

The Relative Influence of Language Use Errors and Coherence Breaks in EFL Student Essay Evaluation: Content Versus Clarity¹

Hiroe Kobayashi

With the recent emphasis placed on process writing that focuses on content, ESL and EFL teachers have still been observed to show a great deal of concern with errors in L2 student essays (Farghal, 1990; Fathman and Whalley, 1990; Leki, 1990; Zamel, 1985). In responding to student essays, many teachers correct language-specific errors the students make, giving little attention to the content of the writing (Zamel, 1985), and in evaluating the essays, they often base their judgments on the basis of "the number of errors they [students] had rather than the meaning or logic they embraced" (Farghal, 1992, p.42). Thus, whereas sentence-level errors are an inevitable part of L2 student writing, the teachers' high concern with error has led many L2 student writers to view freedom from error as being the principal characteristic of good writing (Devine, Railey and Boshoff, 1993).

To help improve L2 student writing, a number of studies have investigated the effect of teacher feedback on sentence-level errors in terms of how effective the feedback is (Fathman & Whalley, 1990; Hendrickson, 1978; Lalande, 1982; Semke, 1984) and also how/when it should be given (Cardelle & Corno, 1981; Robb, Ross and Shortreed, 1986). However, these studies provide no conclusive evidence for the relation between teacher feedback and the improvement of L2 student writing ability, although Fathman and Whalley (1990) showed that feedback on grammatical errors exerted a greater effect on grammatical accuracy than feedback on content does on the improvement of content.

Whereas L2 student essays are often assessed in terms of a variety of criteria such as content, organization and language use (see ESL Composition Profile, for example, Jacobs et al, 1981), the presence of errors can significantly affect the evaluation of overall quality (Farghal 1992; Kobayashi and Rinnert, 1993; Sweedler-Brown, 1993). Sweedler-Brown (1993), for example, found that sentence-level error was responsible for the significantly lower holistic scores of L2 student essays for the original than for the corrected versions. Similarly, Kobayashi and Rinnert (1993) found that overall holistic scores improved considerably for the same essays when all errors were eradicated.

Possible effect of error on the evaluation of L2 writing is not limited to overall quality. McDaniel, according to Sweedler-Brown, found that raters untrained in evaluating ESL writing were affected by error when they evaluated ESL essays, having difficulty in distinguishing between the quality of various features in the essays and scored all features of the essays similarly (1993, p.5). Sweedler-Brown, on the other hand, found no such observable effect except a small increase in scores for organization and content development for the corrected essays, but she cautioned that when a rater is presented a mixed batch of NS (native speaker) and ESL essays, the salience of the ESL errors, which are absent in NS essays, apparently will obscure the rater's perception of competence in organization and development in the ESL essays (Sweedler-Brown, 1993, pp. 10- 11). In relation to L2 writing evaluation, these studies raise the question of whether L2 errors affect a rater's judgment of other features of L2 writing, such as content and organization.

Present Study

The present study aims to investigate how two types of textual features, language use errors (syntactic and lexical errors) and coherence breaks (disrupted sequences of ideas within paragraphs), affect readers' assessment of EFL student essays, with regard to two specific writing qualities: *quality of content* and *clarity of meaning*. More specifically, following Kobayashi and Rinnert's study (1993) that found a significant relation be-

tween language use errors and overall quality, this study further explores the relative influence of such surface errors on these two more specific qualities of L2 writing. In addition to sentence-level features, the paragraph-level feature, coherence break, which has not been as widely studied in terms of its effect on writing evaluation, was also examined in the present study, because coherence has increasingly gained attention in the areas of L2 writing pedagogy and research (for example, see Connor & Johns, 1990).

To investigate possible effects of language use errors and coherence breaks, the two measures of *quality of content* and *clarity of meaning* were selected. *Content*, constituting one of the principal criteria for analytical evaluation of L2 student essays (Jacobs et al, 1981), is assigned scores independently from other criteria, namely, organization and language use. Thus, it is important to examine if there is any effect of such sentence-level and paragraph-level feature on readers' assessment of content. *Clarity*, on the other hand, is not used as a major, distinct component for writing assessment, but presumably relates to language use, organization and coherence. In fact, this relation was supported by the finding that clarity was highly correlated with the two measures of language use and logical connection (Kobayashi and Rinnert, 1993). To investigate this relation in more depth, clarity was also included in the present study. Lastly, like the original study, the present study included four groups of readers with different backgrounds (two English teacher groups - native English and native Japanese - and two Japanese university student groups with differing degrees of writing experiences) to examine whether there are any significant differences among these groups.

The present study asks the following three research questions:

- (1) Do language use errors and coherence breaks within paragraphs affect assessment of EFL student writing by the four groups in terms of content or clarity? If so, in what ways?
- (2) Which criterion, content or clarity, shows a larger mean difference when essays with language use errors are compared with error free essays?
- (3) Which criterion, content or clarity, shows a larger mean difference

when essays with coherence breaks within paragraphs are compared with those having no such problems?

Data Base

The present study is based on data collected for a previous study (Kobayashi and Rinnert, 1993). The data consist of evaluation responses made by 465 readers making up the four groups: 127 inexperienced (IS) and 128 experienced (ES) Japanese university students, and 104 native Japanese (JT) and 106 native English (ET) teachers. The inexperienced students, mostly sophomores, had received no specific English writing instruction at the university, whereas the experienced students, drawn from five different schools, did have at least one semester of such instruction.² Most of the teachers (90%) in both the Japanese and native English groups were teaching at college and universities in Japan; a majority of the native English speakers (80%) were from the U.S. and Canada, the rest from other English speaking countries.³

In order to elicit readers' responses, sixteen different versions of two original student essays, which were written on two different topics (Topic 1: the disadvantages of owning a car; Topic 2: the comparison of cars and bicycles), were constructed in terms of three discourse features: culturally influenced rhetorical pattern, language use errors and coherence breaks. Whereas all the versions reflected the characteristics of either the American or Japanese rhetorical pattern (see Kobayashi and Rinnert, 1993, for detailed characteristics of each pattern), they also reflected one of four textual features: error free (EF), syntactic/ lexical errors (S), disrupted sequences of ideas within paragraphs (D) and both syntactic/ lexical errors and disrupted sequences of ideas combined (DS). The syntactic/ lexical errors included 25 typical errors, drawn from those Japanese students actually made, consisting of a variety of problems (e.g., word order, connectives and verb form).⁴ Coherence breaks in this study corresponds to one of Wikborg's (1990) topic structuring problems: "misleading disposition (ordering of materials)" (p. 134). Disrupted sequences were created by shifting the order of two of the sentences in the body paragraphs of each

essay, without making any other changes; thus, it resulted in mixing references to present and past in the essay on the TV topic and scrambling advantages and disadvantages of each vehicle in the car vs. bicycle essay. It should be noted that while all the versions were constructed to reflect one of these textual features, the content of the two original essays was held the same according to the given topic. (In the Appendix, two of the sample essays are shown.)

By asking each reader to evaluate two essays (one on each topic, paired in terms of different discourse features) on a 10-point scale in terms of *quality of content* and *clarity of meaning* as well as the five other measures (e.g., overall quality and language use), a total of 930 evaluation responses were collected. These responses were first analyzed using a 4 x 2 x 4 (feature x topic x group) factorial design separately for each criterion (content and clarity), and also using a 4 x 2 (feature x criteria) design with repeated measures on one factor (criteria). The factor of rhetorical pattern was excluded from the present study mainly because this factor was not found to be significant in the earlier study.

Results

Content

Tables 1 and 2 show means and standard deviations (SDs) of content and clarity scores, respectively. Table 3 indicates the results of a three way ANOVA on content scores, and Table 4 shows those on clarity scores.

Regarding content, the results revealed significant main effects for three factors: feature [$F(3, 461) = 13.10, p < .000$, see Table 3], topic [$F(1, 461) = 41.45, p < .000$] and group [$F(3, 461) = 3.39, p < .018$], and also a significant interaction between feature and group [$F(9, 461) = 2.22, p < .019$]. Overall, error-free versions (EF) were rated highest, versions with disrupted sequences of ideas (D) were the next highest, and those with syntactic and lexical errors (S) the third, whereas versions with both syntactic and lexical errors and disrupted sequences of ideas combined (DS) were rated lowest (mean scores: 7.13 for EF; 6.73 for D; 6.43 for S and 6.201 for DS). This tendency holds true for the two different topics, although the TV

Table 1: Content Scores: Means and SD Group and Topic for Each Feature

Group	TV Topic				Car vs Bicycle Topic			
	EF	S	D	DS	EF	S	D	DS
IS	6.90 (1.64)	6.84 (1.97)	7.19 (1.77)	7.74 (1.65)	7.22 (2.00)	6.50 (1.85)	6.47 (1.85)	6.00 (2.05)
ES	7.72 (1.65)	7.38 (1.62)	7.50 (1.55)	6.88 (1.54)	6.88 (1.83)	6.66 (1.26)	7.06 (1.54)	5.97 (1.58)
JT	7.93 (1.22)	6.22 (2.14)	7.41 (1.34)	6.64 (1.68)	7.23 (1.86)	5.96 (1.99)	6.30 (1.82)	6.00 (1.66)
ET	7.31 (1.49)	6.85 (1.62)	7.17 (1.49)	6.04 (1.88)	7.08 (1.65)	5.74 (1.51)	5.70 (1.66)	5.04 (1.69)
	EF: error free D: disrupted sequences of ideas				S: syntactic and lexical errors DS: syntactic & disrupted			

Table 2: Clarity Scores: Means and SDs by Group and Topic for Each Feature

Group	TV Topic				Car vs Bicycle Topic			
	EF	S	D	DS	EF	S	D	DS
ES	(1.70)	(1.71)	(1.67)	(1.44)	(1.56)	(1.41)	(1.95)	(1.81)
JT	8.28 (1.49)	7.59 (1.82)	7.94 (1.48)	7.22 (1.56)	8.09 (1.73)	7.69 (1.20)	7.66 (1.81)	6.63 (2.24)
ET	8.39 (0.88)	5.78 (1.81)	7.52 (1.53)	5.96 (1.68)	8.27 (1.43)	6.46 (1.82)	7.04 (2.01)	6.56 (1.76)
	8.42 (0.99)	6.48 (1.77)	7.72 (1.68)	5.16 (2.14)	8.12 (1.58)	6.44 (1.67)	6.48 (2.03)	5.12 (2.14)
	EF: error free D: disrupted sequences of ideas				S: syntactic and lexical errors DS: syntactic & disrupted			

Table 3: Three-Way ANOVA of Dependent Measures (Content Scores)

Source	SS	df	MS	F	p
Feature	118.76	3	39.59	13.10	0.000*
Group	30.74	3	10.25	3.39	0.018*
Feature x Group	60.42	9	6.71	2.22	0.019*
Topic	125.24	1	125.24	41.45	0.000*
Feature x Topic	65.85	3	5.28	1.75	0.156
Group x Topic	3.59	3	1.20	0.40	0.756
Feature x Group x Topic	38.37	9	4.26	1.41	0.179

* $p < .01$ **Table 4: Three-Way ANOVA of Dependent Measures (Content Scores)**

Source	SS	df	MS	F	p
Feature	373.17	3	124.39	40.62	0.000*
Group	127.31	3	42.44	13.86	0.018*
Feature x Group	168.13	9	18.68	6.10	0.000*
Topic	0.08	1	0.08	0.027	0.870
Feature x Topic	47.03	3	15.68	5.12	0.002*
Group x Topic	16.88	3	5.63	1.84	0.139
Feature x Group x Topic	21.33	9	2.37	0.77	0.641

* $p < .01$

topic was evaluated significantly higher than the Car topic (mean scores: 7.11 for TV and 6.37 for Car).

The results of Scheffé tests, which were used to compare mean scores for each of three pairs of two features, indicate that error-free versions (EF) significantly differed from versions with syntactic and lexical errors (S) [$F(2, 468) = 17.898, p < .01$] and also those with syntactic and lexical errors and disrupted sequences of ideas combined [$F(2, 468) = 30.259, p < .01$], but it did not differ from versions with disrupted sequences of ideas (D).

The four groups differed in their assessment of content; overall, inexperienced students gave the highest ratings, and native English teachers gave the lowest ratings, whereas experienced students and Japanese teachers fell between the two (mean scores: 6.85 for IS; 6.72 for JT; 6.54 for ES and 6.37 for ET). In spite of this overall tendency, the significant interaction between feature and group revealed that there was a contrasting tendency between the teachers and students: whereas both Japanese and native English teachers gave higher ratings to error-free versions (EF) than the students, they gave severely lower ratings to versions with syntactic and lexical errors (S) than the latter groups (EF: 7.59 and 7.19 for JT and ET vs. 7.05 and 6.78 for IS and ES; S: 6.09 and 6.28 for JT and ET vs. 6.67 and 6.60 for IS and ES, respectively). Further, apart from these between-group differences, the results of additional Scheffé tests for within-individual group differences confirmed that Japanese and native English teachers made a significant distinction between error-free versions and each of the other two versions, S and DS, whereas no such distinctions were made by the students except one differentiation between EF and DS by experienced students. In short, these results suggest that syntactic and lexical errors and also this feature together with disrupted sequences of ideas within paragraphs affect the assessment of content, particularly by both Japanese and native English teachers, but not as much by the students.

Clarity

Regarding clarity, the results of a three way ANOVA indicate large

main effects for two factors: feature [$F(3, 461) = 40.62, p < .000$, see Table 4] and group [$F(3, 461) = 13.86, p < .000$] and a significant interaction between feature and group [$F(9, 461) = 6.10, p < .000$]. Topic was not found to be a significant factor for clarity, but this factor significantly interacted with feature [$F(3, 461) = 5.12, p < .002$].

The overall influence of feature on clarity resembles that on content: the four versions were rated in the descending order: EF > D > S > DS (mean scores: 8.06, 7.37, 6.89 and 6.42 for the respective features), and there were significant differences in the ratings between error free versions and each of the other three features [EF vs. S: $F(2, 464) = 46.11, p < .01$; EF vs. D: $F(2, 458) = 15.90, p < .01$; EF vs. DS: $F(2, 454) = 74.58, p < .01$], including one such difference between EF and D versions. Further, regarding between-group and within-group differences, the assessment of clarity showed a similar pattern to that of content. In spite of the overall tendency of the four groups' ratings of essays (IS > ES > JT > ET; mean scores: 7.74, 7.63, 7.03, 6.76 for the respective groups), the teacher groups gave higher ratings to error-free versions than the latter groups did, whereas they gave considerably lower ratings to versions with syntactic and lexical errors. As demonstrated for content, the teachers also significantly distinguished error-free versions from each of the three features on clarity, whereas no such distinctions were made by the students except one between EF and DS for experienced students.

Nevertheless, the significant interaction between feature and topic indicates that there was one major difference between content and clarity. That is, whereas the general tendency of mean score differences (EF > D > S > DS) holds true for the TV topic, it does not for the Car topic: syntactic and lexical error versions were evaluated slightly higher than versions with disrupted sequences of ideas. More important, relatively large differences, though not quite significant, were found between S and D versions on the two topics. Whereas the Car topic was rated higher than the TV topic for S versions, this tendency was reversed for D versions, with the TV topic being rated higher than the former (S: 6.57 and 7.21 for TV and Car; D: 7.68 and 7.04 for TV and Car, respectively).

In short, the above findings suggest that similar to the effect on content, syntactic and lexical errors and disrupted sequences of ideas with paragraphs affect the assessment of clarity by teachers and somewhat by experienced students; however, the effect of these factors appears to be influenced by the kinds of topic assigned.

Content vs. Clarity

The findings of the study so far indicate that the two discourse features affect the assessment of L2 student essays in terms of both content and clarity. This section focuses on comparing the magnitude of the effect of the two features on the evaluation of content and clarity to examine if there is a significant difference between the two. Thus, a two way ANOVA (feature x criteria) with repeated measures on one factor was applied for analysis, with only the results of interaction between the two factors being reported below.

(1) EF vs. S

Figure 1 displays mean clarity scores of error-free and syntactic and lexical error versions, 8.06 and 6.89, respectively, and also mean content scores, 7.13 and 6.43, for the two versions. The results of the statistical analysis indicate that there was a significant interaction between feature and criteria [$F(1, 460) = 11.91, p < .001$], which can be interpreted as indicating that mean score differences between these two features [1.17 for clarity and 0.70 for content] were large enough to be significant. This suggests that the effect of syntactic and lexical errors on clarity is larger than that on content.

(2) EF vs. D

Similar to the EF vs. S comparison, Figure 2 shows mean clarity scores on the two versions 8.06 for EF and 7.37 for D, and also mean content scores, 7.13 and 6.73 for the respective versions. The statistical analysis yielded the same result that feature significantly interacted with criteria [$F(1, 464) = 4.60, p < .032$], which indicates that the mean score difference between EF and D were significantly larger for clarity than for content [0.69

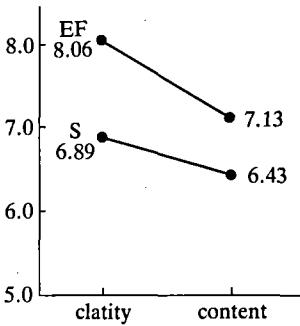


Figure 1: Feature x Criteria (EF vs. S)

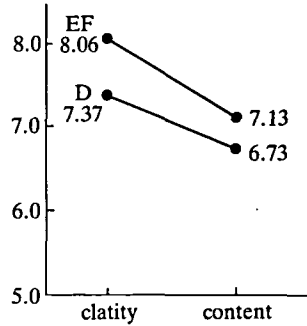


Figure 2: Feature x Criteria (EF vs. D)

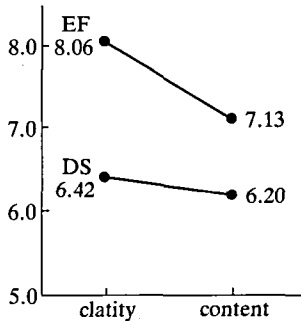


Figure 3: Feature x Criteria (EF vs. DS)

and 0.40, respectively]. This provides evidence that like syntactic and lexic-al errors, disrupted sequences of ideas within paragraphs exert more effect on the assessment of clarity than of content.

(3) EF vs. DS

This comparison showed the same results as the above. Since the two features (S and D) were combined, mean score differences between error-free versions and those with the combined features were even larger than those shown in the above comparisons [1.69 for clarity and 0.93 for content, also see Figure 3 for mean scores]. Thus, the significant interaction between feature and criteria [$F(1, 450) = 21.81, p < .000$] suggests that the combined feature of S and D, too, exert considerably more influence

on the assessment of clarity than of content.

In all, the major findings of the present study are summarized below:

- (1) Language use errors and coherence breaks affected the readers' assessment of student English essays in terms of content and clarity. More specifically, whereas these features (S and D) as well as the two features combined (DS) affected their assessment of clarity, only language use errors and the two features combined (DS) influenced that of content.
- (2) Topic affected content assessment, but this factor appeared to influence clarity assessment only when essays contained either language use errors or coherence problems.
- (3) Language use errors and coherence breaks affected the assessment of content and clarity by both native English and Japanese teachers, but had little effect on that by Japanese university students.
- (4) The effect of the two features appeared to be larger on the assessment of clarity than on that of content.

Discussion

The most significant finding of the present study is that when student essays contain syntactic and lexical errors or coherence breaks, readers are affected in their judgments of both clarity and content, with the judgments on the former criterion likely to be more influenced than that of the latter. The reason for the larger effect of the two text features on clarity than on content could be explained by the fact that clarity has a more direct relation with those features. Whereas content is related to what the writer has actually said about a given topic, clarity, which can be seen as an overall intelligibility of meaning, is connected to the communication of the content (or the writer's intended meaning) on both global and local levels. Language use errors, including such errors least acceptable to readers [Vann, Meyer & Lorenz, 1984 Santos, 1988] as wrong lexical choice and wrong word order as well as problematic use of connectives seriously interfere with the intelligibility of meaning on a local level. Thus, no matter how well the content is developed, such surface errors cause readers trouble in comprehending exactly what the writer wants to convey, and readers, parti-

cularly teachers, appear to be most troubled by this problem.

Like language use errors, disrupted sequences of ideas within paragraphs which make logical connections between ideas unclear also create difficulty in following the content. However, as opposed to surface errors impeding the comprehensibility of the writer's ideas on a local level including both intersentential and sentential errors, coherence breaks in this study influence the comprehensibility only on an intersentential level because disrupted sequences were constructed by switching the order of two of the error free sentences within paragraphs (essays with coherence problems were made error free in terms of syntactic and lexical choice). Because this error free characteristic left individual sentences comprehensible, readers were not troubled in grasping the meaning on a sentential level. This may explain why disrupted sequences of ideas did not hamper clarity as much as language use errors did, and also did not interfere as much with readers' understanding the content. However, it should be noted that the coherence problems that the present study dealt with are only one type of coherence breaks identified by Wikborg (1990); thus, the possible effects of other types such as uncertain inference ties and misleading paragraph division on the evaluation of clarity should also be examined in future studies.

Whereas content was less affected than clarity by the two text features, it was heavily influenced by the kinds of topic assigned. With regard to content, the TV topic was rated higher than the Car topic because for the former topic, the writer explained the social effects of a TV on family life, and ideas were presented in an analytical way. On the other hand, for the Car topic, the writer presented advantages and disadvantages of cars and bicycles through comparison and contrast, which could have given readers the impression that writing as such had clarity on a global level, yet lacked depth in the content. The analyses of readers' comments [Kobayashi and Rinnert, 1993], in fact, revealed that readers based their judgments regarding content on a variety of features, namely, writer's intention, originality, depth, balance, and example/support, and they gave more critical comments to the Car topic than the TV topic, including comments such as "a lack of depth" and "superficial." The readers' comments correspond to the

finding that lower ratings were given to the Car than the TV topic.

With regard to group differences, the two groups of English teachers were strongly affected in their judgments of clarity and content, whereas neither group of students was so affected. Because of their higher English language ability, the teachers easily perceived syntactic and lexical errors or coherence breaks in student essays so that they were apparently more troubled by these problems in comprehending the content. Thus, they were less lenient toward language use and coherence problems in making their judgments, though there may have been some difference between native and Japanese English teachers in their perception of such surface features.⁵ Further in relation to the teachers' high concern with errors, their positive vs. negative reinforcement may in part contribute to creating large differences in the teachers' ratings between error free versions and each of those containing errors (S and DS). In the role of judge, the teachers praise error free essays as "very good" perhaps to give further encouragement to student writers, whereas they may penalize essays with many errors by giving severely low scores. In contrast, the students did not make such distinctions consistently in their assessment on the two criteria, partly because their English ability was not adequate to perceive various problems in the essays, and partly because the group of experienced students, in particular, did not constitute a homogeneous group, due to a wide range of English ability and amount of writing experience among them.

Implications

The findings discussed above can be considered to have significant implications for writing assessment and writing research. In North America, several scoring scales, notably the ESL composition profile (Jacob et al, 1982) and Michigan writing assessment, are widely used by many college level ESL programs (Hamp-Lyon, 1991) to determine ESL students' writing competency for proper placement in writing programs. These scales are often based on analytical scoring, which assesses each of subcomponents of writing such as content, organization and language use independently (see Hamp-Lyon, 1991, for a complete scoring guide for the scales). However,

as the present study shows, this kind of analytical scoring could be negatively influenced by the presence of language use errors and coherence problems in student essays.

In relation to the assessment of ESL students' writing, Carlson (1991) addresses this problem, particularly how to deal with the most salient feature, errors, and emphasizes the need for training readers to be able to differentiate "the errors that signal serious writing problems and those that are easily corrected" (Carlson, 1991, p. 307). She further maintains that a specific writing program sets the criteria and standards in dealing with errors. In spite of this suggestion, however, Sweedler-Brown (1993) found that experienced English writing instructors untrained in ESL placed more emphasis on the ESL sentence-level errors than the essays' rhetorical features despite the faculty's reported criteria preferring rhetorical features over grammatical accuracy, whereas they did make a clear distinction between these two features in their analytical evaluation (see also Cummings, 1990 for this distinction).

The assessment of essays is essentially based on subjective judgments. Despite the amount of experience evaluating English essays, whether among ESL students' or native English college students', an individual reader brings his or her own perspective to this evaluation. As a result, some readers react to certain errors more severely than others, and such reaction may trigger judgments that are not in accordance with the criteria and standards of the assessment program (Carlson, 1991, p.306). Given the individual variation in the assessment of essays, which was also shown in our previous study (Kobayashi and Rinnert, 1993), the only solution for minimizing the individual differences among raters would be to raise the readers' awareness of this problem, discuss criteria thoroughly and run rigorous training sessions until agreement is achieved between the readers involved.

This kind of reader training is also necessary for researchers in both ESL and EFL situations when they use an analytical scoring scale to measure students' writing ability for their study, for example, to compare the writing ability of native and nonnative university students (Campbell, 1991)

or to determine the relationship between the frequency of revision and quality of writing (Kobayashi, 1991). Because the scoring provides critical data for research, researchers should be aware of the relative influence of language use errors and coherence problems on their evaluation and make an effort to eliminate artificial influences of such factors. Only careful evaluation of student' writing ability can lead to accurate research findings.

Lastly, the limitations of the present study should be mentioned. This study used the same versions of student essays constructed for the previous study (Kobayashi and Rinnert, 1993), where a number of variables were controlled in order to examine the effect of such variables on readers' evaluation. We knew it would have been better to use actual versions of student essays instead of manipulated ones; however, it was not possible because of the variety of variables included for systematic statistical analyses (see the discussion in Kobayashi and Rinnert, 1993). In spite of the fact that many variables (readers' backgrounds, writing experience and text features) were involved, however, the present study did not include detailed analyses for each variable; for example, in relation to readers' background, possible differences between the two student groups, as well as between the two teacher groups, can be further explored in depth. For writing assessment, it is suggested that alternative research designs should deal with a small number of variables for focused investigation, involving actual student essays.

Notes

1. I am grateful to Carol Rinnert of Hiroshima City University for her generosity in allowing me to use our shared data for the present study and also her constructive comments for the improvement of this paper.
2. A group of students drawn from one school, though freshman, received intensive writing instruction. Among other students, the amount of instruction varied from one to several semesters.
3. The rest of the native speakers of English were from the U.K., Ireland, Australia and New Zealand.
4. Twenty-five errors include the following: four word choice, two word

order, four adverbial connective, two word class, three verb form, two number agreement, three preposition choice, one pronoun, and four article errors.

5. The earlier study (Kobayashi & Rinnert, 1993) suggests that native English teachers' sensitivity to text features such as cohesive ties and choice of vivid words might have led to their higher ratings of the Japanese rhetorical pattern for the Car topic than its American counterpart, whereas Japanese teachers might not have noticed those features as much.

References

- Campbell, C. (1990). Writing with others' words: using background text in academic compositions. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 21 1-230). Cambridge: Cambridge University Press.
- Cardelle, M., and Corno, L. (1981). Effects on second language learning of variations in written feedback on homework assignments. *TESOL Quarterly*, 15, 251-261 .
- Carlson, S. B. (1991). Program evaluation procedures: Reporting the program publicly within the political context. In L. Hamp-Lyons (Ed.), *Assessing second language writing in academic contexts* (pp.293-320). Norwood, N.J.: Ablex.
- Cummings, A. (1990). Expertise in evaluating second language composition. *Language Testing*, 7, 31-51.
- Connor, U. & A.M. Johns (Eds.), *Coherence in writing: Research and pedagogical perspectives*. Alexandria, VA: TESOL.
- Devine, J., Railey, K., and Boshoff, P. (1993). The implication of cognitive models in L1 and L2 writing. *Journal of Second Language Writing*, 2 (3), 203-225.
- Fathman, A., and Whalley, E. (1990). Teacher response to student writing: focus on form versus content. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom*(pp. 178-190). Cambridge: Cambridge University Press.
- Fargahal, M. (1992). Naturalness and the notion of cohesion in EFL writing classes. *IRAL*, 30 (1), 45-50.

- Hamp-Lyon, L. (1991). Scoring procedures for ESL contexts. In L. Hamp-Lyons (Ed.), *Assessing second language writing in academic contexts* (pp. 241-276). Ablex, Norwood: N.J.
- Henrickson, J. (1978). Error correction in foreign language teaching: Recent theory, research, and practice. *Modern Language Journal*, 62, 387-398.
- Jacobs, H. L., Stephen. A. Z., Wormuth, D. R., Hartfield, V. F. and Hughey, J. B. (1981). *Testing ESL composition: A practical approach*. Rowley: Newbury House.
- Lalande, J. (1982). Reducing composition errors: An experiment. *Modern Language Journal*, 66, 140-149.
- Leki, I. (1990). Coaching from the margins. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp.57-68). Cambridge: Cambridge University Press.
- Kobayashi, H. and Rinnert, C. (1993). Factors affecting composition evaluation: Discourse features and readers' background. Paper presented at the 1993 IATEFL conference in Swansea, Wales and JALT 93 in Omiya, Japan.
- Kobayashi, H. (1991). Revisions by EFL college writers in Japan: Text quality and revision strategies, *Studies in Language and Culture*, 16,50-65. Hiroshima University Faculty of Integrated Arts and Sciences.
- Robb, T., Ross, S., and Shortreed, I. (1986). Salience of feedback on error and its effect on EFL writing quality. *TESOL Quarterly*, 20, 83-95.
- Santos, T. (1988). Professors' reactions to the academic writing of nonnative-speaking students. *TESOL Quarterly*, 22 (1), 69-90.
- Semke, H. D. (1982). The effects of the red pen. *Foreign Language Annals*, 17, 195-202.
- Sweedler-Brown, C.O. (1993). ESL essay evaluation: The influence of sentence-level and rhetorical features. *Journal of Second Language Writing*, 2 (1). 3-17.
- Vann, R.J., Meyer, D. E., & Lorenz, F. O. (1984). Error gravity: A study of faculty opinion of ESL errors. *TESOL Quarterly*, 18 (3), 427-440. 79-101.

- Wikborg, E. (1990). Types of coherence breaks in Swedish student writing: Misleading paragraph division. In U. Connor, & A. M. Johns (Eds.), *Coherence in writing: Research and pedagogical perspectives* (pp. 131-148). Alexandria, VA: TESOL.
- Zamel, V. (1985). Responding to student writing. *TESOL Quarterly*, 19, (2), 79-101.

Appendix

Essay 1 (American rhetorical pattern for the TV topic with syntactic/lexical errors)

In our society, almost family owns at least one TV set. TV function to send information throughout the society; in addition, we can enjoy a variety of programs for entertainment and education. However, TV set prevents from communicating with family members and this problem gets worse when families own multiple TV sets.

There has been a big change in the families communication way. In old days, family members enjoyed to talk over the table, even after dinner was over. Children are used to talk about what had happened in school or who they had played with after school. Parents, otherwise, usually listened such reports, while give comments and advice when necessary. While now most children try to finish dinner as fast as they can to rush to a TV set for their favorite program such as cartoon, or they try to watch it while eating their food very slowly. Then there is not much conversation going on between children and parents at the table.

This problem worsen when each household has more than one TV set. When children are unable to see their favorite cartoons on one TV set, they quickly turn to another one which usually placed in a different room in the house. Parents, to, would often do exact same when their favorite baseball game is not on one TV set. Since family members spend less time together before the TV, they have less ideas of exchanging for TV programs, too.

Although TV is fun to watch, it often creates a physical and mental

distance among family members. This can lead to great loss of mutual understanding among them. At last, many families are coming to be victims of technology, allowing themselves to be controlled by their own TV sets.

Essay 2 (Japanese pattern for the Car and bicycle topic with syntactic / lexical errors and disrupted sequences of ideas within paragraphs)

Today, we live in a fast moving society where people rush to work and to play. When there are public traffic facilities such as trains and subways to keep up with the fast pace in society, many people find it more convenient to have their own vehicles. In fact, cars and bicycles are both very popular.

Cars can transport a long distance quickly, so we can get many things done for either business or shopping. While driving cars is often inconvenient. When riding, a car can also protect us from stormy weather, so that we don't have to worry about getting wet or blown off the vehicles. In a traffic jam, cars can hardly move forward and it takes a lot of time to get to our destination. Also, it is time consuming to find a motorpool for a car in a busy area.

On the other hand, bicycles are not comfortable in bad weather. Taking a bicycle does not use gas, and it does not cost expensive insurance and maintenance. Otherwise, the biggest merit of bicycles is economy. The price of a bicycle is so reasonable that almost anyone can afford to buy one. Taking a bicycle provides the opportunity for enjoyable exercise. We get wet to the skin in a rainy day and feel cold in freezing winter. It is great to be exposed to the wind during riding under a blue sky. It makes us feel fit and healthy.

As we have seen, both vehicles are sides of good and bad each other. Cars are fast, but sometimes inconvenient, while bicycles are economical, yet influenced by weather. Then we cannot choose which one is better. It depends on the lifestyle we have and our own personal needs.