An Assessment of English Language Learner Writing

John N. Wendel

Abstract

This paper presents the results of an assessment of English-as-a-second-language learner writing. Narratives were collected from first-year university students at the beginning and at the end of the academic school year at Bunkyo Gakuin University. A sample of 72 narratives (36 from each collection period) were randomly selected and evaluated for both measures of development and achievement of course goals. Although the results confirmed statistically significant differences between the April and December narratives, the gains were modest.

INTRODUCTION

An assessment of narratives written by students enrolled in the first-year writing course at Bunkyo Gakuin University was undertaken in order to track developments in the students' written communication skills and to evaluate the effectiveness of the first-year university writing course. All students in Bunkyo Gakuin University's Faculty of Foreign Languages are obliged to take the writing course as a part of their core English language curriculum requirements. In total, students must complete two years of English speaking and reading, and three years of writing. The writing and speaking courses are conducted entirely in English, while the reading courses are largely given in Japanese with English texts. The writing and reading classes each meet once a week for 90 minutes; the speaking classes, twice each week for a total of three hours.

In his discussion on assessments of second-language learner writing, Hughes (2003) lists three considerations that should be taken into account in the assessment procedure.

1. We have to set up writing tasks that are properly representative of the population of tasks that we should expect the students to be able to perform.

2. The tasks should elicit valid samples of writing (i.e., which truly represent the students' ability).

3. It is essential that the samples of writing can and will be scored validly and reliably. (Hughes, 2003, p.83)

As the final goal of the three year writing curriculum is to have students write a proper academic essay, the ability to compose a simple narrative is certainly within the purview of expected sub-skills. Also, a narrative was thought to be the most effective vehicle for eliciting a sample that is representative of the students' skills because of its intuitive appeal (story-telling figures significantly in any conversation among friends) and because few, if any, of these students have prior training in academic writing, in either Japanese or English. For this assessment, the learners were provided with a prompt in the form of an eight-frame cartoon which meant that the individual learner narratives could be easily compared, not only with each other, but across administrations because the same prompt was used in both April and December. Every effort was made to ensure that the scoring was valid and reliable.

The question arises as to what in the narratives should be measured. For this study, the main focus is not the students' writing proficiency, but their development along several well established dimensions: fluency, accuracy, and complexity. In their treatment of measures of SLA writing, Wolfe-Quintero, Inagaki, and Kim (1998) point to the many studies that have used such measures to evaluate the effectiveness of instruction.

"In second language acquisition research studies, developmental measures of fluency, accuracy, and complexity have also been used as dependent measures for examining the effect of a pedagogical treatment on either oral or written language use." (Wolfe-Quintero, Inagaki, and Kim, 1998, p.3)

There is also something very appealing about the tripartite indices of fluency, accuracy, and complexity for investigating written (and oral) second-language studies because, as Wolfe-Quintero, Inagaki, and Kim (1998) pointed out, this classification captures common assumptions about the routes of development in second-language writing.

1) "Second language learners write more fluently, or write more in the same amount of time, as they become more proficient.

2) Second language learners write more accurately, or produce fewer errors in their writing, as they become more proficient.

3) Second language learners write more grammatically and lexically complex sentences as they become more proficient." (Wolfe-Quintero, Inagaki, & Kim, 1998, p. 3)

Researchers have used these measures of development in many areas in second-language studies, including the evaluation of writing programs (e.g., Carlisle, 1989), task type (e.g., Foster and Skehan, 1996), and off-line planning (e.g., Crookes, 1989; Ortega, 1999; Wendel, 1997, 1999).

Although certain of the measures adopted in the present study target specific first-year

Writing course objectives, the majority of the measures can, in fact, be interpreted more broadly as indices of development (or lack thereof) in the students' written communication skills that reflect, to an indeterminate degree, the effectiveness of the English skills curriculum as a whole. Thus, this assessment is also evaluative of the curriculum as well as the Writing course. However, because of the many limitations in this study (two short writing tasks, small sample size, etc.— see Appendix A), all findings must be interpreted judiciously.

This assessment addresses the following questions:

1. What changes are evident in students' written communication skills after an eight-month interval?

2. Does the extent of the changes in students' written communication skills depend on initial skill level or are the changes uniform across skill levels?

3. Do all students evidence change in their written communication skills in a positive direction?

4. What is the relationship between the TOEIC Bridge scores and the measures of students' written communication skills?

5. What insights do text analyses of the narratives contribute to our understanding of developments in the students' written communication skills?

PROCEDURE

In April 2006[1] , students were given 20 minutes to write a story based on an eight-frame cartoon (see Appendix B) during their first regularly scheduled Writing class; the same task was given again (using the same cartoon) in their last Writing class of the calendar year in December 2006.[2]   The students were not aware that they would be asked to write on the same material twice. Of the 230 narratives collected in April, 36 were selected at random (three from each of 12 university Writing classes) and then, after the December administration, paired up with their December authors' narratives, making a total of 72 narratives (a sample size of about 15% of the total) for evaluation[3] . All identifying marks (names, student ID numbers) were masked from the original narratives and two copies of each narrative were made to prepare two sets of the 72 narratives. Each set was thoroughly mixed and then given to two raters[4] for evaluation. The raters were not aware of the authorship of the narratives, the classes from which the narratives were selected, or the time (i.e., April or December) the narratives were written.

Each narrative was evaluated on the nine measures below. Global is the only holistic scale, Discourse and Mechanics are analytic scales, and the remainder (Sentences, Words, etc.) are quantitative measures. Rubrics for Global, Discourse, and Mechanics are found in Appendix C.

1. Global (Glob)— the principal measure of interest in this assessment, Global is a holistic measure on a six-point scale (see Appendix) that evaluates each narrative for its overall worth as a piece of story-writing and includes such notions as coherence, cohesion, organization, productivity, sophistication of expression and range of vocabulary, grammatical complexity and accuracy, naturalness and appropriateness. Global does not include considerations of spelling, formatting, paragraphing, margination, capitalization, and punctuation.

2. Discourse (Dis)— a six-point analytic measure which specifically targets the use of discourse markers such as *then, after, next, before, first, second, and then*; sentence-initial dependent clauses such as *Before leaving the restaurant, After he said that*; coordinate clauses with serial verbs (V, V, *and* V), and other devises (e.g., mid-sentence adverbial clauses, use of connecting words such as *so, suddenly*, or *just*)— in short, the sorts of devices that explicitly move the narrative (action or event structure) forward in time or link events and characters within or across sentences. This is most generally understood as a consideration of discourse grammar.

3. Mechanics (Mech)— a five-point analytic measure that evaluates each narrative for spelling, formatting, paragraphing, margination, capitalization, and punctuation.

4. Sentences (S)— the number of sentences in each narrative.

5. Words (W)— the sum of the number of words in each sentence of the narrative.

6. Clauses/Sentence (Cl/S)— the average number of clauses for each sentence, arrived at by dividing the number of clauses by the number of sentences for each narrative.

7. Words/Sentence (W/S)— the average number of words for each sentence, arrived at by dividing the number of words by the number of sentences for each narrative.

8. Percent error free sentences (PerEr)— the percentage of error free sentences in a narrative, arrived at by dividing the number of error free sentences by the total number of sentences for each narrative.

9. Error free sentences (ErFr)— the number sentences in a narrative that are free of grammatical errors. This measure does not consider errors in appropriateness, spelling, capitalization, punctuation, or how effectively a sentence is contextualized within the narrative.

The above measures were grouped into four areas of interest:

- ・ Criterion measures: Global, Discourse, Mechanics
- ・ Grammatical Complexity measures: Words/Sentence, Clauses/Sentence
- ・ Productivity measures: Words, Sentences
- ・ Grammatical Accuracy measures: Error free sentences, Percent error free sentences

The Criterion measures are so named because these three measures assess three first-year Writing course objectives: Global — to develop the overall coherence and organizational properties of a piece of writing; Discourse — to develop the students' use of discourse markers and cohesive devices; Mechanics — and to instruct students in appropriate mechanics and formatting procedures for essay writing in English.

Raters were free to add a '+' to a criterion score for a narrative that was found to be intermediate between two points. In such cases, 0.5 was added to numerical score: thus a mark of 2+ was entered as 2.5, 3+ as 3.5, etc. The raters' scores for the criterion measures were averaged; all other measures were arrived at consensually by the raters. Inter-rater reliability for the criterion measures was calculated using the Pearson Product Moment correlation statistic: Global (r=0.83); Discourse (r=0.73); Mechanics (r=0.85).

RESULTS

Question #1: Changes from April to December

To address the first research question (What changes are evident in students' written communication skills after an eight-month interval?), the 72 narratives were evaluated on all the measures described above. These measures will be tested for statistical significance: the nine measures (i.e., Glob, Dis, Mech, Cl/S, W/S, S, W, PerEr, ErFr) are the dependent variables; the time the narratives were written is the independent variable with two levels, April and December. The null hypothesis (H_0) of no differences across April and December will be tested with a repeated-measures multivariate analysis of variance procedure (SPSS Release 8.0, 1998).

As Table 1 below shows, the means for the December narratives were higher in every measure than those of the April narratives indicating a favorable trend towards improvement and/or development in each area. (The 'Change in mean' row in Table 1 shows the difference between the December and April means. All the values are positive.) Some of the differences are quite obviously negligible (e.g., PerEr=0.4) while others are meaningful (e.g., Glob=0.77 or

Table 1. Descriptive statistics for April and December narratives (N=72)

	Glob	Dis	Mech	S	W	Cl/S	W/S	PerEr	ErFr
April (n=36)									
Mean	**3.64**	**3.19**	**2.96**	**10.8**	**81.7**	**1.45**	**7.8**	**42.7**	**4.69**
Sd Dev	0.75	0.98	0.99	5.2	37.0	0.24	1.5	25.1	3.85
December (n=36)									
Mean	**4.41**	**3.93**	**3.99**	**14.1**	**113.7**	**1.61**	**8.2**	**43.1**	**6.14**
Sd Dev	0.50	0.92	0.77	4.1	30.5	0.26	1.5	20.0	3.62
Change in Mean	**0.77**	**0.74**	**1.03**	**3.3**	**32.0**	**0.16**	**0.4**	**0.4**	**1.45**

W=32).

　　　　Table 2 presents the correlations between all the dependent measures. Global is significantly correlated with all the other measures. That Global is moderately correlated with Discourse (r=0.76) is not surprising as Discourse is a measure of devices that are implicated in story cohesion and grammatical complexity, both factors being strongly represented in Global. The highest correlation in the table, Sentences and Words (r=0.88), is simply a logical outcome of the measures themselves: the more sentences you have in a composition, the more words you tend to have. In fact, many of the higher correlations can be explained in terms of length (i.e., the overall length of a given narrative). Length figures to some significant extent in many measures: Global, because a successful story will have to be necessarily longer to be a candidate to be evaluated better; Discourse, because the more sentences you have, the greater are the opportunities to use a variety of discourse markers appropriately, and so on.

Table 2. Matrix of Pearson correlations (N=72)

	Glob	Dis	Mech	S	W	Cl/S	W/S	PerEr	ErFr
Glob	1.00								
Dis	0.76**	1.00							
Mech	0.61**	0.50**	1.00						
S	0.54**	0.32**	0.37**	1.00					
W	0.67**	0.57**	0.45**	0.88**	1.00				
Cl/S	0.51**	0.60**	0.36**	0.13	0.50**	1.00			
W/S	0.28**	0.45**	0.16	-0.28**	0.19	0.71**	1.00		
PerEr	0.37**	0.918	0.16	0.06	0.06	-0.14	-0.37	1.00	
ErFr	0.57**	0.29**	0.34**	0.70**	0.60**	-0.03	-0.19	0.71**	1.00

　**p<0.01 (two-tailed)
　*p<0.05 (two-tailed)

　　　　This raises the issue of why all of the measures were included in this analysis, a few of which are strongly correlated (for example, Global and Discourse or Words/Sentence and Clauses/Sentence), if they seem to be measuring the same thing? The fact is, we want to know if the students, after eight months of English language instruction representing about 200 hours classroom time, are writing overall longer narratives, whether this be manifested in terms of higher numbers of shorter sentences (resulting in high numbers of sentences, but proportionately lower numbers of words) or fewer longer sentences (resulting in lower numbers of sentences, but proportionately higher numbers of words). The same goes for our interest in grammatical complexity: we want to know if the sentences are overall longer in terms of both numbers of words (i.e., words/sentence) and numbers of clauses (i.e., clauses/sentence): the former captures overall sentence length including such additions as adjectives, prepositional phrases, noun

phrases, adverbial phrases while the latter focuses more narrowly on grammatically dependent strings that include finite or nonfinite verb elements.

　　　　For protection against unequal variances among the pairs of dependent variables (e.g., April, Words=1372.7; December, Words=931.3) and multicollinearity (e.g., Sentence and Words, r=0.88), the comparatively robust Pillai's Trace test statistic was used for the MANOVA procedure along with a conservative overall significance level of $P=0.01$. All nine measures above were analyzed statistically using a repeated-measures, within-subjects MANOVA design; the result was statistically significant (F=8.45, $P=0.0005$) and the null hypothesis H_0 of no differences between the April and December narratives was rejected (see Table 3).

Table 3. Multivariate test result (N=36)

Source	Pillai's trace	F	df	p
Time	.738	8.450	9	.0005*

*p<.01

　　　　Subsequent univariate tests (see Table 4) resulted in statistical significance for seven of the nine measures: Global (F=41.77, p=.0005), Discourse (F=15.09, p=.0005), Mechanics (F=32.39, p=.0005), Words (F=31.58, p=.0005), Sentences (F=19.02, p=.0005), Error free sentences (F=9.22, p=.004), and Clauses per sentence (F=11.16, p=.002).　A nonsignificant result was found for both Words/Sentence (F=2.12, p=.155) and Percent Error Free sentences

Table 4. Univariate test results for Time (N=36)

Source	Sum of squares	df	Mean square	F	p
Global	10.70	1	10.695	41.77	.0005*
Error	8.96	35	.27		
Discourse	9.753	1	9.753	15.09	.0005*
Error	22.62	35	.65		
Mechanics	19.27	1	19.27	32.39	.0005*
Error	20.82	35	.60		
Sentences	203.35	1	203.35	19.02	.0005*
Error	374.15	35	10.69		
Words	18432.00	1	18432.00	31.58	.0005*
Error	20426.00	35	583.60		
Clause/S	.487	1	.487	11.164	.002*
Error	1.526	35	.0436		
Words/S	2.21	1	2.21	2.16	.115
Error	36.60	35	1.05		
Percent Error	2.16	1	2.16	.009	.926
Error	8560.01	35	244.57		
Error Free	37.56	1	37.56	9.23	.004*
Error	142.44	35	4.07		

*p<.01

(F=0.009, *p*=.926).

A significant value is a necessary initial step in the statistical procedure, but just as important is determining the amount of the variance in the dependent variables that is attributable to the independent variable, in this case Time. This is one of several evaluations of the meaningfulness of the difference across Time levels. The eta squared statistic gives this information and is directly translatable into a percentage figure (Hatch & Lazaraton, 1991 pp. 354-355). Eta squared was calculated for each of the seven statistically significant results above: Global (0.544 or 54.4%), Dis (0.301 or 30.1%), Mech (0.481 or 48.1%), Sentence (0.352 or 35.2%), Words (0.474 or 47.4%), Clause/S (0.242 or 24.2%), and Error Free (0.209 or 20.9%). Thus we can say that the 54.4% of the variability found in, for example, Global is accounted for by Time; in Discourse, 30.1%, etc. These figures are all on the high end and demonstrate healthy relationships between the dependent variables and the effect of Time.

The gains in the criterion scores, if not impressive, are certainly worthy of note. On average, both Global and Discourse increased around three-quarters of a point (0.77 and 0.74 respectively), while Mechanics rose a full point (1.03). Taking our cue from the Global scale rubric (see Appendix for rubrics for the three criterion measures), the basic story structure of the December narratives was overall more transparent, the vocabulary more effectively deployed, and sentences more skillfully crafted than those of the April narratives. The use of discourse markers improved from Åesome use of discourse markers' to 'use of discourse markers essentially appropriate'; and the formatting and mechanics of the December narratives essentially followed all conventions, where 'infrequent slips do not cause a problem in the reading experience.'

Turning our attention to the quantitative measures, perhaps the most obvious gains were found in the productivity: the narratives were longer both in terms of sentences (from 10.8 to 14.1 — see Table 1) and in terms of numbers of words (from 81.7 to 113.7). This suggests students had achieved a greater confidence and facility with the use of written English. There was a small, but significant, gain in terms of grammatical complexity: the difference between the April and December narratives was 0.16 on average. This tendency towards greater complexity may be another confidence factor: students are now bolder in the use of sentence embedding devices. (Or alternatively it may be attributable to students' overall greater use of reported or direct speech in the December narratives. Analysis would show.) The other complexity measure was nonsignificant, of course, and the average increase of less than half a word per sentence (0.4 — see Table 1) is nothing more than background noise.

No gains were realized in the grammatical accuracy measures: as can easily be determined by casual inspection, the proportion of error free to flawed sentences was

nonsignificant; and the statistically significant difference between the April and December narratives in numbers of error free sentences (4.69 versus 6.14?see Table 1) can easily be explained in terms of overall gains in sentence productivity (i.e., the students in December were writing more sentences overall, so there would be, logically, a greater number of error free sentences). Sentence-level grammatical issues, thus, remain a problem with these students.[5] Informal observation reveals that grammatical deficits are found in all areas, particularly in article use, verb morphology and verb tense/aspect considerations, use of particles and prepositions with verbal phrases, and subject/verb agreement.

Question #2: Changes from initial skill level

To address the second question, the data were examined using the students' April Global scores as a grouping variable and creating three writing skill levels: those with the low April Global scores (n=12; range: 2.50-3.25), those with intermediate April Scores (n=12; range: 3.25-3.75), and those with the high April scores (n=12; range: 4.00-5.50). The purpose of this analysis was to evaluate the changes from the initial skill level along the nine measures. Table 5 presents the

**Table 5**. Descriptive statistics for three skill groups: low, intermediate, high

April (n=36)April─High Score Group (n=12)								
Mean	4.5	4.00	3.54	13.2	105.3	1.5	8.37	57.0	7.3
Sd Dev	0.5	1.02	0.94	7.2	46.9	0.3	1.45	17.3	4.5
December								
Mean	4.6	4.35	3.98	13.8	120.8	1.8	9.01	41.2	6.0
Sd Dev	0.4	0.76	0.89	4.0	26.9	0.2	1.51	17.4	4.2
Change in Mean	0.1	0.35	0.44	0.6	15.5	0.3	0.64	-15.8	-1.3
April─Intermediate Score Group (n=12)								
Mean	3.6	3.15	2.96	11.0	78.3	1.5	7.24	43.1	4.4
Sd Dev	0.2	0.59	1.00	3.8	28.5	0.1	0.98	24.6	2.7
December								
Mean	4.4	3.52	4.06	16.1	118.3	1.6	7.56	44.4	7.2
Sd Dev	0.6	0.93	0.78	4.5	29.2	0.3	1.70	13.8	3.4
Change in Mean	0.8	0.37	1.10	5.1	30.0	0.1	0.32	1.3	2.8
April─Low Score Group (n=12)								
Mean	2.9	2.44	2.35	8.0	61.3	1.3	7.94	28.0	2.3
Sd Dev	0.3	0.56	0.69	2.4	16.7	0.2	1.83	25.3	2.3
December								
Mean	4.2	3.91	3.92	12.5	102.0	1.5	8.03	43.5	5.2
Sd Dev	0.6	0.92	0.68	3.1	34.0	0.2	1.08	27.3	3.2
Change in Mean	1.3	1.47	1.57	4.5	40.7	0.2	0.09	15.5	2.9

descriptive statistics for the three skill groups for all measures.

Interestingly, the low group, on six out of nine measures, realized the greatest change; the high group, on two of nine measures, the least change. The 'Change in Mean' row (Table 5) for each skill group shows the difference between the December and April means.

Figure 1 shows the differences in mean gains for the three criterion measures. Notice, for example, that whereas the overall (i.e., across skill groups) gain from April to December for Global was 0.77 (Table 1), the low group realized the highest gain of 1.3 over the eight-month interval, while the high group realized a negligible 0.1 gain (Table 5). In fact, the gains for all three criterion measures were small for the high group. The two measures for grammatical accuracy (Table 5) show that the high group actually performed worse in their December narratives than the April narratives, the changes in mean negative values (PreEr.=-15.8; ErFr= -1.3), while low group realized the highest gains (PerEr=15.5; ErFr=2.9).

Figure 1. Mean gains across skill levels on criterion measures
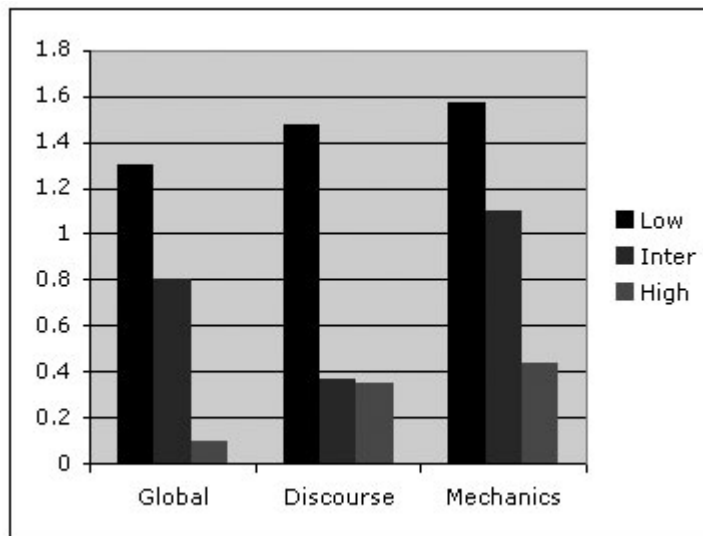
What can reasonably account for the uneven gains in skill level above? One possibility that might explain the changes in the criterion scores is that the scales themselves are biased towards the lower end of the skill level. According to this explanation, the change in, for example, the mean Global score from April to December would tend to be a numerically greater interval for a pair of texts that have a low end April Global score than a high end April score, all things being equal. Another explanation might be that there are, in the minds of the high skill group, perceived thresholds in length and innovation of a 'successful' narrative beyond which a student feels the task does not warrant the added investment of effort. This is a sort of 'been there, done that' syndrome according to which a student feels unrewarded by further exertion.

While these explanations are certainly plausible, they are not convincing. For one thing, the high skill group performed the lowest of the three groups or lower than the low skill group in all the measures except for grammatical complexity suggesting that there is more at work here than biased scales (which applies only to the criterion measures). Also, it is unlikely that these students have reached a level of sophistication in their writing where they would be influenced by any perceived threshold. The last explanation considered here is the simplest: that the needs of the higher group were not being addressed by the English language skills curriculum. This suggests that students who come to the university with initially higher level writing skills may need more challenging material and that instruction should focus on more sophisticated matter than it is at present. This strikes me as the most likely explanation.

Question #3: Direction of change in Global scores

It is also useful to track the direction of the change in the global scores. The global is, after all, the central index for the narratives, being as it is a composite evaluation. How many of the students' Global scores increased? remained the same? decreased? Another question would be, why did anyone's score remain the same? or worse, decrease?

Of the 36 students' December Global scores, four declined (see Table 6 below); all of these students were in the high skill group. Of the two students whose scores remained the same, one student was in the high skill group, the other in the intermediate group. These cases will need to be evaluated in detail to determine why the scores that declined were found in the high skill group. While small differences (as in Case #1 below?a drop of 0.25) or no differences (as in Cases #5, #6 below) may be attributed to the vicissitudes of rater judgments on subjective scales, larger declines are likely to be attributable to other factors such as a lack of incentive, motivation, drop in skill level, a failed strategy, or chance. In any case, that these declines are mainly among the high skill group is not surprising given the performance of the high skill group as a whole in the section above (see Table 5). A close text analysis of the narratives may shed some light on these cases.

Table 6. Global scores that dropped or remained the same

Case #	Apr, Dec Narrative ID	Apr to Dec	Change
1	#68, #43	4.00 to 3.75	-0.25
2	#46, #18	5.50 to 4.75	-0.75
3	#35, #9	5.25 to 4.50	0.75
4	#4, #6	4.75 to 4.00	-0.75
5	#55, #17	3.50 to 3.50	0.0
6	#7, #49	4.50 to 4.50	0.0

Question #4: TOEIC Bridge scores

In addition to the above, TOEIC Bridge scores for both April 2006 and January 2007 administrations were retrieved for 25 of the 36 students (11 students were absent for either the April or January tests: their data were not included in the TOEIC Bridge analyses). This enabled further analyses to be conducted on all dependent measures for these 25 students. Table 7 shows the descriptive statistics on the TOEIC Bridge for the April 2007 and the January 2007 administrations for these 25 students. (Note that the SEM for the TOEIC Bridge is officially published as $\pm$ 9.0.)

Table 7. Descriptive statistics for the TOEIC Bridge

	TOEIC Bridge
Ap—Mean (n=25)	137.7
Std Dev	15.3
Dec—Mean (n=25)	141.5
Std Dev5	14.0
Change in Mean	3.8

Table 8 shows the Pearson correlations between the TOIEC Bridge and the nine measures. A moderate correlation was found between the TOEIC and Global (r=0.66) and, perhaps, TOEIC and W/S (r=0.50); weak correlations were found among the other measures (e.g., TOEIC and Dis, r=0.45; TOEIC and W, r=0.40).

Table 8. TOEIC Bridge Pearson Correlations with Apr. and Dec. Measures (N=50)

	Glob	Dis	Mech	S	W	Cl/S	W/S	PerEr	ErFr
TOEIC	0.66**	0.45**	0.44**	0.21	0.40**	0.49**	0.50**	0.28	0.40**

**p<0.01 (two-tailed)

These correlations, along with the very small overall gain in TOEIC Bridge score across administrations (Change in Mean: 3.8?see Table 7), suggest that gains in dependent variables measured in this study are not matched with equivalent gains in TOEIC. Put another way, with the possible exception of Global, the TOEIC Bridge score is a poor indicator of success of the dependent measures evaluated in this study. In practical terms, this means that one cannot make the claim that because students have a weak performance on their TOEIC Bridge test, they haven't learned anything or haven't made improvements in their English over the course of the academic year. Unequivocal gains were made on several of the measures (e.g., Dis, Mech, S, and W) that are only weakly or moderately related to the TOEIC Bridge score.

Question #5: Text Analysis

To address the last question (What insights do text analyses of the narratives contribute to our understanding of developments in the students' written communication skills?), eight of the students' narratives were analyzed: one April and one December narrative from each of four students. There are many reasons for conducting a text analysis including (1) giving the reader an idea of the relationships between the measures above and the narratives themselves, (2) illustrating the types of changes found in the from April to December, (3) and picking up the slack for those other influences that simply were not included (i.e., operationalized) by the adopted measures.

Highly detailed text analyses are far superior to quantitative analyses when it comes to the very important task of teasing out relationships and patterns among the universe of influences on which a piece of writing is contingent, and uncovering the subtle shifts, resonances, and shades of difference that counting methods are simply not capable of capturing. Importantly, through text analyses, we can infer writer strategies, approaches, missteps, and trouble spots locally within a narrative that may reveal significant undercurrents in a piece of writing. In the case of these narratives, for example, a comprehensive analysis might also evaluate and compare the strategies most and least favored by writers in resolving the problem ending, both in terms of content (the story itself), but also in terms of grammatical devices used. As in Case Studies 'B', 'C', and 'D' below, many students resorted to the use of relative clauses in attempts to connect the identification of the violinist at the restaurant with the man in the last frame. Informally, we can say that the results were uneven, but always interesting. An analysis of these strategies might lead to very specific recommendations for classroom activities intended to develop further students' use of reference in general and relative clause use in particular.

The analyses included here are not comprehensive and are intended mainly to demonstrate the possibilities of this approach for further study. Following each pair of narratives, descriptive statistics have been provided (Tables 9, 11, 13, and 15) and a sequence by sequence comparison of the April and December texts (Tables 10, 12, and 14). Note the use of abbreviations below: Apr=the April text; Dec=the December text; DM=discourse marker; PP=prepositional phrase; RC=relative clause. (See Note 2 for more information about the instructions given to students before writing and the task sheet.)

CASE STUDY 'A' (NARRATIVES #30 & #48)

April story ― #30 (Global=2.50)

One day he is dating with her.

And they are drinking wine.

They listen to violine sound.

Because they interested in violine before.

　Good mood.... They closed our eyes.

And they catch each hands.

　He

They go away in front of caunter.

　She sit the bentch.　He propose her.

He succeed propose for her.　　They

December story 一 #48 (Global=4.00)

　One day Bob went to restrant with his girlfriend. When they drank alchole, bar's master asked them. "Shall I play the violin?" He played the viollin, and Bob and girlfrend holded each hands. Then, he holded her sholuder. They surrounded good mood.

　When they went out bar, they surrounded good mood. Then, they went to park. Bob propose her. Her answer is "OK". Next, they went to her house. He thought that she lives alone, but she lived with father.

Table 9. Results for measures on Case Study 'A'

	Glob	Dis	Mech	S	W	Cl/S	W/S	PerEr	ErFr	TOEIC
Apr	2.50	2.00	2.00	10	51	1.10	5.10	10.0%	1	108
Dec	4.00	4.75	4.00	11	80	1.55	7.27	18.2%	2	126

Commentary on 'A'

The December text received significantly higher scores for all three criterion measures, with gains of 1.50 for Global, 2.75 for Discourse, and 2.00 for Mechanics. Reading through the two texts, it is easy to see why. The December text coheres better, the sentences are connected with discourse markers and basic paragraphing and punctuation conventions are followed. The April text is a succession of one clause sentences, but 4 of the 11 sentences in the December text are multiclausal (note the difference in the Cl/S values: April, 1.10; December, 1.55). Whereas the April texts includes several sentence fragments, the December text does not. In terms of productivity, the April text is at the low end of all the narratives with 51 words compared with 80 words in December, a modest increase. The measures for grammatical accuracy show no meaningful difference. With the exception grammatical accuracy, the measures for the December text are markedly better.

　　　It is interesting, for this pair of texts, to extend our analysis and briefly consider content, style, and the relationship between form and meaning. Apart from following

conventional formatting procedures, the first thing we notice about December is that this writer gives the male character a name, Bob, and identifies the female character with a role, 'his girlfriend.' This approach is more engaging than the April text (an awkward first sentence that denies the characters anything other than reference to their gender: 'One day he is dating with her') and catches the reader's interest for the story that follows. The character roles are better developed and this makes a more fulfilling reading experience. When the sequences are placed side by side (as in the Table 8 below), we can see how each December story element has been rendered with greater precision and detail (e.g., sequence #1, #2, #4). Significantly, the writer gathers two or more events within the same sentence frame (e.g., #4, #5, #7) generating a higher index of syntactic complexity as mentioned above, but also resulting in more natural story-telling rhythms. In December, the writer also comes closer to resolving the problem ending. In sum, an all around better performance in December: greater control over basic syntactic devices and sentence embedding, modest use of discourse markers, and greater productivity.

Table 10. April (#30) and December (#48) story sequences

	April Story	December Story	Remarks
1	*One day he is dating her.*	*One day Bob went to restrant with his girlfriend.*	Setting, male character name, and female role is identified in Dec, but Apr only uses simple pronouns.
2	*And they are drinking wine. They listen to violine sound.*	*When they drank alchole, bar's master asked them. "Shall I play the violin."*	Punctuation aside, an accomplished three-clause sentence instead of two simple sentences. There is a new character role in Dec, that of the bar master who is also the violinist. Not just the violin music, but also the violinist plays active role in Dec story.
3	*Because they interested in violine before.*		Violin is backgrounded in Apr, but plays an active role in Dec, as in next frame.
4	*Good mood⋯. They close our eyes. And they catch each hands.*	*He played the violin, and Bob and girlfrend holded each hands. Then, he holded her sholuder. They surrounded good mood.*	Apr's fragment "Good mood" is expanded to a full sentence, "They surrounded good mood." In Dec, the music is implicated syntactically in the action of holding hands and then linked with "holded her sholuder." Apr is given two details ("close our eyes" "catch each hands") linked by coordinator "and".
5	*They go away in front of caunter. She sit the bentch. He propose her.*	*When they went out bar, they surrounded good mood. Then, they went to park. Bob propose her*	Apr has three simple sentences with no discourse marking, but Dec uses adverbial clause "When⋯" and sentence initial "Then⋯" to link discourse structure and move story forward—a natural feeling.
6	*He succeed propose for her.*	*Her answer is "OK".*	The difference is that in Dec, the writer is telling a story, whereas in Apr, the author reporting a fact.
7		*Next, they went to her house. He thought that she lives alone, but she lived with father.*	The writer comes closer to concluding the story, which includes the resolution to the story's problem ending, although the father is never identified as the restaurant violinist. Sentence initial "Next" moves action to girlfriend's home and sets up the surprise ending. What "He thought" and what he found are linked together with coordinator "but" making a three-clause construction.

CASE STUDY 'B' (NARRATIVES #19 & #44)

April story ― #19 (Global=3.25)

One day I went to the restlunt with my girl friend.

And I ask violinist to play the violin.

So, he began to play.

I hug her sholder.

We left the restlunt.

I talked with her in the park.

At the last, She came my house.

After few minutes, She left my house

An our leter, She came back with a man

The man is the violinist, who I met in the restlunt.

I was surprised and shocked.

December story ― #44 (Global=5.00)

  One day a man had dinner with his girlfriend. Some minutes later, he asked a waiter to play music. The waiter play the violen for the couple. Then, the atmosphere became better, and the atmosphere of the couple became better too. After the dinner, he brought her to a park near his house. He told her "I love you, I need you" again and again. She listend and was silent. He told her "Why don't you go my house" and he took her to his house. Two hours later, soon, her celephon was called. She talk to someone without her boyfriend. She told him "I must go out soon, but I'll be back" Don't worry" He permited her. A few minutes later, she went back his house with another man whom her boyfriends knew. The man is the waiter. Her boyfriend said his self again and again "Why? Why? Why? Why?..."

Table 11. Results for measures on Case Study 'B'

	Glob	Dis	Mech	S	W	Cl/S	W/S	PerEr	ErFr	TOEIC
Apr	3.25	2.50	2.00	11	78	1.27	57.1	72.7	8	134
Dec	5.00	5.25	4.50	16	152	1.63	9.5	68.8	11	na*

  *not available

Commentary on 'B'

With the exception PerEf, there are gains all around in this December narrative. Each criterion measure shows an increase of 1.75 points or more, which is at the high end of the range. The

productivity measures have both increased, the number of words overall nearly doubling (from 78 to 152). Both indices of complexity have risen, a boost of 0.36 for Cl/S-and this in a text having five more sentences than previously. Reading over the December text, we see that the gain in complexity is partly a result of the many bits of dialogue (four additional dependent noun clauses), and partly the result of more coordinate links (three instances of *and* linking two independent clauses: for example, 'Then, the atmosphere became better, *and* the atmosphere of the couple became better too').

It is interesting to compare the sentence initial discourse features (in this case, those that consist of more than one word) across the two texts: there are three examples in April (At the last; After few minutes; An our leter) and four in December (Some minutes later; After the dinner; Two hours later; A few minutes later) and they are remarkably similar in both form and function in spite of having been composed eight months apart. We might wonder why? These phrases are very simple, but effective devices for moving along the action in time. (Formally speaking, the most simple would be no marking at all, but letting the sequencing of the sentences do the work of moving time forward in the story?as in the April text in Case Study 'A' above. The next stage would be the one-word expression fronting a sentence such as *And* and *So* in the April narrative, and *Then* in the December narrative.) These instances represent opportunities for a skilled writer to exploit the resources in the grammar for variety and depth that extends their range of expression. A reasonable place for a student to start are the <u>December alternates</u> (see below) that are offered in place of the <u>December originals</u>. One could argue that the alternates represent a next stage in the development of a student's grammatical armamentarium.

December originals		**December alternates**
1. Some minutes later	-->	A few minutes after they ordered the meal
2. After the dinner	-->	After they left the restaurant
3. Two hours later	-->	After talking together for two hours
4. A few minutes later	-->	Minutes after she left him

Granted there is nothing at all elegant about the alternates, but each one offers just a bit more precision, gives the reader just a bit more story detail to work with, and links the action across sentences more naturally. Of course, the alternates above, once mastered, will themselves be stepping-stones to more elaborate realizations (e.g., for #4: Leaving him momentarily dumbfounded and confused...). In addition to the action-orientation that inheres in its meaning, the verb is a potent force because of the role players and specific relationships among role players that the verb controls in a proposition. This property opens up new perspectives from

which to frame significant details, events, and characters in a piece of writing, and so the sentence initial clause (among other things) is an indispensable part of the writer's tool kit.

Table 12. April (#19) and December (#44) story sequences

	April Story	December Story	Remarks
1	*One day I went to the restlunt with my girl friend.*	*One day a man had dinner with his girlfriend.*	Apr restaurant is exchanged for Dec dinner as the principal locator. On both occasions, a girlfriend.
2	*And I ask violinist to play the violin.* *So, he began to play.*	*Some minutes later, he asked a waiter to play music. The waiter play the violen for the couple.*	Each event is framed in a simple sentence although the Dec sentences have been lengthened by a DM "Some minutes later" and PP "for the couple."
3	*I hug her sholder.*	*Then, the atmosphere became better, and the atmosphere of the couple became better too.*	Apr atmosphere is registered, perhaps, in the boyfriend's hugging her sholder; Dec is alternatively described in terms of environmental atmosphere getting better and the couple's circle of warmth improving, both a result of music.
4	*We left the restlunt.* *I talked with her in the park.*	*After the dinner, he brought her to a park near his house. He told her "I love you, I need you" again and again. She listend and was silent.*	DM "After the dinner." the Apr "I walked with her in the park" has been amplified greatly with direct quoted speech, "'I love you, I need you' again and again" Far more effective a device for engaging the reader. Also, Dec records her reaction to the boyfriend's entreaties.
5	*At the last, She came my house.*	*He told her "Why don't you go my house" and he took her to his house.*	Dec further develops the boyfriend's character through dialogue. Also, Dec realized this event in to actions, whereas Apr is abbreviated.
6	*After few minutes, She left my house*	*Two hours later, soon, her celephon was called. She talk to someone without her boyfriend. She told him "I must go out soon, but I'll be back" Don't worry" He permited her.*	Apr leaves the reader with many questions, but Dec provides background into girlfriend's actions through direct reported speech: again, far greater detail.
7	*An our leter, She came back with a man The man is the violinist, who I met in the restlunt.*	*A few minutes later, she went back his house with another man whom her boyfriends knew. The man is the waiter.*	The only complex clauses in the both Apr and Dec narratives-both of them employing RCs as the device through which to identify the mystery "man" (and presumed lover of his girlfriend) as the violinist/waiter. Dec is strategically the better of the two RCs and more effectively sets up the mystery identification. Interesting that for both Apr and Dec, both begin the sequence with DMs.
8	*I was surprised and shocked.*	*Her boyfriend said his self again and again "Why? Why? Why?..."*	Apr records the fact of surprise and shock, but in a wonderful dramatic twist, Dec dramatizes the boyfriend's shock and surprise in this plaintive cries, "Why? Why? Why? Why?"

CASE STUDY 'C' (NARRATIVES #26 & #47)

April story — #26 (Global=3.25)

One day a man and a woman had dinner at restaurant.

Then a violinist came and play the violin for man and woman.

with the help of violin's music, a man and woman made nice atmosphere.

After out of restaurant, a man propose to a woman in the park.

Go back to the restaurant, a woman run toward something.

In a little while a woman go back, but

December story ― #47 (Global=5.00)

One day Tom and Mary eating dinner at the France restaurant. Then the violinist came to them and played violin for them. The violin tone color is very beautiful so they felt comfortably. After eating dinner, they go to the part. Then Tom proposed to Mary.

After that time they go to Mary's house. Mary kept Tom waiting at the door. After some minute Mary backed to the door with another man. Tom was relly surprised because the man is violinist when played violin for them at restaurant. And Mary said "This is My father." Tom was fall down on the spot.

Table 13. Results for measures on Case Study 'C'

	Glob	Dis	Mech	S	W	Cl/S	W/S	PerEr	ErFr	TOEIC
Apr	3.25	3.25	2.25	5	58	1.40	11.6	0	0	150
Dec	5.00	4.75	4.50	11	102	1.64	9.3	45.5	5	154

Commentary on 'C'

The December values for most of the measures are appreciably better ― even those for grammatical accuracy. Only those for grammatical complexity results are mixed where we see a modest gain in Cl/S, but a slight decline in W/S. Once again, we can appreciate how much better developed a narrative can be with increased productivity: six additional sentences amounting to 44 more words for the December. Overall, the December story is superior to the April one and the Global score rose by 1.75.

One interesting feature about this pair of texts concerns the writer's struggle with the English reference system and basic anaphoric devices associated with it. Notice in the April text, for example, that many nouns that should have been preceded by the definite article *the* are consistently associated with the indefinite *a* instead. The struggle is most obvious with the writer's overuse of the indefinite article with the nouns *man* and *woman*, as in '*a* man and woman made nice atmosphere' and '*a* man propose to *a* woman,' etc. In these and other instances, the writer should have used either the definite article ('*the* man and woman made nice atmosphere' '*the* man propose to *the* woman') or resorted to using subject and accusative pronouns ('*they* made nice atmosphere' 'he propose to *her*'). There are four nouns in the text that do not follow this generalization: 'with *the* help' where the definite article is an invariant

part of the set phrase; '*the* violin' for which either the definite or indefinite article would have sufficed; '*the* park' which should have been preceded by the indefinite article; and '*the* restaurant' which properly refers back to unmarked 'restaurant' in the first sentence.

　　　In December, the writer avoids many of these potential trouble spots by giving each character a name, Tom and Mary (which also has the nice effect of giving the characters a bit of personality and lays the groundwork for the story to come). But somewhere within the eight-month interval, this writer has also learned to use third-person pronouns as the following examples show: 'Then the violinist came to *them* and played violin for *them*' or the third sentence 'The violin tone color is very beautiful so *they* felt comfortably,' etc.　This may represent a step forward along the developmental continuum that, of course, does not show up in our data.　On the other hand, the December narrative does not present even one correct example of the classic first-mention/subsequent-mention noun phrase sequence preceded by *a* and *the* respectively.　There are, in the December text, several noun phrase contexts for which the writer is obliged to provide an indefinite/definite article sequence.　Follow, for example, the two instances for the noun *restaurant*; or, more interesting, follow the fate of the *violinist/violin* noun complex throughout the text.　In both cases, the writer fails to use or associate the appropriate article with the nouns.　So while the writer in the December narrative demonstrates control over the use of third-person pronouns, he or she has not yet worked out the notoriously thorny English article system.

Table 14.　April (#26) and December (#47) story sequences

	April Story	**December Story**	**Remarks**
1	*One day a man and a woman had dinner at restaurant.*	*One day Tom and Mary eating dinner at the France restaurant.*	Apr setting is accomplished, but the slightly expanded setting in Dec with character names is an improvement.
2	*Then a violinist came and play the violin for man and woman.*	*Then the violinist came to them and played violin for them.*	Both sequences are remarkably alike. Apr 'man and woman' are not pronomialized as in Dec 'them.'
3	*with the help of violin's music, a man and woman made nice atmosphere.*	*The violin tone color is very beautiful so they felt comfortably.*	Both sequences alike in intent, but Apr is enhanced with PP 'with the help of.' Apr characters are still associated with indefinite article.
4	*After out of restaurant, a man propose to a woman in the park.*	*After eating dinner, they go to the part. Then Tom proposed to Mary.*	Dec is sequenced more logically: dinner->park ->proposal, but realized in two sentences compared with one sentence for Dec.
5	*Go back to the restaurant, a woman run toward something. In a little while a woman go back, but*		In the Apr sequence, the woman returns to the restaurant, but for what is unclear. In the last fragment, it is not clear where the woman returns to. Perhaps the writer was not sure what to make of the story. The female continues to be identified as 'a woman.'
6		*After that time they go to Mary's house. Mary kept Tom waiting at the door.*	This sequence shows that the writer is now following the story-line in the cartoon, in contrast to Apr where is not clear.

7		*After some minute Mary backed to the door with another man.*	Here the writer competently sets up the arrival of man the in anticipation of the surprise ending.
8		*Tom was relly surprised because the man is violinist when played violin for them at restaurant.*	A wonderful multiclause sentence (the longest in either Apr or Dec narrative) which makes the connection between the violinist and the man using a botched RC. Writer skillfully sets up the next sequence in which relationship of the man to the woman will be revealed.
9		*And Mary said "This is My father."*	The only quoted speech reveals the man's relationship to the woman. 'And…' leads us to the revelation.
10		*Tom was fall down on the spot.*	This last sequence gives the man's reaction. A near-masterful rendering of the English expression of ritual shock, "He fell down on the spot"

CASE STUDY 'D' (NARRATIVES #46 & #18)

April story ― #46 (Global=5.50)

One day a beautiful woman approached me when I was having a dinner. She's name is Bell and I'm sure she like me. We talked a lot. She working at disco. When we having a good time, a man comeing to us who was playing violin. We listened it. We became very close. It was very romantic! After the dinner, we leave the restaurant together. We went to park near restraunt. I love her very much. So I dicided to tell her what I feel. She was sitting in bench and I said that I love you looked at her eyes. Then she said I want you to meet a man. If you meet him, I'll follow you.

Of course I will! She took me a small house. She opened at the door and said please waiting here. Then she was gone from me. "How long had I waiting"?? 30...? 40...? oops. I've only waiting 10 munis!? Wow. I really love her.

I got in sight her. I was happy but not only her. She came to me with big man with a glasses. He was her father. That's my story about my wife. We still love each ather now.

December story ― #18 (Global=4.75)

One day one cupple were having dinner. A man called Jack. He is a business man and he wore a suit. A woman called Stephany. She has a blond hair and she wore a red dress. She is beautiful. She really love him. She wanted him to love her. Another man who was playing the guitar approached them. His music makes them feel good. It was romantic. Stephany looks so beautiful for Jack. He fell in love with her at last. After an hour, they left the restaurant. He proposed her at the park. She was so happy to hear that. She invited him to her house. Then he was surprised because her father was a guitarist who was playing the guitar at the restaurant. This story was planned by Stephany. Her strategy succeed.

Table 15. Results for measures on Case Study 'D

	Glob	**Dis**	**Mech**	**S**	**W**	**Cl/S**	**W/S**	**PerEr**	**ErFr**	**TOEIC**
Apr	5.50	5.00	3.50	28	196	1.54	7.0	50.0	14	156
Dec	4.75	2.50	4.25	20	134	1.40	6.7	70.0	14	166

Commentary on 'D'

This pair of narratives represents one of the four cases where the December Global score declined-in this case, by -0.75. Note that all values for all measures (except PerEr) also fell. In particular, the Discourse and the two productivity measures, S and W, evidenced a considerable decline: -2.50, -8.0, and -62.0 respectively. All things considered, the April story is the more satisfying of the two. It is curious that this writer, who in April would begin the narrative by setting up the story with background details ('a beautiful woman', 'Bell,' 'she like me') competently embedded within two multiclause sentences, would, in December, begin her narrative with the two disjointed sentences 'One day one cupple were having dinner. A man called Jack.' The latter is the vehicle the writer uses to set up a description of the male character in the next sentence 'He is a business man and he wore a suit.' This pair of sentences finds a parallel in the next two sentences, 'A woman called Stephany. She has a blond hair and she wore a red dress.' The December opening is artless, and makes us wonder if this writer is also is the author of the April text. The reason we are convinced of their common authorship is because, in both narratives, the writer freely and exuberantly invents on the characters' manners, clothing, and circumstances. The interesting question is what assemblage of factors account for the lackluster December performance? Compared to April, the December writing is perfunctory: a series of choppy single clause sentences that do lead to up the competently rendered relative clause identifying the man at the door with the guitarist from the restaurant, 'Then he was surprised because her father was a guitarist who was playing the guitar at the restaurant.' (Note, the April violin had become a guitar in December.)

This pair of narratives is presented, without analysis, just to provide the reader a contrast to the previous three pairs of narratives which showed overall positive development from April to December.

SUMMARY OF FINDINGS AND CONCLUSION

Dependent measures

Students made significant progress in their written communication skills, although not in all the areas examined for this assessment. Modest gains were achieved in terms of textual coherence, organization of information, use of discourse markers, and clarity of expression. The practical

outcome of these gains is that students are capable of more effectively communicating intended meanings in their writing. Simply put, students are better writers at the end of the academic year than they were at the beginning. Students made significant improvements in mechanics and formatting which includes following conventions of English paragraphing, spelling, and punctuation. Also strong were the results for productivity which found students writing overall more words and sentences. This finding suggests that students have a greater facility with English and a higher level of confidence in their powers of self-expression. Only weak gains were achieved in grammatical complexity: a small, but significant, increase was found in the numbers of clauses per sentence. The two measures for grammatical accuracy, however, showed no meaningful gains. Students did not evidence any progress in resolving sentence-level grammar problems.

In sum, the results for the dependent measure are mostly positive and encouraging, and confirm that students have indeed moved a step forward as writers in English.[6] At the same time, I do believe that these modest gains, encouraging as they are, in truth represent a call to action, given the length of time and numbers of classroom hours the students had to improve their writing skills from April to December.

Gains across writing skill levels

One very important finding from this assessment is that the gains were not achieved uniformly across the three skill levels (as defined in terms of initial Global score). This analysis revealed that it was the least skilled writers who made the most, and significant, progress on all measures, whereas the gains for the most skilled group were often insignificant or even, in the case of grammatical accuracy, negative. Although other explanations are possible, I believe these findings suggest that while the lower level writers are indeed benefiting from the instruction, higher level students needs' are not being addressed. Students who enter the university with comparatively high level writing skills may need more challenging materials and more attention given over to higher order writing considerations.

TOEIC Bridge scores

Only 25 of the 32 students in this study took both the April 2006 and January 2007 TOEIC Bridge test. Overall, students' scores increased by a small increment across test administrations. The TOIEC Bridge scores from both tests correlated modestly with the chief dependent measure of this assessment (Global), but weakly or not at all with the other eight measures. From these combined findings, a reasonable conclusion is that the TOEIC Bridge scores are a poor indicator of success on the measures adopted by this assessment. Thus, although TOEIC Bridge scores

gains may be disappointingly low, as explained above, students nonetheless made significant progress as writers.

Text analysis

Analysis of the eight student narratives provided a behind the scenes perspective through which to appreciate the significance of each dependent measure, illustrated the range of strategies adopted by students in their writing, and examined several challenges and positive developments in the students writing from the April to the December interval. Although not comprehensive, these analyses have suggested many fruitful directions that future text analyses for this assessment might take. Also, these analyses provide substantive evidence confirming the statistical findings: significant learning has taken place, and students have moved forward as writers in English.

IMPLICATIONS FOR THE LANGUAGE CURRICULUM

Writing

Continue to develop students' fluency in writing through 'free writing' or 'quick writing' exercises. Provide students with more reading opportunities (e.g., a wide variety of models illustrating types of writing) to expose them to as many different styles, approaches, strategies as possible. Try to budget classroom time to be able to include consciousness raising exercises in grammar. Develop activities which focus on sentence combining within the context of a larger piece of writing to develop their formal range for expression. Also, as explained above, students who enter the university with comparatively high level writing skills may need more challenging materials and attention given over to higher order writing skills. For the high skilled group, recommended activities are ones where students are obliged to evaluate whole texts, that direct their attention to gross structural relationships and the organization of large chunks of information within texts, forcing students to see the forest for the trees. Also, move the students quickly through material whenever possible.

Grammar

Students would benefit from grammar instruction that includes explanations in Japanese, but the emphasis should be on exercises that provide opportunities for students to use the new grammar in spontaneous interactions (i.e., communicatively, spoken or written). The use of mechanical grammar drills should minimized. Where possible throughout the English-medium skills courses, emphasis should be placed on providing grammar instruction through consciousness raising exercises. As an example of a technique that narrowly focuses on sentence-level

grammatical problem areas, teachers can copy individual sentences from students' compositions that include common mistakes (e.g., in subject-verb agreement, word order, article use, use of prepositions) and have students collaborate in making changes to them. This is a very effective and highly motivating way of getting students engaged in improving their grammar: the solutions that students come up with are immediately applicable to the writing task at hand.

Reading

Above all, it is likels that the many obstacles students face in their English language development (e.g., sentence-level grammatical deficits, insensitivity to structural and organizational features of English text, narrow range of vocabulary) would be effectively addressed if students were, throughout their university years, required to do large amounts of extensive reading. Research has showed that benefits from extensive reading programs are manifold; our students' English language development would be greatly accelerated through dedicated participation in such a program. Extensive reading should immediately become the cornerstone of the University reading curriculum.

NOTES

1. I am most grateful to the teachers of Bunkyo Gakuin University whose cooperation in administering the writing task during regular class times made this assessment possible.

2. Teachers gave the students oral instructions not to worry about spelling and grammar too much, but to focus instead on relating the story itself: meaning over form. This was also done to encourage the students to take risks and not waste time recalling details of spelling and specific points of grammar. I cannot speculate on the actual effect of this instruction on the students' writing. But I do feel that any newly acquired grammar internalized over the eight-month interval would have, all things being equal, found its way into the students' December narratives even if the students were focusing on meaning. Note, too, that the task sheets given to students had printed on the first line at the top "One day..." to prompt students to frame their narrative in the past tense, to start their narrative off in typical story-telling style, and to give the students a common point of departure. This is the reason all the narratives begin with "One day..."

3. For the April administration, 240 students were present; for the December, 170 students. Thirty-six narratives were first selected at random from the April classes and then paired up with 36 from the December classes. In the event that a student attended class in April but not in December (meaning that a paired match couldn't be made), a different April narrative was selected at random from the same class and matched with the same author of a December narrative. The sample size of 15.6% was calculated on the basis of the 36 narratives selected from the original pool of 230 April narratives.

4. The two raters were Eleanor Kelly and myself. Many thanks to Eleanor Kelly for the long hours she put into evaluating these narratives! Thanks also to Bruce Brinkman for additional assistance in the selection and handling of the narratives.

5. I note that the two grammar measures (PerEr, ErFr) are very coarse and that measures targeting

specific grammatical systems such as reference, articles, subject-verb agreement, verb tense and aspect, might have, because, of their greater sensitivity, picked up developments (positive or negative) in students' grammar from April to December.

6. There were many things competing for students' time during the period this assessment took place: apart from all the social adjustments that attend the first year of university life, first-year students typically carry a 14-18 course load per semester, only four of which are English skills courses; and most first year students have part-time jobs in evenings and on weekends. Although I haven't made these considerations a part of our analysis, these factors should also be given some weight when evaluating gains or deficits in the students' English language development.

REFERENCES

Carlisle, Robert (1989) The writing of anglo and hispanic elementary school students in bilingual, submersion, and regular programs. *Studies in Second Language Acquisition*: 11, 257-280

Crookes, Graham (1989) Planning and interlanguage variability. *Studies in Second Language Acquisition*: 11, 367-388

Hatch, Evelyn and Anne Lazaraton (1991) *The Research Manual: Design and Statistics for Applied Linguistic*s. Newbury House: New York

Hughes, Arthur (2003) *Testing for Language Learners 2nd Edition* Cambridge University Press

Ortega, Lourdes (1999) Planning and focus on form in L2 oral performance. *Studies in Second Language Acquisition*: 21, 109-148

Weigle, Sara (2002) *Assessing Writing*. Cambridge University Press.

Wendel, John N. (1997) *Planning and Second-Language Narrative Production*. Unpublished Doctoral Dissertation. Temple University. Philadelphia, Pennsylvania

Wendel, John N. (1999) Planning and production: A focus on syntactic complexity. *Pragmatics and Language Learning*: 9, 195-211

Wolfe-Quintero, Kate, Shunji Inagaki & Hae-Young Kim (1998) *Second Language Development in Writing: Measures of Fluency, Accuracy and Complexity*. Second Language Teaching and Curriculum Center: University of Hawai'i Press.

APPENDICES

A. Limitations and weaknesses in this assessment

1. Frequency of assessment: Only two 20-minute administrations: April & December.

2. Task type: There was only one task type in this assessment. The task type was limited to narrative.

3. Few measures: Targeted only a few measures of a whole universe of possible measures. No measures for vocabulary were included; no measures for propositional phrases; no measures targeting specific grammar points (e.g., subject-verb agreement, article use, verb morphology); no measures for accuracy in discourse level grammar.

4. Sample size: A modest sample size of 15%.

5. No instruction: The narrative genre was not a specific course or curriculum objective. Student were not given specific instruction during the year in narrative or story-writing.

6. Practice effect: The practice effect must be taken into account. Although a full eight months had passed (and, an important detail, students were not permitted to keep the cartoon after the April

writing task), it is likely that some of them may have been able to embellish the story with more details, more engaging characterizations, during the eight month period.

7. Task instructions: For this assessment, students were told to focus, in their relating of the story, on communicating the story itself and not to worry about spelling and grammar. Although I have no reason to beliefe this was the case, this particular instruction may have backfired and some students may have felt that 'anything goes.' The result would have been a less than representative sample of their writing abilities.

8. Out-of-the-classroom influences: Lots of outside experiences impact on the student English-language development: the Chat Lounge (a large lounge area on campus where students may visit with native or near-native speakers during class hours: some student go often; others, never or rarely), conversation schools outside university, independent study and reading, experiences abroad (home stay, study), etc.

9. Test-retest: Giving the same cartoon to write about twice may have been demotivating for some students. This may be especially true for the more skilled writers who may have felt that they'd already written in detail about the story and saw no reason to do so again.

B. The eight-frame cartoon

For copyright reasons, the original cartoon has not been included. Instead, a short, frame-by-frame description is given as below.

〔Background〕

This is essentially a love story with a quirky ending which leads to a number of possible interpretations. By design, there is no running narrative or dialogue provided in the individual frames. The cartoon has no title. Students only have the sequence of pictures to go by and must imagine what the characters might be thinking or saying. The cartoon is a black-and-white print of the sort you might find in a 1950s issue of the New Yorker magazine.

〔Frame 1〕

An elegantly dressed lady and gentleman are seated at a table-for-two in an high-end, cozy restaurant. They are enjoying their dinner with wine and smiling together.

〔Frame 2〕

An older man, a violinist, has approached their table, violin at the ready. He has a mustache and is wearing glasses. The couple regard him agreeably.

〔Frame 3〕

While the violinist begins to play, the couple join hands across the table, their eyes closed in quiet enjoyment, their smiles in full bloom.

〔Frame 4〕

Now the couple have arranged themselves at one end of the table. The gentleman has his arm around her shoulder and the lady's head rests on his. The violinist plays passionately for the couple's enjoyment.

〔Frame 5〕

The couple, now in their winter coats, are passing the cashier's desk and leaving the restaurant. They have their arms around each other and appear deeply happy.

〔Frame 6〕

The couple have entered a park and the lady is now seated on a park bench. In front of the lady, the

gentleman is on bended knee either professing his love to her or proposing marriage to her. As he addresses her, his hands and arms gesturing meaningfully, her gaze is directed unassumingly to a point in the low distance.

〔Frame 7〕

The couple have just entered a home, apparently that of the lady. She leads him into the foyer. The gentleman has just crossed the threshold, his hand still on the door knob. They both have pleased and expectant expressions on their faces.

〔Frame 8〕

Still in the foyer, the lady reappears from within the home with another man at her side, his arm around her. The man is clearly the mustachioed, bespectacled violinist from the restaurant. He is extending his hand to the gentleman for a handshake. The gentleman's expression is one of profound bewilderment and confusion, though his hand has been extended, in a polite, but perhaps automatic, gesture to meet the violinist's.

C. Criterion measures rubrics

1. Global

Global Scale Rubric

1 — little or no response, too little to be evaluated.

2 — incoherent, can't follow the story, totally incomplete story line.

3 — story difficult to follow but can make out inklings of a structure, incomplete storyline, only basic vocabulary and often inappropriate, sentences awkwardly combined.

4 — basic story structure is clear (little difficulty in making out overall story, occasional problems), characters can be identified across sentences with usually appropriate use of pronouns, limited range of vocabulary, sentences adequately, but occasional problems

5 — story structure is clear, characters are embellished somewhat, vocabulary choices are often effective, sentences fairly well combined although may not be natural.

6 — story written with naturalness, characters are developed, writing is engaging. very effective choice of words, sentences combined in natural and logical way.

2. Discourse

Discourse Scale Rubric

1 — little or no response, cannot be evaluated.

2 — no use of discourse markers at all. Writer relies solely on order of sentences to more event structure, action, story along.

3 — some use of discourse markers, but may be inappropriate or awkward reducing their discourse value.

4 — use of discourse markers essentially appropriate, perhaps basic, though may be awkward.

5 — good use of discourse markers, natural, though may include repetitions of same types demonstrating limited range.

6 — elaborate use of discourse markers, shows inventiveness, effectively deployed throughout.

3. Mechanics

Mechanics Scale Rubric

1 — little or no response, too little to be evaluated.

2 — formatting and/or mechanics severely interfering with reading experience (e.g., listing).

3 — basically following conventions, but occasional slips in ways that cause the reader to work too hard in following the story.

4 — essentially following all conventions, infrequently slips do not cause a problem in the reading experience.

5 — conventions observed throughout, reading experience is fluent.