

Raters' background reflected in EFL writing evaluation

Soo-im Lee

英作文における採点者間の評価基準の違いとその要因

李 洙 任

Abstract

Who are ideal teachers in teaching EFL writing? It is arguable to decide who can be more qualified to teach writing; native speakers of English who have intuitive judgment in writing or Japanese teachers of English who have new-native speaker proficiency in English and have the experience of going through the complex process of acquiring English as a foreign language. The purpose of this study is investigate the differences in the criteria of writing assessment among the raters who have different backgrounds. Three native speaking raters and three Japanese raters participated in this study and their assessment criteria and types of feedback were compared to see what assumptions and expectations each group have toward students' writings. The native group relied on their intuitive judgment to measure the degree of naturalness and looked at writing tasks as a means of communication while the Japanese group was more concerned with sentence-level linguistic sophistication and used accuracy as the base line of their assessment. This study implies that the philosophies each rater has in terms of writing tasks reflect different expectation and assumptions.

Key words : Writing assessment, Language Measurement, Testing

(Received September 6, 1997)

抄 録

英作文を指導、採点するのに最も理想的な資質を得ている教師とはどのような教師をさすのであろうか。直観的に英作文の質や間違いを指摘できるネイティブ/スピーカーとネイティブレベルの言語能力をもつ日本人教師のどちらのグループが英作文を教える点において理想的な資質をもちえているのであろうか。本研究では英作文のクラスを担当している3人の日本人教師と3人のネイティブスピーカーの英作文における評価基準の違いに焦点をあて、その違いの原因となる採点者間の英作文における指導価値観を究明する。英作文の授業において、教師の指導法と評価方法の相関は高い。日本人教師は作文指導において文法や句読点の正しい文章を書くことを目的とした指導を行うことが多く、どうしても、テスト内容は穴埋め、整序問題、日本語を英語に訳すなどが中心となり、訳においての正確性を要求する。一方、ネイティブ・スピーカーにおいては、構成や論理性、統一性を重視する修辭的な文章を書くことを目的とする作文指導が多く、正確性の他に他の要因をも問う自由作文などがそのテスト内容となる。本研究では、2つのグループ間に評価基準の違いが見られ、その違いは各採点者の指導価値観から反映されたものと理解された。又グループ間においても海外留学経験有無などが要因となり評価基準が異なることが判明された。

キーワード : ライティング・アセスメント, ラングエッジ・メジャメント, テスティング
(1997年9月6日 受理)

Statement of the problem

The youth of Japan today are showing a keen interest in studying abroad, particularly America. Also with the expansion of Japanese business interests in overseas markets, companies have come to realize that international experience is an important and, in many cases, necessary asset. Japanese students are welcome in foreign countries, but in addition to the language problems, they frequently face great difficulties with the new study skills which are critically important to succeed in these foreign countries (Lee, 1990).

Most EFL teaching in Japan is responding to these urgent needs by offering more classes particularly designed for English for Academic Purposes (EAP). EAP curriculum is designed to teach classes such as reading, writing, TOEFL preparation, and study skills for the students' future academic task in study abroad, therefore the goals and objectives of these classes are more clearly specified than other EFL classes in Japan which aim at the improvement of general language proficiency.

EAP instructors in Japan are most concerned with how they can teach EAP curriculum to Japanese students effectively and efficiently. Choice of teachers for each class is one of their concerns, for example ample discussions are built on who should teach these EAP related classes at various schools. Many schools employ Japanese instructors for academic reading and TOEFL preparation courses because Japanese instructors are able to teach syntactic complexity using Japanese language more easily than native instructors. Courses for study skills, which are more or less combined skills of EAP, are often taught by native speaking teachers, and for example note-taking skill along with academic listening is one of the areas to be taught by native speakers in this area.

However, whether a native instructor or a Japanese instructor should be teaching academic writing is a difficult choice. If the purpose of a writing course is to teach expository writing emphasizing paragraph organization, coherence and cohesion and western rhetorical patterns, and to prepare our students to write academic writing as their final goal, it might be easy to conclude that native speakers would be better teachers than Japanese teachers. In terms of grammatical accuracy in writing, Hendickson (1984) proposed basic questions about correcting errors in L2 students' writing. He questioned when and how students' errors in compositions should be corrected and his final question was of who should correct them. Japanese teachers might teach better in terms of grammatical instruction based on translation exercises. However, Kobayashi (1992) claimed in his study that native speakers intuitively could correct more errors in L2 students' compositions than Japanese. He

also found apparent superiority of the native English speaking instructors in editing and correcting compared to Japanese teachers. The unstated question of whether L2 composition should conform to native speakers' expectations arises here. Before concluding the answer to the question, it is important to study what their expectations are and how their expectations differ from those of Japanese teachers.

A brief look at the writing courses in the U.S.A. and Japan

During the past 25 years of TESOL history, various approaches for writing have been introduced in the U.S.A., and the 1970s saw the development of more than sentence combining and controlled composition. New concerns replaced old. In place of "accuracy" and "patterns" came "process," "making meaning," "invention," and "multiple drafts" (Raimes, 1991). With this innovative ideology there was a clear concept that writing tasks should help students learn. The focus on the writer as language learner and creator of text has led to a "process approach" (White & Arndt, 1991; Zamel, 1982).

Although much of the research in communicative competence has focused on oral skills, the evolution of communicative competence was reflected in reading and subsequently in writing tasks also. Communicative writing was employed by focusing on the audience and the purpose of writing, and one of the textbooks following this approach, Leki's textbook, *Academic Writing: Techniques and Tasks* (1989), gives various topics in which students can express their thoughts and feelings.

With the idea that communicative competence focuses on language in use rather than on language acquisition (White, 1988), journal writing began to appear as a communication device that gave everyone in a class the chance to "talk" and receive feedback from a teacher. Advocates of this approach value the idea that journal writing is a creative task, and the purpose is not to display knowledge, but to discover knowledge. Students learn by doing, not by performing what they have learned elsewhere (Keio university, SFC, 1993).

While ESL teachers in the U.S.A. began to focus on "fluency" based on students' creativity along with "accuracy" in writing curriculum, EFL teachers in Japan still persisted in maintaining the traditional approach which is mainly based on translation tasks of single sentences from Japanese to English. Japan has a long tradition of employing the *yakudoku* approach to reading. Kawasumi (1975) gives a definition of *yakudoku*: Yaku means "translation" and doku means "reading." *Yakudoku* is defined as a technique or a mental process for reading a foreign language in which the target language sentence is first translated word-by-word, and the resulting

translation reordered to match Japanese word order as part of the process of reading comprehension. There have been many criticisms of the *yakudoku* approach particularly when the communicative approach began to be emphasized in English education in Japan. As a justification of the widespread practice of *yakudoku*, many analysts refer to its easiness for the teacher (Hino, 1988 ; Tazaki, 1978 ; Tajima, 1978). This *yakudoku* approach greatly affected writing curriculum and consequently translation tasks from Japanese to English of single sentences with emphasis on grammar, syntax, and mechanics were adopted as the main approach to writing. It emphasized accuracy ignoring the other important elements in writing tasks such as fluency, original ideas, communicative factors, purpose, and audience while the rest of the world started to explore various other approaches. Most Japanese instructors learned English in the traditional way so they tend to teach in the same way that they were taught. Unfortunately this conservatism has hampered the development of the writing curriculum in Japan. The rapid increase of students who study abroad is demanding fundamental changes in the curriculum of writing and also more textbooks which emphasize communicative writing are available than traditional and translation oriented writing textbooks today.

Assessment in writing

Writing evaluation is one of the crucial factors which affects the writing curriculum and instruction. Weir (1993) categorizes two different approaches for assessing writing ability. The first approach is the writing test type which can be focused on more specific 'discrete' elements, e.g. grammar, vocabulary, spelling, punctuation and orthography, and attempts can be made to test these formal elements separated by the use of indirect and objective tests such as the section of the structure and written expression in TOEFL test (Test of English as a Foreign Language). Japanese teachers of English working at high school or even at university levels acquired writing skills based mainly on the translation work mentioned above or cloze tests to ascertain grammar knowledge or words learned before.

The second type of writing test is a more direct, extended writing task. These involve the production of continuous texts of at least 100 words, with the writer being given some room for individual interpretation. Hughes (1989) emphasized the following three important specifications for writing tests in order to make the assessment valid and reliable:

- 1) Teachers have to set writing tasks that are properly representative of the population of tasks that they should expect the students to be able to perform.

- 2) The tasks should elicit samples of writing which truly represent the students' ability.
- 3) It is essential that the samples of writing can and will be scored reliably.

Hughes continues that in order to judge whether the tasks are representative of the tasks which teachers expect students to be able to perform, they have to be clear at the outset just what these tasks are so that students should be able to perform. Assessment of writing quality determines proficiency level placement, provides diagnostic criteria for selection of syllabus components, and affects the determination of final course marks (Kobayashi & Rinnert, 1996). Therefore, it is the test designer's responsibility to implement a good quality test in terms of validity and reliability.

Impressionistic scoring involves the assignment of a single score to a piece of writing on the basis of an overall impression of it. This kind of scoring has the advantage of being very rapid. According to Reid (1993 a), experienced scorers can judge a one-page piece of writing in just a couple of minutes or even less (scorers of the new TOEFL Test of Written English (TWE) will apparently have just one and a half minutes for each mark of a composition). The reliability tends to be low if the judges' assessment criteria is different from one to another. In the case of TWE each paper receives two independent scores, and any discrepancies are resolved by a third reading. Raters are strictly trained to follow TWE guide. Low intra-rater reliability hampers improvements in the interrater reliability, therefore it is significantly important to have a consensus among judges in writing assessment.

This study focused on the difference of the assessment criteria among raters who have different backgrounds; that is, the research question for this study is to see how raters' differences in background affect their assessment of writing. Difference in background is derived from many factors. Different learning experience mentioned earlier is one of the factors contributing to this difference. Rhetorical patterns in the first language which differs from culture to culture is also an important factor.

The various cultural rhetorical patterns in first languages and the influence of first language patterns on second language writing has been explored in past research represented by Kaplan (1966) and Hinds (1990). Kaplan states that logic which is the basis of rhetoric, evolves out of a culture; it is not universal. Logic varies from culture to culture and it changes even from time to time in one culture. Hinds offers another perspective of dichotomy, inductive versus deductive style in writing, though he claims that it is not a valid parameter for evaluating texts across languages. He states that while English writing tends to be deductive to make a

coherent composition, Japanese and other oriental languages tend to have an inductive style in writing. He attributes the difference to that of the reader's expectation. Shiokawa and Yoffe (1996) studied whether or not inductive organization is the prevalent model of expression in the written production of the Japanese learners of English. Their finding was that Japanese EFL learners without proper training in logical organization in writing display no specific organizational pattern, inductive or deductive. Hind's assertion that *ki-sho-ten-ketsu* (introduction-development-turn-conclusion) is the natural way of Japanese rhetoric which in turn affects their writing organization in English was unfounded in their study. They concluded that *ki-sho-ten-ketsu* is one of the patterns present in Japanese writing, however, it is not accepted as appropriate within the Japanese academic community, for example Japanese scholars follow a certain format which is commonly used in their own fields when writing academic papers. This study implies that Japanese writers' lack of coherence in their organization is not necessarily derived from cultural factor.

There is ample research on the perceptions of the readers in writing. Schwartz (1984) found that when two pieces of discourse are read by two different readers, the very text that pleases one reader may irritate the other. Mendelsohn and Cumminig (1987) stated that there were numerous factors which affect raters' perceptions of students' writing. If this finding is significantly related to raters' subjectivity in assessment, it causes a serious problem affecting the reliability in writing assessment. Brown (1991) found that while both the English faculty and ESL teachers considered content the most important feature, the English faculty paid more attention to sentence-level features such as cohesion and syntax, as opposed to the ESL teachers' stronger emphasis on organization. He suggested the two groups of teachers should cooperate with each other to lead their international students to academic success. Santos (1988) investigated if there was a significant difference in assessment between the faculty of the physical sciences and those in humanities and social sciences and she found that those in the physical sciences gave more severe judgments of language use errors than those in the humanities and social sciences. It means that professors have different expectations and assumptions from students depending on their majors. Kobayashi and Rinnert (1996) introduced L1 rhetorical features on readers' perceptions of L2 writing and they investigated how readers with different backgrounds (differing L1, academic status, and amounts of writing instruction) evaluated Japanese university students' English compositions with different culturally influenced rhetorical patterns (Japanese vs. American English) as well as two other features (coherence breaks and language use errors). Their findings show that other factors in their writing ability, including coherence and

language use, culturally influenced rhetorical pattern affected assessment of EFL student writing on an analysis of effects topic.

Teachers' feedback in writing

Variability in assessment of writing among the raters who have different background is the main focus for this study. In order to investigate the raters' mental mode in the assessment process, the effect of raters' feedback written on the L2 students' compositions was also studied. Freedman (1984) discussed about the significant influence of the readers' expectation and assumptions on their responses to student writing.

Zamel (1985) carried out an in-depth analysis on teachers' responses to student writing. She claims that teachers' marks and comments usually take the form of abstract and vague prescriptions and directives that students find difficult to interpret. She continues that ESL teachers rarely seem to expect students to revise the text beyond the surface level. This statement can be interpreted that teachers' comments are closely related to the validity of their measurement. Their responses are more likely used as detailed explanations to students how their papers were assessed. It is also assumed that such comments were also used by the raters themselves to reach their final assessment. For example, if there were numerous corrections for the mistakes made by a student, the rater also will convince himself or herself to give a low final grade by looking over him or her own feedback before deciding the final score. The student also will be convinced to obtain a low final score because of the numerous corrections. Therefore, it is concluded that the teachers' feedback which appeared as the interlinear comments is closely related to the validity of the assessment.

There are other types of feedback besides correction. Encouragement is an effective approach to improve the quality of students' writing, and even grammatical accuracy was improved when students received positive feedback from their teachers (Cardelle and Corno, 1981; Robb, Ross, and Shortreed, 1984). Fathman and Whalley (1990) found that students showed significant improvement in grammatical accuracy if the errors were underlined instead of giving corrections, and that comments on both grammar and content can be given at the same time without overburdening the students. Correction of errors is effective when the feedback concerning the error is clear; that is, the response must adequately describe the problem and suggest methods of correction. Moreover, students should be ready to learn, to commit to the change, and have the appropriate background knowledge to be able to revise their writing (Reid, 1993b). However, in many EFL contexts, the

majority of raters of English writing may be Japanese speakers whose expectations differ substantially from those of North American readers.

The purpose of the study

The purpose of this study is to investigate to see the differences in the criteria of writing assessment among the raters who have different backgrounds. This study investigated the following five research questions:

- (1) Do the two groups' raters score each component differently?
- (2) What expectations and assumptions of both groups of raters appeared in their feedback?
- (3) What are other variables besides language background (L1 or L2) which affect their assessment?
- (4) What are the students' reactions to different types of feedback?
- (5) Who is worth more; Japanese teachers or native speaking teachers for writing?

Method

The goal of this study was narrowed down to provide a descriptive and interpretive explanatory account of what the raters do when proceeding with their assessment of writing. In-depth analysis on their scoring and feedback written on the students' composition were the focus of this study. Some descriptive data; the means of the raters and the group mean in each criteria and the frequency of the significant types of responses were collected and analyzed to investigate the variability between the groups in their assessment in a triangulation method. Retrospective interviews with the raters and the students who wrote the compositions were conducted to validate the findings from the first data. Since the small number of subjects curtails the study's suitability for making generalizations, this descriptive data should be understood as being one of a naturally occurring phenomena without manipulation.

Writers

Five university students were asked to participate in this study. The students, who had taken TOEFL before, were chosen to know exactly about their language proficiency level. Their TOEFL scores ranged from 440 to 520 (Table 1). Three students among them have taken a course of expository writing before this study, however they had learned writing with the strict controlled approach which was through mainly translation tasks. Since two students had never taken any writing courses before, a short lecture (one hour) on paragraph writing was given to them. In the lecture they were taught how to develop paragraphs having a topic sentence

in each paragraph and supporting sentences to support the topic sentence within the same paragraph. It was emphasized the importance of logical development in paragraph writing, but no specific explanation of the different rhetorical patterns in English and Japanese was given. They were to choose any topic on which to write an 800-word-long type written composition without time limit. They were informed of, but not given an explanation of definitions of the six criteria; overall quality, grammar, word choice, coherence and cohesion, mechanics and creativity, upon which their papers would be evaluated. After their compositions were evaluated, the scores and feedback given by the raters were shown to the students and a 15 minute interview was given to them one-on-one. The following three questions were asked during the interview:

1. What do you think of the scores they obtained for their writing.
2. What do you think of the feedback they received from the raters.
3. Which type of feedback do they find useful to improve their writing skill?

Raters

Table 1 Demographic characteristics of the writers

students	gender	age	TOEFL	writing experience
S1	female	18	440	none
S2	male	20	475	1 year
S3	female	20	478	1 year
S4	male	22	467	none
S5	female	21	520	1 year

Three American instructors, all native speakers (E group) of English and three Japanese instructors (J group) participated as raters in this study. These subjects currently teach writing courses at the university level and have at least three years of experience teaching writing courses. In order to protect their privacy, all the raters' names are coded. During the initial meeting with them, the purpose of the study was fully explained to them and also how the research would be conducted because researchers are not only ethically responsible to their subjects, but also to other constituencies. The subjects were not informed of any data as to the students and the compositions were labeled with an identification number. They were instructed to read the five compositions once and rate them according to their first impression on six 5-point scales; overall quality, grammar, word choice, coherence and cohesion, mechanics, and creativity. The scaling guide was explained verbally to give ideas of criteria for scoring such as 1 for poor, 2 for needs improvement, 3 for

satisfactory, 4 for good and 5 for excellent. No clear definitions of the six components were explained. Next, the raters were asked to give feedback in the same way as they usually give in their writing classes and they were told to refer to their own feedback during the retrospective interview which was given two weeks after their task. The interview was conducted on a one-on-one basis for 20 minutes and the questions were asked based on the analysis of their scores and feedback. The other variables such as age, gender, academic status, and academic discipline might affect the readers' perceptions of the students' writings. Therefore, this information was included in the following descriptions of the raters.

Table 2 The description of raters

	age	nationality, gender	the length in Japan	Japanese proficiency	teaching experience	speciality
J1	45	Japanese/m	7 months		15 years	literature
J2	56	Japanese/m	9 months		25 years	linguistics
J3	28	Japanese/f	6 years		3 years	TESL
E1	29	American/f	2 years	basic	2 years	economics
E2	34	American/m	10 years	intermedia	7 years	journalism
E3	45	American/m	17 years	advanced	15 years	literature

Table 3 reports the means and the standard deviations of the scores given by the six raters for each of the six components. Table 4 shows that the mean comparison between the two groups in the six areas. The mean comparisons on the five students' writing performance among the six raters are indicated in Table 5.

The next section will analyze in detail the raters responses to the six areas.

Table 3 Summary statistics of rater's performance on the six components

	mean/SD	overall	grammar	word choic	coherence	mechanics	creativity
J1	3.37(0.85)	3.2	3.2	3.8	3.0	3.2	3.8
J2	3.27(0.79)	3.2	3.2	3.4	2.6	3.2	3.6
J3	3.17(0.70)	3.2	2.8	3.6	2.4	3.4	3.6
E1	3.57(0.73)	3.8	2.8	2.8	3.4	2.6	4.2
E2	3.30(0.75)	3.6	2.6	2.6	3.0	3.2	3.8
E3	3.70(0.84)	3.2	2.8	2.4	3.6	3.2	3.6

Table 4 Comparative analysis between the two groups in the mean scores

componen	J group	E group	mean difference
overall quality	3.2	3.5	0.3
grammar	3.1	2.7	0.4
word choice	3.6	2.6	1.0
coherence	2.7	3.3	0.6
machanics	3.3	3.0	0.3
creativity	3.7	3.9	0.2

Table 5 The mean comparisons on the students' performance

	J1	J2	J3	E1	E2	E3	the means
S1	3.0	3.2	3.0	3.2	2.7	3.2	3.1
S2	3.2	3.3	3.2	3.2	2.5	3.3	3.1
S3	3.5	3.0	3.2	2.8	3.5	3.0	3.2
S4	3.5	3.3	3.2	3.5	3.5	3.2	3.4
S5	3.7	3.2	3.3	3.7	3.5	3.0	3.4

Overall quality

The means of the six raters in overall quality didn't differ greatly from each other (J1;3.2, J2;3.2, J3;3.2, E1;3.8, E2;3.6, E3;3.2). The mean difference between the average scores of the two groups in overall quality was 0.3, which can be considered as a small variation. Even though the mean scores of the three Japanese raters were the same (3.2), they scored the five compositions quite differently. It means there are individual variations in assessing the overall quality among the raters. The results from the retrospective interview revealed that J1 and J2 used linguistic sophistication in sentence-levels as their major judgment to determine overall quality. They had the similar idea that accuracy should be the base line for their assessment and the content should be considered as an additional value to the base line. They stated that the final grade was determined by counting the number of mistakes or corrections. Even though they didn't count precisely, the mistakes or corrections were the dominant image of the papers. J3 relied on syntactical and lexical accuracy as the main factor to decide her scores in overall quality, but she also paid attention to the quality of paragraph organization. She stated, however, that it was difficult to pay attention to the both factors of accuracy and the quality of organization at the same time. Therefore, she first read the compositions without paying attention to the

errors so much, and at the second time of her reading she concentrated on correcting the errors. This might indicate that she could not combine the top-down approach with bottom-up approach as a reader when she identified a number of errors in the compositions. The errors distracted her from reading continuously to evaluate the quality of organization.

The E group also showed some variations in the scores of overall quality, but the common factor all three raters in the E group used to determine the overall quality was organization and coherence and cohesion. E1 and E2 emphasized naturalness of the flow as the most important criteria to measure their overall quality of the compositions. They relied on their intuitive judgment to evaluate naturalness and they looked at the writing tasks as a means of communication. If the message is clear enough, the writing can be considered as good writing even though there are some surface level mistakes. E2 added the importance of originality in the students' composition and he stated that the overall grade was not necessarily an average of the five sub components. E3 stated that the most important factor to determine the quality of the compositions depended on the paragraph organization, but he added poor linguistic features might hamper him from reading smoothly, and consequently it might influence him while evaluating the organization. It seemed that E3 shared a similar view with J3 considering the importance of accuracy in addition to the quality of organization.

The overall quality was referred to by J2, J3, E1, E2, and E3 in their general comments which appeared in the end of the paper using four or five sentences. J1 was the only rater who didn't give any overall comments in sentences. All raters felt that assigning a single grade for overall quality holistically was the most difficult task because some papers were strong in surface-level, yet weak in organization and development. J1 and E2 included the strengths and weaknesses of the students' writing in their overall comments as follows;

J 1 : Overall quality is good, especially the content is excellent, but many mistakes with articles and tenses lowered the quality of the writing.

E 2 : You are a creative writer, but you tend to mix present and past tenses and this is confusing for the reader.

Grammar

The mean comparison of the raters and between the group for grammar also indicated a small variation (J1; 3.2, J2; 3.2, J3; 2.8, E1; 2.8, E2; 2.6, E3; 2.8, the mean difference between the groups; 0.4). As past research revealed, Japanese raters

respond to grammar errors much more frequently than do native speaking raters (Kobayashi, 1992). The variations in the types of corrections of grammatical errors made by Japanese raters were also identified. J1 and J2 corrected almost all of the grammatical errors that occurred in the compositions and they admitted that they browsed through the mistakes and corrections when they decided the overall score. Their ways of correcting errors were different from each other. J1 simply corrected the errors, but J2 included grammatical instructions to explain why they were grammatically wrong using Japanese language in addition to giving the right answers. The following is one of the instructions he made for the students.

J 2 : Fortunately, I was decided to study abroad from 9 February.

Naturally our meals are make by myself.

Intransitive verbs (verbs that do not take a direct object) cannot be made passive.

Only transitive verbs (verbs with object and indirect objects) can be made passive.

J2 included suggestive responses to change the original forms to more authentic English sentences by stating "Your sentence is grammatically acceptable, but native speakers would not write like this, so I suggest that this sentence would carry more accurate meanings of your opinions." J3 made the least grammatical corrections among the Japanese raters, but her mean score was lower than the ones of the other Japanese raters (2.8). She did not provide full corrections on the errors, and some minor errors were underlined. She corrected only serious errors which affect the comprehensibility of the text, and she provided grammatical explanations on those global errors in her general comment written at the bottom of the papers. She stated that Japanese students were under tremendous pressure when writing English composition and the pressure is usually caused by their strong consciousness of grammar mistakes, and in order to moderate the students' pressure she uses underlines to indicate the mistakes as implicit corrections. She continued that such implicit corrections are proven more effective in her experience because the students would repeatedly continue making similar mistakes caused by their carelessness so that implicit corrections help the students to become more independent learners. The careful observation on the E group's assessment in grammar revealed that they gave many fewer corrections for grammatical errors than did the Japanese raters, but the three raters' grades for this sub component were not high (E1; 2.8, E2; 2.6, E3; 2.8, the group mean; 2.7). The three raters in the E group shared the same opinion that the students' grammatical errors which appeared frequently did not annoy them so much as long as the content is clear enough and the surface-level problem did not

affect their overall evaluation either, however since they were asked to evaluate this sub component of grammar analytically their close observation on this aspect led them to the low grades. One of the native speaking raters, E2 focused on the errors which frequently occurred in the compositions and corrected only those errors. He stated that even native speakers often made grammatical errors in both areas of speaking and writing, so a high level sophistication of grammar from learners of second language should not be expected. The other rater, E1 emphasized that the grammatical mistakes Japanese learners often make are articles and tenses and if she starts correcting all the local errors, the job would be endless and also she would lose concentration on understanding the text, therefore the best solution is to give a separate instruction to review these areas which cause difficulties to Japanese learners. E3 was the only rater who used the abbreviated marks to indicate the errors and he said he usually gave a brief explanation for what these abbreviated marks meant in the beginning of his writing classes. He added that he would concentrate on judging the content of the writing rather than spending too much time correcting the students' grammatical errors.

Word Choice

The criteria of word choice showed the second largest difference between the two groups. The average score of the J group and E groups were 3.6 and 2.6 respectively and the mean difference between the group was 1.0. The three native speaking raters gave corrections or suggestions on word choice much more frequently than did their Japanese counterparts. This finding matches with the results of Kobayashi's research (1992) on native and nonnative reactions to ESL compositions. He concludes that the errors of words are probably due to semantic transfer from Japanese and he implies that the loan word is a potential pitfall faced by Japanese ESL teachers. He added that since today's Japanese language finds itself flooded with innumerable loan words of English origin, it is highly likely that many Japanese learners of English fail to notice that such words may have different semantic properties and connotations from those of the original English word. Japanese ESL instructors, as well as learners, could overlook such an error because they take the meaning of the word for granted. On the other hand, native speakers of English detect such odd usage at a glance, especially the two raters of E2 and E3 who have been in Japan longer than E1. They detected more loan words from Japan because they shared a great number of schemata with the writers, which facilitated their comprehension. E2 and E3 provided a variety of alternative lexical items. The followings are some of the examples of made by the two raters.

This is really Japanese. They have "a lot of pride."

E 2 : **People care about how people look at them, and have high pride to maintain a good reputation.**

Again, it sounds like Japanese. Change to "in the eyes of their teachers."

E 3 : **Students want to have good reputations from their teachers.**

Peterson (1988) argues that a sentence starting with "especially" is one of the most frequently observable expressions in English compositions written by Japanese. Two Japanese raters, J1 and J3, left this awkward expression uncorrected, but J2 whose speciality is linguistics was more sensitive to word choice than the other raters in the J group. He included detailed explanations of why these words were more appropriate as follows:

If you want to say "tanitsuminzoku", "homogeneous" is a right word

J2: **Japan is a single nation.**

"chuumokuwo atsumeru" = drawing attentions

J2: **The news is gathering attentions today.**

Do you mean "jimaku" ?, then subtitles or captions are more commonly used in English.

J2: **Among many countries only Japan uses superimpotion to show foreign movies.**

Coherence and Cohesion

Coherence in written text is a complex concept, involving a multitude of reader- and text-based features. Perhaps because of this, writing instructors and textbooks often discuss coherence in a vague or incomplete manner. However, it is often heard from university professors in America that foreign students' academic writing is often "incoherent." John (1986) introduces three principles to guide instructors in teaching the concept of coherence:

- 1) Coherence is text based and consists of the ordering and interlinking of propositions within a text by use of appropriate information structure (including cohesion).
- 2) At the same time, coherence is reader based. The audience and the assignment must be consistently considered as the discourse is produced and revised.
- 3) Instructors have an obligation to teach coherence comprehensively, that is, to take into account these two approaches (text based and reader based), at a minimum.

The findings of this study revealed that coherence and cohesion showed the widest range in the assessment among the raters. The average scores of the J group

and the E group were 2.7 and 3.3 respectively and the mean difference was 0.6. The reason behind this variation is seemingly due to the different interpretations of coherence and cohesion among the raters. All the three raters in the J group focused on the topic sentence in each paragraph and how well it was developed in the written work to judge the quality of cohesion and coherence. Even though the students received a one hour explanation of how paragraphs are developed by focusing on topic sentences and supporting sentences, there were many coherence problems identified by the Japanese raters. J2 and J3 provided further explanations of how they evaluated the quality of cohesion and coherence as follows:

- J 2 : The most important aspect to evaluate cohesion and coherence is to see how each sentence is sticking to the central idea. The appropriate usage of conjunctions and reference items such as this, the, or it are the key points for me when judging the quality of paragraph organizations.
- J 3 : I always pay attention to how successfully the students are linking sentences through use of vocabulary. It is important to see how these related words appear throughout the paragraph? The inking of vocabulary is the key point to see the organization of the students' paragraphs.

While the interpretation of the J group on coherence was defined as text based, the raters in the E group were likely to be reader based oriented when they assessed the quality of coherence and cohesion in the students' writings. A text cannot be considered separately from the reader and that experience requires successful interaction between the reader and the discourse to be processed (Carrell, 1982; Rumelhart, 1977). As a reader based approach requires the raters to go through an interactive and interpretive process while they are reading the students' writing products. Therefore, they believe that successful writers must continuously keep the intended audience in their mind. E3 elaborates this point as follows:

- E 3 : When I finish the students' writing, I always turn the paper down and try to summarize the central idea introduced through the written work. Good quality papers in terms of coherence and cohesion give me an easy time to summarize the ideas, on the other hand I have a hard time to pin point the central ideas after I read poorly organized papers.

J3 added in her comments that when she saw some cases in which reference items are wrongly used, she tended to give lower grades for this criteria. Too much emphasis on text based coherence is probably the major reason why the J group's average grade was lower than that of the E group.

Mechanics

Most of the students had a basic knowledge of English mechanics, therefore all the raters agreed that the quality of mechanics was satisfactory. S1 had the largest number of mechanical problems, specifically he made three mistakes when splitting words based on syllable at the end of line and four raters of J1, J2, E2, and E3 stated that the mechanical errors unconsciously or subconsciously affected their assessment in the section of overall quality and consequently. S1 gained the lowest average score among the five subjects (Table 5).

Creativity

All of the six raters agreed that even in the case of a poorly written paper, if the content included personal views of the student they found the writing interesting. Personal experience, belief and emotional appeals provoked the readers' attention and response. Japanese preference for balancing ideas rather than taking one side (Harder, 1980) was identified in the compositions of S1 and S2. and, E2 responded to the S2's writing as follows:

E 2 : You lost a point on personal voice. You don't sound confident enough.

The rater E1 found most of the students' compositions original and interesting. He stated that he enjoyed reading the compositions because the views originated in Japanese culture and Japanese ways of thinking were interesting for him. Since his length of staying in Japan is relatively shorter than the other two raters, the information provided in the writings seemed to provoke his attention.

Raters' feedback

The various types of comments were adapted by the six raters because the subjects were instructed to give feedback in the usually way. Their comments appeared in the following forms: 1) question marks, 2) crossed lines over original form, 3) underlined lines 4) bracketing original form, 5) brackets to insert new words or expressions, 6) ticked marks to insert some words, 7) arrows to change specific words or expressions, and 8) abbreviated words to indicate the type of errors such as "art" for articles and "spe" for spelling errors. Since these marks were used arbitrarily, they were not focused on as part of this study. The usage of writing tools such as red pencils or green pens seemed to be habitual behavior even though E3 claimed that he particularly used a green pen because of the effectiveness of a less authoritative image using green rather than red.

Table 6 The types of responses

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8
J1	145	93	45	13	22	9	4	0
J2	205	114	44	28	24	10	8	12
J3	105	70	27	4	19	6	0	14
E1	71	55	10	0	8	18	0	19
E2	89	58	8	0	6	21	0	23
E3	107	83	22	0	10	22	0	29

Item 1=time consumption (minutes)

Item 2=the average number of total feedback

Item 3=the average number of corrections

Item 4=the average number of suggestions

Item 5=the average number of grammar explanations

Item 6=the average number of alternative words or expressions

Item 7=the average number of changing locations of words or expressions

Item 8=the average number of personal responses

Table 7 The mean comparison of the types of responses between the two groups

	J group	E group
Item 1	152	89
Item 2	96	65
Item 3	39	13
Item 4	15	0
Item 5	22	8
Item 6	8	20
Item 7	4	0
Item 8	15	24

Table 6 shows a variety of feedback made by the six raters including the time spent assessing and writing their responses. The feedback identified in the compositions were categorized under the following types: (1) corrections, (2) suggestions on surface errors, (3) grammar explanations, (4) alternative words or expressions, (5) changing locations of words or expressions, and (6) personal responses to the opinions of the writers.

Table 6 also shows the types of responses which were identified in the raters' comments and also the frequency of the occurrence of the responses. Table 7 shows the mean comparisons of these types of responses between the groups. Time difference in scoring the five papers showed a large variation among the raters (J1; 145 minutes, J2; 205, J3; 105, E1; 71, E2; 89, E3; 107). The J group spent almost two times longer time than did the native group (The average time consumption of the J group and E group were 152 minutes and 89 minutes respectively). The reason was that the

J group gave 70% more feedback than did the E group (the average number of responses for each group was J group; 96 and E group; 65). The J group's feedback mainly consisted of corrections (J group; 59, E group; 13) while personal reactions were more predominant in the E group's feedback (J group; 5, E group; 24). The J group made quite a large number of grammatical explanations (J group; 22) while the E group made quite a few lexical explanations (E group; 20). J1 and J2 changed the locations of the words or expressions because they thought the sophistication of the sentences would improve if the location of a certain words or expressions have been changed. It seemed that they were sensitive to the style of the sentences because of their specialized majors (J1; literature, J2; linguistics).

The retrospective interview validated the two groups of native and Japanese raters as having the following contrastive views of students' writings. J1 and J2 stated that they had felt obligated to correct as many errors as possible and they concentrated to evaluate the papers accurately. J3 also shared the similar view to the other raters, but she made less corrections than the other two because she wanted to enjoy reading the students' writing to discover the strengths rather than the weaknesses in students' papers. All the three raters in the E group said they had enjoyed reading the compositions as readers and they had felt that the main responsibility as a rater was to improve the quality of the papers so that they made more personal reactions which are sometimes critical opinions. On the other hand, the J group didn't make so many personal reactions as the E group, but they gave a large number of suggestions to improve sentence-level features. However, the two raters J3 and E3 demonstrated a balanced view on both sentence-level and also content-level features.

A close qualitative observation of the data shows that the J group used more negative reactions and less positive feedback for their comments than did the E group. The observed responses in the J group's assessment which were categorized as negative responses were as follows:

1. You had better review basic grammar. It is very important.
2. This expression is awkward. Use an English-English dictionary to learn correct ways of using the word.
3. These are careless mistakes.
4. Getting away from the central theme.
5. This is really confusing. Make your point clear.
6. Organize your ideas more. It's out of focus.
7. Review the basic rules of punctuation.

8. This is an embarrassing error.

Some of the negative responses above overlapped with the responses of asking for clarification, but compared to the comments made by the E group and even though the responses had the purpose of asking the students to clarify their intention, the J group made more direct, instructional, and command type comments.

The factor to validate this tendency is that the J group could figure out the students' language proficiency level considering the subjects' age (The raters were informed of the students' age). They had a clear idea of what knowledge of English an average college student should have. They used it as a standard level and evaluated the students' writing based on the criteria. Therefore, the negative feedback implies that the students should not make such mistakes if they are university students. The E group, on the other hand, naturally had less knowledge of this aspect except E3, so that they just read the students' compositions without any prior knowledge as to the students' language proficiency. One of the native raters, E1 confessed that she had a sense of admiration for the students' ability because she herself learned French as her second language, but her ability in French is not good enough to write this level of compositions.

Instead of using negative responses, the native group used responses of "asking for clarification" frequently such as "I don't understand your point. Clarify this point." "How is this sentence related to the main topic?" "Who said this? Is this your opinion?" By looking at this characteristic of the native group's evaluation it indicates that they looked at the students' writing as work in process and numerous positive phrases were used to guide the students to future revision.

Pure praise such as "Good" "good idea" "yes" were often written in interlinear or in margins by the E group. Praises aimed at features of the text such as "good examples" "It's well described." "good expressions". were also used. The comments which indicate the raters' interest or concern such as "This is interesting." "Really?" "It's amazing!" The J group admitted that this type of comments would be useful to motivate the students, but they simply didn't use them because they didn't know these types of reactions were possible to use. J3 gave her comment on this matter:

J 3 : Teachers in Japan are still a symbol of authority, but in an innovative view , teachers should be facilitators for students, but this concept is hardly rooted in the Japanese educational system. Composition teachers, therefore, have the biggest duty to find students' errors and correct them as language teachers, but I personally admit that this kind of positive feedback will definitely motivate the students and eventually they will improve their writing skill by having less pressure. It is the most important environ-

ment Japanese learners should be given; an anxiety free learning environment.

Students' reaction

The five students who wrote the compositions were shown the papers corrected by the six raters. It was not explained who the raters were and they were just asked how they interpreted these scores and comments. All the five students stated that they would have preferred their fully corrected papers by J1 and J2 and they had a doubt if the raters truly read their work if the papers were not corrected enough. Also, the full corrections which appeared in the J1 and J2's evaluation were helpful for the students to understand the meanings of the scores, however they could not relate the scores with the comments written by the E group. They stated they did not have confidence in writing sentences in English so that they expect thorough corrections by raters. However, they admitted the fact that the personal reactions made by the E group had succeeded in motivating them and they strongly felt the urge to improve the paper to better quality one. They felt as if they were communicating with the reader and they gained a sense of achievement. On the other hand, the comments made by the J group were sometimes too impersonal and they felt slightly intimidated and even lost confidence in their own writing skill.

The characteristics of their views appeared in their feedback are summarized as follows;

- 1) The J group used linguistic accuracy as the main factor to determine the holistic scores of overall quality, while the E group relied on the quality of the organization.
- 2) The main feedback made by the J group were corrections while the E group made numerous comments to ask for more clarifications to the writers.
- 3) The J group were sensitive to grammatical errors while the E group paid attention to the lexical errors.
- 4) The E group was looking at the students' writing as work in process while the J group considered the work as final products.
- 5) The E group was reading the students' writing as readers, on the other hand the J group were evaluating them as language teachers.
- 6) The E group expected the students to revise according to the changes that they wanted to impose on the text, but the J group evaluated more objectively and impersonally.

There were some variations among the raters in both groups. It seemed that J 3 had slightly different views from the ones of the other two Japanese raters. It is

more likely that J3 shared a similar view with that of E 3. This can be explained that J3's experience of learning writing skills mainly in the U.S.A. had influenced her assumptions and perceptions toward writing. Also E 3 who has been in Japan for a longer time than the other two raters seemed to have been influenced by the education environment where he has been, therefore, he was more sensitive to accuracy than were the other two raters.

There were many more marks indicating unknown intentions by the use of marks, underlines, and exclamations marks by the E group and almost no such ambiguous comments in the J group's evaluations. The J group had all corrections consistent and made the purpose of the corrections clear enough that the students received the message.

This study implies that the philosophies each rater has in term of writing tasks are diverse, and they reflect different expectations and assumptions, different pedagogical objectives, and different perspectives in their assessment. Apparently J1 and J2 in the J group who had more experience of the traditional approach attempted to improve the students' accuracy with their great attention to the discrete elements in students' writing as language teachers and the assumptions reflected in their responses observed earlier. E3 in the native group who has been in Japan for 17 years was obviously influenced by the Japanese teaching approach so that he was more sensitive to accuracy and sentence-level errors. On the other hand, J3 who had a 6 year experience of studying in America paid attention to the cohesion and coherence quality by focusing on the linking words to see the natural flow in the context. E1 and E2 in the native group put an emphasis on the students' fluency rather than accuracy and attempted to have communication with the writers without discrete language elements. The dichotomy between fluency and accuracy has existed in language teaching for decades (Lennon, 1991), and on which should the language teacher concentrate?

The final question in this study is that who is worth more; Japanese teachers or native speaking teachers for writing. Japanese teachers' top priority is to use English accurately and appropriately and in terms of fluency Japanese teachers are heavily handicapped compared to native speaking teachers. However, Medgyes (1990) gives strong encouragement to non-native teachers by raising the following reasons;

1. Non-native teachers can serve as imitable models of the successful learner of English.
2. Non-native teachers can teach learning strategies more effectively.

3. Non-native teachers can provide learners with more information about the English language.
4. Non-native teachers are more able to anticipate language difficulties.
5. Non-native teachers can be more empathetic to the needs and problems of their learners.
6. Non-native teachers can benefit from sharing the learners' mother tongue.

Lee (1989) puts an emphasis on the collaborative work of Japanese teachers and Native speaking teachers to supplement each others' weaknesses. Precise evaluation of the students' needs are important before EAP curriculum is implemented. If students need more practice on accuracy before improving fluency in writing courses, students should be provided surface-level instructions and if their basic competence is proven they should be given ample opportunities to practice free writing. In the case of Japanese students they are strictly trained to have grammatical understanding in their writing, but their limited knowledge of academic forms and audience expectations may result in a serious obstacle to success when they study abroad, therefore EAP curriculum designers should demonstrate clear a understanding of their future needs and implement new objectives without any biases and prior assumptions.

References

- Brown, D. J. (1991). Do English and EFL faculties rate writing samples differently? *TESOL Quarterly*, Vol. 25, No. 4, pp. 569-603.
- Cardelle, M & Corno, L. (1981). Effects on second language learning of variations in written feedback on homework assignments. *TESOL Quarterly*, Vol. 15, No. 3, pp. 251-261.
- Carrell, P. L. (1982). Cohesion is not coherence. *TESOL Quarterly*, Vol. 16, No. 4, pp. 479-488.
- Cumming, A. (1990). Expertise in evaluating second language composition. *Language Testing*, (7), pp. 31-51.
- Fathman, A. K. & Whalley, E. (1990). Teacher response to student writing: Focus on form versus content. In B. Kroll (Eds.), *Second Language Writing: Research insights for the classroom*. pp. 178-190. New York; Cambridge University Press.
- Freedman, S. W. (1984). The registers of students and professional expository writing: influences on teachers' responses. In *New Directions in Composition Research*. Richard Beach and Lillian S. Bridwell (Eds.), pp. 334-347, New York: Guilford Press.
- Harder, B. (1980). Cultural value differences in the stylistic problems of English compositions written at the 20th International Conference of Orientalist, Tokyo.
- Hendrickson, J. M. (1984). The treatment of error in written work. In S. McKay (Ed.), *Composing in a second language*. pp. 145-159, Rowley, MA: Newbury House.
- Hinds, J. (1990). Inductive, deductive, quasi-inductive: Writing in Japanese, Korean, Chinese, and Thai. In U. Connor and A. Johns (Eds.), *Coherence in Writing*. pp. 87-110. Alexandria, VA: TESOL.

- Hino, N. (1988). Yakudoku; Japan's dominant tradition in foreign language learning. *JALT Journal*, Vol. 10, No. 1&2.
- Hughes, A. (1989). *Testing for Language Teachers*. Cambridge Handbooks for Language Teachers.
- Johns, A. (1986). Coherence and academic writing; Some definitions and suggestions for teaching. *TESOL Quarterly*, Vol. 20, No.2. pp. 247-265.
- Kaplan, B. R. (1966). Cultural thought patterns in inter-cultural education. *Language Teaching*, Vol. 16, pp. 1-20.
- Kawasumi, T. (1975). Yakudoku no rekishi, *The English Teachers' Magazine*. July, Special issue, pp. 14-19.
- Keio University, (1993). *Journal writing: Pedagogical Perspectives*, Keio University, SFC (Shonan Fujisawa Campus), Institute of Languages and Communication.
- Kobayashi, H. & Rinnert, C. (1996). Factors affecting composition evaluation in an EFL context: Cultural rhetorical pattern and readers' background, *Language Learning*, Vol.46, No.3, September: pp. 397-437.
- Kobayashi, T. (1992). Native and nonnative reactions to ESL compositions, *TESOL Quarterly*, Vol.26, No.1. pp. 81-111.
- Lee, S. (1989). The effectiveness of team-teaching methods in utilizing Japanese and native speaking instructors, *Osaka Jogakuin Kiyo*, No. 20.
- Lee, S. (1990). The necessity of a pre-academic program in Japan. *Osaka Jogakuin-Kiyo*, No. 21.
- Leki, I. (1989). *Academic writing: Techniques and Tasks*. New York: St. Martins Press.
- Lennon, P. (1991). Error: Some problems of definition, identification, and distinction. *Applied Linguistics*, 12 (2), pp. 180-196.
- Medgyes, P. (1990). Native or non-native: who's worth more? In Tricia Hedge & Norman Whitney, *Power pedagogy and practice*, Hong Kong, Oxford University Press. pp. 31-42.
- Mendelsohn, D., & Cumming, A. (1987). Professors' ratings of language use and rhetorical organizations in ESL compositions, *TESOL Canada Journal/ Revue TESL du Canada*, 5 (1), 9-26.
- Raimes, A. (1991). Out of the woods: Emerging traditions in the teaching of writing. *TESOL Quarterly*, Vol. 25 No. 3, pp. 407-431.
- Reid, J. (1993a). *Teaching ESL Writing*. Engliwood Cliffs, NJ: Prentice Hall.
- Reid, J. (1993b). *The process of paragraph writing (2nd. ed.)*. Englewood Cliffs, NJ: Prentice Hall.
- Rumelhart, D. (1977). Toward an interactive model of reading. In s. Dornic (Ed.), *Attention and Performance*. Vol. 6, pp. 33-58. Hillsdale, NJ: Lawrence Erlbaum.
- Robb, T. N., Ross, S. & Shortreed, I. (1984). A large-scale study of feedback methods in EFL composition. Paper presented at the 18th Annual TESOL convention, Houston, March 1984.
- Peterson, M. (1988). *Japanese English*. Tokyo: Iwanami.
- Santos, T. (1988). Professors' reactions to the academic writing of non-native speaking students. *TESOL Quarterly*, Vol. 22, No. 1, pp. 69-90.
- Schwartz, M. (1984). *Response to writing; A college-wide perspective*. *College English* 46 (1): 55-62.
- Shiokawa, H. & Yoffe, L. (1996). Rhetorical patterns in the composition of Japanese EFL learners. The JACET 35th Annual Convention Theme: Cross-Cultural Interaction and the

Teaching of English, Ohtani university; Kyoto.

Tajima, K. (1978). The grammar–translation method: Its historical and social background. In I. Koike, M. Matsuyama, Y. Igarashi, & K. Suzuki (eds.), *The Teaching of English in Japan*. Tokyo; Taishukan.

Tazaki, K. (1978). *Theories of English Teaching*. Tokyo; Taishukan.

Weir, C. (1993). *Understanding & Developing Language Tests*. Prentice Hall International (UK) Ltd.

White, R.V. (1988). *The ELT Curriculum, Design, Innovation and Management*. Mass; Blackwell.

White, R.V. & Arndt, V. (1991). *Process writing*. London: Longman.

Zamel, V. (1982). Writing: The process of discovering meaning. *TESOL Quarterly*, Vol. 16, No. 2, pp. 195–209.

Zamel, V. (1985). Professors' reactions to the academic writing of non native speaking students. *TESOL Quarterly*, Vol. 19, No. 2, pp. 79–101.