ** p< .01

performed on the data to check for significance in differences among means and to test the discriminatory power of measures.

3. RESULTS

3.1 Correlations

As can be seen in Table 2, the significant correlations between objective measures and the CELT scores ranged from .230 to .647. The significant correlations were above .40 on six measures in each of the two modes of discourse. Correlating significantly with the CELT scores in both modes were five out of the ten measures: the total number of words per composition, mean T-unit length, the total number of words in EFTs, the total number of EFTs, and the percentage of EFTs.

It should be noted that the differences in the correlations between objective measures and the CELT scores were not statistically significant in a comparison of the descriptive mode and the argumentative mode, although five out of ten measures showed slightly higher significant correlations above .40 in description than in argument. That is, the validity of ten objective measures did not differ significantly between modes of discourse.

3.2 ANOVAs

The means and standard deviations for the objective measures are reported in Table 3. Since the total number of T-units had the lowest correlation with CELT scores (r= .230 for description; r=

Table 3. Means, SDs and the results of ANOVAs for each of the objective measures

Objective measure	Group	Mode Description Mean (SD)	Mode Argument Mean (SD)	Mode (A) F ratio	Proficiency level (B) F ratio	A X B Interaction F ratio
Total # of words	low	95.40(25.49)	139.80(47.02)	86.48**	16.11**	5.06*
	middle	122.60(24.96)	151.43(43.07)			
	high	144.60(33.03)	212.60(76.06)			
	all	121.30(32.42)	163.81(61.17)			
Mean T-unit length	low	7.63(1.28)	9.17(1.40)	70.95**	31.28**	4.18*
	middle	8.41(1.66)	9.37(1.72)			
	high	9.80(1.78)	12.74(1.59)			
	all	8.56(1.79)	10.41(2.12)			
Mean clause length	low	6.97(0.88)	7.47(1.42)	5.73*	12.83**	0.21
	middle	7.19(1.12)	7.50(1.28)			
	high	8.13(1.44)	8.81(1.42)			
	all	7.39(1.25)	7.82(1.47)			
Mean # of clauses per T-unit	low	1.09(0.09)	1.24(0.18)	66.21**	12.15**	2.70
	middle	1.17(0.13)	1.32(0.14)			
	high	1.20(0.13)	1.47(0.23)			
	all	1.16(0.13)	1.34(0.19)			
% of dependent clauses	low	7.87(7.08)	18.06(11.56)	67.26**	12.81**	1.17
	middle	13.54(8.95)	23.48(8.52)			
	high	15.82(8.66)	30.47(9.35)			
	all	12.69(8.94)	23.87(10.53)			
Total # of words in EFTs	low	12.90(8.83)	16.15(16.39)	6.91*	23.58**	1.42
	middle	26.08(17.85)	33.53(24.03)			
	high	47.05(20.20)	64.75(51.78)			
	all	28.03(20.74)	36.99(36.52)			
Total # of EFTs	low	2.00(1.48)	2.10(2.05)	0.05	18.37**	0.04
	middle	3.68(2.58)	3.90(2.39)			
	high	6.00(2.79)	5.95(4.69)			
	all	3.84(2.80)	3.96(3.35)			
Mean EFT length	low	5.65(2.71)	6.71(4.04)	17.62**	12.48**	1.39
	middle	6.94(2.88)	8.38(2.66)			
	high	8.07(1.38)	10.72(2.24)			
	all	6.90(2.68)	8.55(3.30)			
% of EFTs	low	17.08(13.95)	13.25(10.90)	1.52	18.44**	2.06
	middle	23.03(13.55)	26.01(15.81)			
	high	39.37(14.77)	32.41(16.09)			
	all	25.62(16.24)	24.42(16.37)			

Note: L=Low, M=Middle, H=High
  * p< .05   ** p< .01

24

.217 for argument), it was excluded from further analysis.

3.2.1 Mode of discourse

As indicated in Table 3, ANOVAs showed that the main effect for mode of discourse was significant for seven of the nine dependent variables: ignoring proficiency level, argument scored significantly higher than description on the total number of words per composition, mean T-unit length, mean clause length, mean number of clauses per T-unit, the percentage of dependent clauses, the total number of words in EFTs, and mean EFT length.

No significant main effects for mode of discourse were found in two measures, the total number of EFTs and the percentage of EFTs: there were no significant differences between modes on these two measures.

There were significant interactions between mode and proficiency on two of the nine dependent variables, i.e., the total number of words, $F(2, 77)=5.06$, $p< .01$; and mean T-unit length, $F(2, 27)=4.18$, $p< .05$. The significant interactions between mode of discourse and proficiency level indicated that at each of the three proficiency levels , argument scored significantly ($p< .01$) higher than description with respect to these two measures.

3.2.2 Proficiency level

As shown in Table 3, the main effect for proficiency level was significant ($p< .05$) for all nine dependent variables. Pairwise comparisons among means except the total number of words and mean T-unit length using the LSD procedure revealed that regardless of mode of discourse, the measures except mean clause length and the mean number of clauses per T-unit increased linearly and significantly with language proficiency level and discriminated between all adjacent levels: high > middle > low ,$p< .05$. Mean clause length and the mean number of clauses per T-unit increased linearly and significantly, but failed to discriminate between the low and middle groups though the high group outscored significantly ($p< .05$) both middle and low groups: high > middle = low.

There was a significant interaction between proficiency and mode on the total number of words, $F(2, 77)=5.06$, $p<$ .01. There were significant differences among proficiency levels: $F(2, 77)=$ 18.79, $p<$ .01 for description; $F(2, 77)=12.07$, $p<$ .01 for argument. Pairwise comparisons of means on the total number of words per composition revealed that in description, the total number of words discriminated significantly between all three adjacent proficiency levels: high > middle > low, $p<$ .05; however, in argument, the total number of words did not distinguish between the low and middle groups although it distinguished significantly ($p<$ .05) between the low and high groups and between the middle and high groups : high > middle = low.

The interaction effect of proficiency and mode on mean T-unit length was also significant, $F(2, 77)=4.18$, $p<$ .05. There were significant differences among proficiency levels: $F(2, 77)=10.83$, $p<$ .01 for description; $F(2, 77)=31.63$, $p<$ .01 for argument. Mean T-unit length discriminated significantly ($p<$ .05) between the low and high groups and between the middle and high in each mode although it failed to discriminate between the low and middle groups in either mode of discourse. For the significant inter-action between proficiency and mode for mean T-unit length, there was a slightly greater variance or dispersion (in scores) in argument than in description in mean T-unit length between the low and high groups and between the middle and high groups.

The interaction effect of proficiency and mode on all dependent variables except two measures (the total number of words and mean T-unit length) was not significant.

4. DISCUSSION

Several features of this study merit discussion.

4.1 Mode of discourse

The effect of mode of discourse was clear. Regardless of proficiency level, there were significant differences between

26

modes on all measures except the total number of EFTs and the percentage of EFTs. In the total sample, when comparing argument and description, argument produced longer writing, more syntactically complex writing (longer T-units, longer clauses, a greater number of clauses per T-unit, and a higher percentage of dependent clauses), a greater number of words in EFTs, and longer EFTs. The finding that argument produced longer T-units supports the findings of several earlier studies in L1 (San Jose 1972; Perron 1977; Crowhurst and Piche 1979) and in L2 (Dvorak 1987). It may be that the nature of argument causes students to produce a high degree of subordination, which lengthens T-units (Crowhurst 1978: 92).

Surprisingly, for the total sample, argument did not produce less accurate language than description in terms of the percentage of EFTs: the percentage of EFTs was not significantly lower for argument than for description. This finding differs from that of Dvorak (1987), who found that argument produced less accurate language than narration. One likely explanation for the finding in the present study is that description was as considerably demanding for the students as argument: the description task involved vocabulary that was not accessible to students, and this provoked errors. Moreover, another explanation is that our criteria for judging a T-unit to be error-free were more stringent than Dvorak's (1987), where spelling and punctuation were disregarded as criteria in considering a T-unit to be error-free.

4.2 Proficiency level

As in previous studies, the main effect of proficiency level indicated that in the total sample, fluency (the total number of words), syntactic complexity (e.g., mean T-unit length) and accuracy (e.g., the percentage of EFTs) increased linearly and significantly with levels of language proficiency. Regardless of mode, the percentage of dependent clauses and all error-free measures discriminated between all adjacent levels. However, it

was found that mean clause length did not discriminate between all adjacent proficiency levels. The finding for mean clause length confirms that of previous research involving Japanese EFL students (Hirano 1990; (in press); Kadota 1990).

The interaction effects are perhaps the most interesting findings. It is the presence of the significant interactions of proficiency and mode in the total number of words and mean T-unit length measures that is of interest. It seems important to note that the discriminating power of only these two measures was different between the two modes. With respect to the total number of words, description, which distinguished between all adjacent levels of proficiency, was superior to argument in discriminating between all adjacent levels. If the total number of words produced by students under restricted time pressure is used to gauge overall EFL proficiency, description rather than argument appears to be suitable as a writing task. Why didn't the total number of words discriminate between all adjacent levels in argumentation? One likely explanation is that in argument, where students write without restricted word limits and can avoid vocabulary that is not available to them, even the low group was able to produce more vocabulary in this mode, so that the difference (variance or dispersion of the score in the total number of words) between the low and middle groups was smaller in argumentative writing than in description. As a result, the difference between the low and middle groups was not statistically significant in argument. Furthermore, the low group may have lacked the vocabulary required in description, a task which controls the content, eliminating variation in writing. Thus, it seems plausible that the difference in the total number of words between adjacent levels of proficiency might have more clearly reflected different proficiency levels of the subjects in description than in argument.

More interesting is the fact that the interaction between

28

proficiency and mode was significant with respect to mean T-unit length. This finding means that for mean T-unit length, argument is a superior method for assessing proficiency. That is, concerning mean T-unit length, there were slightly greater differences between the low and high groups (3.57 words per T-unit) and between the middle and high groups (2.87 words per T-unit) in argument than that between the low and high groups (2.17 words per T-unit) and that between the middle and high groups (1.39 words per T-unit) in description, although this measure failed to discriminate between the low and middle groups in each mode. One explanation for the greater differences for argument is that argument tends to have writers make maximum use of their syntactic resources. Crowhurst and Piche (1979) suggest that studies of syntactic development should be based on "writing which requires subjects to make maximum use of their syntactic skill. The evidence suggests that argumentative writing is one such kind of writing" (p.108).

In view of the evidence that significant interactions between proficiency level and mode were found for the total number of words and mean T-unit length, these two measures are mode-sensitive, and are therefore of limited usefulness as an index of Japanese EFL development unless the mode of discourse factor is taken into account.

With respect to mean T-unit length, the findings conflict. In the argumentative mode, Hirano (1990; (in press)) found mean T-unit length discriminated between all adjacent levels of proficiency, while the present study found that mean T-unit length did not operate in the same way between levels although it showed significant correlations with the CELT scores(r= .585)in Table 2.

Finally, the absence of significant differences between the modes in correlations with the CELT scores and that of interactions of proficiency and mode in any of the four error-free objective measures should be noted. The results suggest that with

the exception of mean EFT length, EFT measures, i.e., modified T-unit measures, appear to be more effective as indicators of Japanese EFL proficiency than unmodified T-unit measures (e.g., mean T-unit length). The lack of interactions indicates that no error-free measures were affected by mode of discourse in terms of their discriminatory power. Furthermore, regardless of mode of discourse, all these measures discriminated between all adjacent proficiency levels. With the exception of mean EFT length, these error-free measures showed correlations which were above .04 with CELT scores. With the exception of mean EFT length, the superiority of error-free measures over unadapted T-unit measures agrees with past ESL research (Larsen-Freeman and Strom 1977; Larsen-Freeman 1978, 1983; Vann 1979).

The higher correlation with CELT scores than in other measures, in addition to the absence of significant differences between the two modes in terms of discriminatory power indicated that the total number of words in EFT was the most powerful and appropriate indicator of Japanese EFL proficiency. It showed the highest significant correlation with the CELT score in description. The great sensitivity of the total number of words in EFTs is confirmed by the findings of Khered (1984).

One possible reason for the lack of significant interaction between proficiency and mode on all error-free measures may be that subjects had to write in each mode under restricted time pressure. Given longer periods of writing time, the results might have been different. Further research is needed to investigate the effect of mode of discourse on the efficacy of objective measures when students are given longer periods of time.

5. SUMMARY AND CONCLUSION

This study was designed to investigate (1) whether mode of discourse would have significant effects on the fluency, complexity and accuracy of writing produced by Japanese EFL

30

university students; and (2) whether significant differences in the effectiveness (i.e., the validity and discriminatory power) of objective measures related to T-units would occur between modes of discourse, in order to see which mode is superior in which measure when assessing proficiency.

The major findings of this study are summarized as follows:

(1) In the total sample, mode of discourse had significant effects on the fluency (the total number of words), complexity (unmodified T-unit measures) and accuracy of the writing of the students. Regardless of proficiency level, argument produced significantly more complex writing (longer clauses, a greater number of clauses per T-unit, a higher percentage of EFTs), a greater number of words in EFTs, and longer mean EFT length than did description. At each level of proficiency, argument produced significantly longer writing and longer T-units than description.

(2) Whereas the differences between modes in correlations between objective measures and CELT scores were not significant, there were significant differences between modes in discriminatory power with respect to the total number of words and mean T-unit length. Mode of discourse significantly affected the discriminatory power of the two measures. Thus, there is clear need to control mode of discourse in studies involving the establishment of an index of EFL development.

1. Concerning the total number of words, description was superior to argument in discriminating between levels of proficiency.

2. With respect to mean T-unit length, argument was superior to description in discrimination of proficiency levels.

(3) With the exception of mean T-unit length, error-free objective measures, or measures that take the absence of errors into account, appear to be more suitable for assessing Japanese EFL proficiency.

A great deal more research like this is necessary before these findings can be generalized to include all Japanese EFL students.

The present study raises some questions that should be addressed in future studies. For example, what effect does the amount of prior writing experience exert on objective measures? Does the effect of mode of discourse on objective measures differ between students under communicative English instruction and those who have had formal instruction which focuses on correctness? Would it be more beneficial to construct an index of EFL development based on coherence/cohesion of writing or features of communicative discourse rather than surface syntactic features ?

Larsen-Freeman (1983) states that "the questions involved in the construction of an index are formidable" (p.301). However, measures based on T-unit analysis, with the advantages of objectivity, and ease of application, would be very beneficial to teachers and researchers as an aid in assessing EFL proficiency.

The results of this study offer some insight into the search for a developmental yardstick by which to gauge global EFL proficiency. The results of the present study need to be supported by further research in which mode is not confounded with topic and data is gathered from more than a single sample of writing from individual students who have received practice at writing under time pressure.

NOTES

1. A simple or a complex sentence constitutes one T-unit, while a compound sentence consists of more than one T-unit. For example,the sentence "No work has been done on this theme, / so it is obvious the results of the study will be beneficial to teachers" constitutes 2 T-units and involves 11 words/T-unit.

32

2. Harris, D. and L.Palmer. *CELT--A Comprehensive English Language Test for Learners of English: Form A.* New York: McGraw Hill Book Company, 1986.
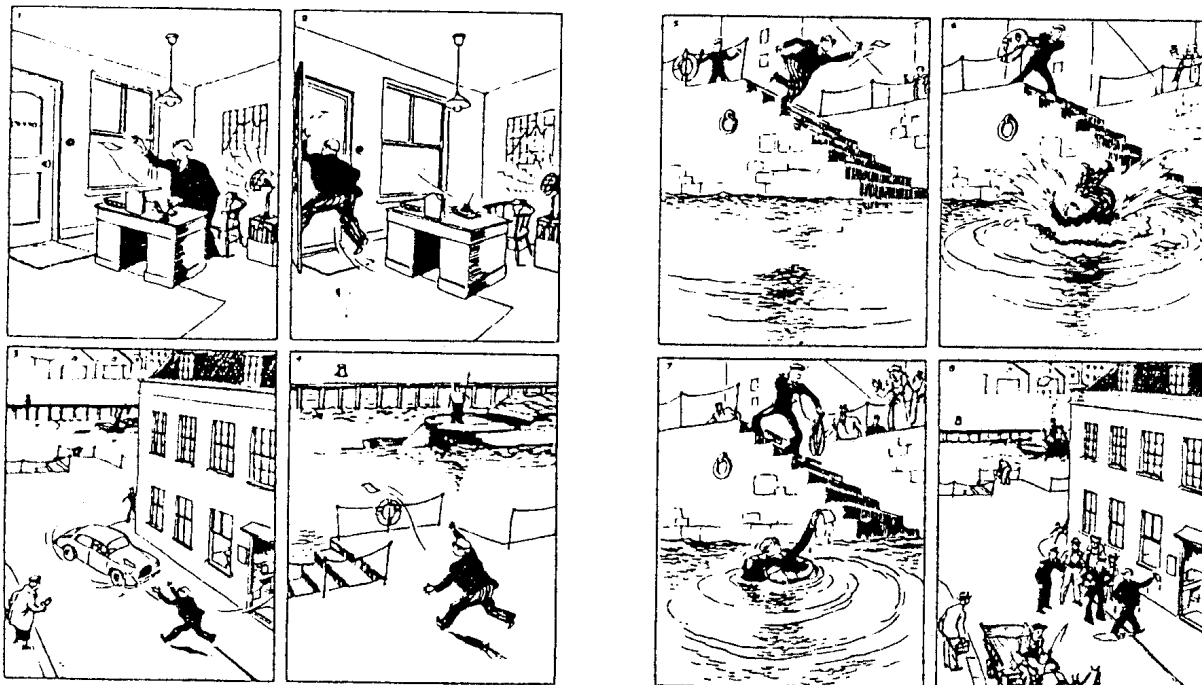3. Hill, L.A. *Picture Composition Book.* London: Longman, 1960. 26-27 (Picture Story 12)

REFERENCES

Cooper, T."Measuring written syntactic patterns of second language learners of German." Journal of Educational Research 69 (1976): 176-183.

Crowhurst, M. The Effect of Audience and Mode of Discourse on the Syntactic Complexity of the Writing of Sixth and Tenth Grades. Unpublished Ph. D Dissertation. University of Minnesota, 1978.

---. "Syntactic complexity in narration and argument at three grade levels." Canadian Journal of Education 5 (1980): 6-13.

---. and G. Piche. "Audience and mode of discourse effects on syntactic complexity in writing at two levels." Researchin the Teaching of English 13 (1979): 101-109.

Dvorak, T. "Is written FL like oral FL?" In Cross-Linguistic Series on Second Language Research. Ed. R. Anderson. Rowley, MA: Newbury House, 1987. 79-91.

Harrington, M. "The T-unit as a measure of JSL oral proficiency." Bulletin of the ICU Summer Institute in Linguistics 19 (1986): 49-56.

Harris, D. and L. Palmer. CELT--A Comprehensive English Language Test for Learners of English: Form A. New York: McGraw Hill Book Company, 1986.

Hill, L.A. Picture Composition Book. London: Longman, 1960. 26-27.

Hirano, Kinue. "Gengo nouryoku no kyakkanteki shihyou no datousei-- nihonjin EFL daigakusei no bawai" (The validity of objective measures of languageproficiency in Japanese university students). Bulletin of Joetsu University of Education 9 (1990): 65-77.

---. "The effect of audience on the efficacy of objective measures of EFL proficiency in Japanese university students." ARELE 2 (in press).

Hunt, K. Grammatical Structures Written at Three Grade Levels. NCTE Research Report No. 3. Urbana, IL: The National Council of Teachers of English, 1965.

---. "Syntactic maturity in schoolchildren and adults." Monographs of the Society for Research in Child Development 35 (1970): 1- 67.

Kadota, S."Sakubun hyouka ni okeru kyakkan teki shihyou no datousei"(The validity of objective measures in evaluating writing). Eigo Eibun Gaku Shinchou(New Current International)

(1990):339- 350.

Kameen, P. "Syntactic skill and ESL writing quality." In On TESOL '79. Eds. C. Yorio et al. Washington, D.C.:TESOL, 1979. 343-350.

Khered, M. Measuring the Syntactic Maturity in the Written English of Arab Students at Four Proficiency Levels and Establishing EFL Index of Development. UnpublishedPh. D. Dissertation. University of Kansas, 1983.

Kotani,Y. etal. "Astudy of the measurement of EFL writing--can we specify an analytic scoring item which shows high correlation with impressionistic scoring." JACET Bulletin 20 (1989): 17-35.

Larsen-Freeman, D. "An ESL index of development." TESOL Quarterly 12(1978): 439-447.

---."Assessing global second language proficiency."In Classroom-Oriented Research in Second Language. Eds. H. W. Seliger & M. Long. Rowley, Mass: Newbury House, 1983. 287-304.

---. and V. Strom." The construction of a second language acquisition index of development." LL 37 (1977): 123-134.

Monroe, J. H. "Measuring and enhancing syntactic fluency in French." The French Review XLVIII (1975): 1023-1031.

Perron,J. Written Syntactic Complexity and the Mode ofDiscourse. Eric Document No. ED 139 009, 1977.

San Jose, C. Grammatical Structures in Four Modes of Writing at Fourth Grade Level. Unpublished Ph.D. Dissertation.Syracuse University, 1972.

Scott, M. and G. Tucker."Error analysis and English-language strategies of Arab students." LL 24 (1974): 69-97.

Smith, W. and B. Swan. "Adjusting syntactic structures to varied levels of audience." Journal of Experimental Education 46 (1978): 29-34.

Tomita, Y. "T-unit o mochiita koukousei no jiyuu eisaku bun noryoku no sokutei (Assessing the writing ability of high school students with the use of T-units). Step Bulletin 2 (1990): 14-28.

Vann, R."Oral and written syntactic relationships in second language learning." In On TESOL '79. Eds. C.Yorio et al. Washington, D.C.: TESOL, 1979. 322-329.

34

APPENDIX

Describe the picture story in English in 20 minutes.

(Hill 1960: 26-27)