A Composition Program Using Criterion^{sm}

Rebecca L. Benoit

Chiben Gakuen Wakayama

For senior high teachers searching for more efficient ways to teach composition skills for Japanese university entrance English exam practice, essay-grading software can provide assistance. Though there are several types of software available that can do this, some are too expensive and impractical for most teachers to consider implementing (Hearst 2000). However, Criterion^{sm} (by ETS Technologies) may provide teachers with a cost-effective option. Moreover, 60 of the 99 prompts are similar to past Japanese university entrance exam questions. As of 2002, over 60 of Japan's national and public universities have free-composition questions on their entrance exams. Results from on-going use of Criterion^{sm} by 350 Japanese students at a private academic senior high school confirm that the program has been beneficial in stimulating revision and positive attitudes toward writing.

大学入試対策として、自由英作文を指導している高等学校の教師にとって、英作文を自動評価するソフトウェアは効率的で非常に有効である。いくつかのソフトウェアが存在するが、なかには、非常に高価すぎて実際には導入しかねるものもある。しかし、ETSの開発によるクライテリオン（Criterion^{sm}）は、少ない費用で大きな効果が得られる優れたソフトウェアである。同ソフトの扱う99の練習問題のうち、実に60題が大学入試問題に類似している。平成14年現在で、60校以上の国公立大学の入試で自由英作文が出題されている実状をうけ、智辯学園和歌山高等学校では、現在クライテリオン（Criterion^{sm}）を使用して自由英作文の指導をしている。その結果、生徒自身がコンピュータを使って自分の英作文を何度も推敲するなど、積極的な活動が可能になった。最後に、大学入試に向けた自由英作文の指導を実りのあるものとするため、各大学の入試担当者に自由英作文の採点基準を公表することを提案する。

Introduction

The Japanese university English entrance exam (JUEEE) is often a controlling force in determining curriculum priorities at the senior high level. Gilfert (1999) reported that one area often neglected in English classes is communicative writing. This is a growing concern since as of 2002 over 60 of Japan's 95 public and national universities (and a select number of prestigious private institutions) have free English composition

(FEC) questions on their entrance exams (see Appendix, figures 1 and 2). Although these questions do not carry the weight of translation or reading comprehension sections, they still constitute a significant percentage of the exam that high school teachers cannot afford to ignore; particularly since answering these questions can be difficult to most students without adequate, or indeed any, preparation. Gilfert's article proposes that native speakers assist students with writing, which can be challenging with limited time and large classes. In some cases, writing instructors teach hundreds of students, making it difficult to give feedback on thousands of essays a year.

Chiben Gakuen Wakayama & Criterion[sm]

Much of the discussion surrounding product-oriented writing programs involves their failure to produce long-term improvements in student writing, and that teacher feedback tends to produce an "inconclusive, sometimes contradictory" impact on student writing (Leki, 1997, p. 66). In order to prepare students for exams and establish a writing curriculum, Chiben Gakuen Wakayama's English department needed to answer the following questions: How could students write better-organized essays with richer vocabulary? How could our limited time as raters be put to better use? Finally, how would we get our students to adopt a process-oriented approach in order to improve the quality of their writing? Though our students were aware of the importance of writing, very few revised their essays based on our corrections or comments. Graders spent hours correcting compositions for students who simply put them back in their desks without anything more than a cursory glance. There simply had to be a more effective way to give students writing training and practice without overloading teachers.

Computerized grading and feedback of essays seemed as though it might be a possible solution. Reid states that "students react positively to CALL use; they find revision easier, they enjoy working with computers, and they believe that helps to improve their writing" (1993, p.43). Although, he warns, "teacher feedback remains the most important part of the ESL writing class". Criterion[sm], a browser-based application by ETS Technologies, provides us with an answer to our composition curriculum problems.

Whether one believes in the validity of using computers for essay grading or not, it is, and will likely remain, a practical part of writing assessment for the foreseeable future (see Cushing Weigle, 2002; Reid, 1993; Ferris and Hedgecock, 1998; Cunningham, 2000, for more on benefits and drawbacks of computers for L2 composition). Though there are several types of essay-grading software available (e.g. IEA, PEG, MITRE), most are either too expensive or impractical for most schools to consider adopting (for a comparison see Hearst, 2000; Little, 2001).

Criterion[sm] provides our program with objective holistic scoring and feedback on grammar, usage, mechanics, style and organization within 30 seconds of an essay's submission (using e-rater[TM], the scoring engine, which generates the holistic score). We have found that this instant feedback allows students to revise their essays in class. Prior to our adoption of Criterion[sm], students had to wait several days to weeks for feedback on their essays which significantly inhibited their motivation to revise.

In addition, Criterion[sm]'s composition manual gives students the freedom to explore essay instruction, examples and advice at their own pace; thus "giving teachers more time to help their students improve their overall writing skills" (Smith quoted in ETS press release, 2001) and students extensive reading practice. Furthermore, the program's instructor comment and annotation

features ("Pop-Up Notes"), allow teachers and students to exchange e-mail messages when in-class conferencing is impossible. Teachers can also customize the program to allow all the students in a certain class to view each other's work, thereby giving students model essays to study (Criterion^{sm} provides model essays on all topics at every score level). Additionally, instructors can design custom prompts based on past JUEEE-FEC questions for students to practice. Although Criterion^{sm} cannot grade these essays, students still have access to all the feedback analysis features and teachers can input sample answers to these questions from Japanese publishers or write their own. Furthermore, in order to help students with limited vocabularies avoid repetition of words (Criterion^{sm} cites this as a problem in nearly 95 percent of our students' essays) we encourage them to use the thesaurus in Microsoft Word. To reduce the amount of time teachers spend commenting on surface errors, Criterion^{sm} does an initial screening of essays. If, after several drafts, certain grammatical problems persist, instructors explain how to correct these mistakes. At a base price of $6.50 USD per student for unlimited use of up to 111 questions, Criterion^{sm} is more affordable than many commercially available writing textbooks. Finally, since the service is web-based, teachers and students can work from any computer with Internet access and at any time; thus reducing problems with lost essays and composition paperwork.

Since most Chiben students prepare for the exams of Japan's top public, national and private universities, an additional consideration for the administration was whether Criterion^{sm} questions were similar to past JUEEE-FEC questions and were satisfied to learn that 60 of the 111 prompts are comparable in theme. Criterion^{sm} thus provides our students with a cost-effective, relevant preparation tool for exam practice.

Process-oriented approach: portfolios

By the time our students graduate, they will have written at least 40 short essays ranging from 100 to 400 words, which we estimate as the minimum needed for students to write well in English. While this number may seem impressive compared to that of other high schools without FEC programs, it is still far from adequately preparing them for the one to five-page, reference-supported university assignments that await them. Additional challenges facing writing instructors is that our students have no essay training in their L1 until their last year of high school, when they write 12 short essays over the course of the year. Insufficient composition training in L1 is another possible explanation why many of our students have reported initial difficulty writing in English (Takagi, 2001; Kasuya, 1999, p. 1). Furthermore, English writing teachers only have two hours a month to teach the basics of composition (one in the classroom and one in the computer lab for Criterion^{sm}).

All 265 of our second grade senior high students (and 90 third graders) have a year to revise their Criterion^{sm} essays until they are satisfied with their work, and write an additional ten essays on paper from topics in their reading textbooks. Criterion^{sm} scores constitute up to 15 percent of their reading course grade. The rationale for delayed assessment was that "extending the amount of time allotted to producing an essay might affect the level of mastery exhibited" (Kroll, 1997, p. 140).

Procedure

Students receive guided writing worksheets for homework. They begin by outlining in either Japanese or English, based upon whether they have "topic-area knowledge" of the question

in their L1 or not (Friedlander in Kroll, 1997, p. 124). These worksheets include reminders about using logical connectors (e.g. first, second) and a variety of vocabulary related to the question. Since "without being more conscious of the differences in Japanese and English rhetorical patterns, student writers cannot convey their ideas to native speakers of English clearly," these worksheets are also designed to give students an understanding of rhetorical differences between English and Japanese (Takagi, 2001).

Students are required to achieve a minimum score of 2 to 'pass' each essay (the maximum possible score is 6). The score feature is highly motivational and the excitement it generates in class often results in further motivating them to improve their own work in order to increase their scores. Both teachers and students try to foster the attitude that "the act of composing should become the result of a genuine need to express one's personal feeling, experience, or reactions, all within a climate of encouragement" (Zamel quoted in Reid, 1993, p. 31).

This process of writing also leads students to self-discovery. For example, in their first essay, students were asked, "If you could make one important change to a school that you attended, what change would you make?" Initially, some said that there was nothing they wanted to change with their current situation. However, through the process of brainstorming with classmates, they began thinking about possible school improvements, and proposed constructive changes through their writing.

Working from Reid's (1993) claim that "imposed change may force students to *perform*, but it will not motivate them to *become*" (p. 124, emphasis in the original), comments given to students about revision are prefaced with the understanding that making changes in future submissions is voluntary. On average they revise 3 to 5 times and some return to earlier essays later

in the year as their writing ability develops. Instructor feedback is initially motivational, and then proceeds to address rhetorical structure and content development. When teachers respond to content, students have a greater tendency to take the advice seriously (Reid, p. 163) and certainly once students realize that revising according to instructor comments are likely to result in a higher score, they are even more motivated to do so.

Benefits and limitations

The debate regarding the validity of grading criteria, the e-rater's vulnerability to being 'fooled', or the ethics of having students motivated by grades assigned by a computer certainly warrant future research, though such a discussion is beyond the scope of this paper (for more on tricking the e-rater™, see Powers et al., 2001 and Little, 2001 p. 11, Yang et al., 2001). In 1999, Wilson reported inter-rater reliability between e-rater™ and trained faculty readers between .87 to .94 but Burstein et al. now state this at .93 to .94 (2002) while, inter-rater reliability (human to human) ranged from .87 to .93 in 1999 (Wilson).

Despite the advantages Criterion℠ offers, it is not a panacea. Students are routinely reminded that Criterion℠ is not "perfect" or an ultimate source of authority on their writing. Recently, a second-grade student said the following,

> "I scored a 6, but I think it is only because I wrote many words. I know my ideas don't go in a straight line, so I want to fix it, but I don't know what to write" (Sakaguchi).

This student's suspicion of her score's validity demonstrates a potential grading-technology shortcoming. At the same time, the realization that her essay lacked focus and that, despite a

perfect score of 6 she was still motivated to seek teacher advice on how to improve her composition demonstrates her emerging consciousness as a writer. These and many other similar incidents illustrate the positive effect that the application has on our students, at least in terms of their attitude towards writing and, in particular, the process of revision. Ideally, this is the effect we would like to achieve with all our students.

Although it is still early to analyze grading and writing quality improvements from a full year, student attitudes to Criterionᴿᴹ are clear: they enjoy using it, look forward to our classes, and enjoy themselves. Students are also requesting more "Criterionᴿᴹ lessons"; at the end of the year, some students asked whether they could use Criterionᴿᴹ next year. During the two-week trial in 2001, when asked the question, "Did you enjoy using the Criterionᴿᴹ?" All 20 of the 29 students who attended the 2 sessions replied "yes" (9 students were absent) and when asked the question, "If you could use Criterionᴿᴹ in university, would you?" 23 of the 29 students also responded positively.

Backwash from JUEEE–FEC questions

During a university trial entrance exam designed to imitate the exams of Tokyo, Kyoto, Osaka and Kobe universities offered nationally by Z-KAI in November 2002, 262 of our 265 students answered the FEC question on the exam. In 2001 (prior to use of Criterionᴿᴹ), Chiben students averaged 9.7% higher than the national average on the FEC question while in 2002 the average was 18.9% higher than the national score. On a similar exam given by Benesee in February 2003, our students averaged 13.7% higher than the national average (an increase from 10.5% in the previous year). There may be a correlation between these results and their Criterionᴿᴹ practice (see Appendix figure 6).

Though we were delighted to see that nearly all of our students had applied what we had taught them about writing, we later learned from a company representative that Z-KAI graders considered length more important than content, going even so far as to give students whom had written more than 105 words a score of zero. On another practice exam by Yoyogi Zeminar for Osaka University (October 2002), a student wrote an essay clearly answering the question (see Appendix figure 4 and 5 for question, grading criteria and sample student essay) and scored zero. While the company admitted to having no contact with Osaka University in establishing their grading criteria, this did little to calm our upset student. The effect that these exams have on students should not be underestimated; students are often under the impression that company-run tests are as real as the actual exams. Consequently, without information from universities, we are left wondering what we should advise students to do on exams with unclear FEC directions. This issue further complicates teaching, as several Japanese teachers of English (JTEs) I have worked with over the years believe that grammar and word-length are more important to university graders correcting essays than ideas or rhetoric, though some university native English composition instructors consider the opposite to be true (Guest 2003, p. 14). This certainly makes composition instruction difficult in team-teaching contexts when JTEs and native English teachers hold opposing views on what aspects of writing should be emphasized in class.

Conclusion

Teaching students how to write well in English is more important than ever. Since JUEEE-FEC questions inevitably create backwash, it certainly would be helpful for universities to do as ETS has done with the TWE portion of the TOEFL

test and make grading criteria and model answers available to examinees. Alternatively, exam writers might consider establishing a consensus on FEC questions with the companies, which hold the mock exams in that university's name. If all of this is impossible, at the very least giving more comprehensive instructions which indicate the importance of organization, or whether ideas are more important than grammar, would certainly be more fair (as Fukui Prefectural and Sophia Universities are doing (see Appendix, figure 3). Gilfert's recommendation to encourage native SHS teachers to give composition instruction is a good one; those of us involved in preparing students for these exams have a counter-proposal: make the grading criteria for these essay questions more transparent so we can better prepare students for the exams and beyond.

References

Burstein, J. et al. (2002). *Computer Analysis of Essays. Research paper for ETS*. Retrieved December 2002 from <http://www.ets.org/research/dload/ncmefinal.pdf>.

Cunningham, K. (2000). Integrating CALL into the Writing Curriculum. *The Internet TESL Journal*, Vol. VI, No. 5, May 2000.

Cushing Weigle, S. (2002). *Assessing Writing*. Cambridge: Cambridge University Press.

ETS. (2001). *Prentice Hall School and ETS Technologies Launch Web-based Essay Scoring Service: e-rater(tm) Grades Essays Online in Seconds*. Retrieved December 2002 from <http://www.ets.org/news/01020101.html>.

Ferris, D. & Hedgecock, J.S. (1998). *Teaching ESL Composition: Purpose, Process, and Practice*. Mahwah, NJ: Lawrence Erlbaum Associates.

Friedlander, A. (1997). Composing in English: Effects of a first language on writing in English as a second language. In B. Kroll (ed.), *Second Language Writing, Research insights for the classroom* (pp. 109- 125). Cambridge: CUP.

Guest, M. (2003, February 28). Indirectly Speaking: When is 'accuracy' not really accurate? [The Language Lab] *The Daily Yomiuri*, p. 14.

Gilfert, S. (1999). Let's Write in English: Teacher, we never learned that. *The Internet TESL Journal*, Vol. V, No. 4.

Hearst, M. (2000). *The Debate on Automated Essay Grading*. Retrieved December 2002 from <www.knowledge-technologies.com/presskit/KAT_IEEEdebate.pdf>.

Kroll, B. (1997). The Teaching of Topical Structure Analysis as a Revision Strategy for ESL Writers. In B. Kroll (ed.), *Second Language Writing, Research insights for the classroom* (pp. 140- 154). Cambridge: CUP.

Leki, I. (1997). Coaching from the margins: issues in written response. In B. Kroll (ed.), *Second Language Writing, Research insights for the classroom* (pp. 57- 68). Cambridge: CUP.

Little, J. (2001). *Computerized Evaluation of Essays*. Retrieved December 2002 from <http://newark.rutgers.edu/~gshafer/little.pdf>.

Powers, D. et al. (2001). *Stumping e-rater. Challenging the validity of automatic essay scoring*. Retrieved December 2002 from <http://www.ets.org/research/dload/powers_0103.pdf>.

Reid, J. (1993). *Teaching ESL Writing*. Englewood Cliffs, New Jersey: Prentice Hall Regents.

Sakaguchi, M. (2002). Private communication in class to author.

Takagi, A. (2001). The need for change in English writing instruction in Japan. *The Language Teacher*, 25(7), 5-9.

Kasuya, M. (1999). *Teaching English textual patterns to Japanese students*. Student paper for University of Birmingham's M.A. TESL/TEFL program. Retrieved December 2002 from <http://www.cels.bham.ac.uk/resources/essays/MichikoM4.PDF>.

Wilson, D. (1999). The state of the GMAT. *Business Week*, November 11th. Retrieved December 2002 from <http://www.businessweek.com/bschools/originals/bs91111.htm>.

Yang, Y., Buckendahl, & C., Juszkiewicz, P. (2001). *A Review of Strategies for Validating Computer Automated Scoring*. Paper presented by the Gallup Organization (Nebraska) at the annual meeting of Midwestern Educational Research Association, Chicago. Retrieved December 2002 from <http://www.unl.edu/BIACO/MWERA/MWERA%20CAS%202001.pdf>.

Appendix

Figure 1: List of public and national universities which have or have had FEC questions on their JUEEE as of 2002:

Aichi Teachers' College, Akita University, Ehime University, Fukui University, Fukui Prefectural University, Fukui Medical University, Fukuoka Teachers' College, Gifu University, Gunma University, Hirosaki University, Hiroshima University, Hitotsubashi University, Ibaragi University, Iwate University, Kagawa University, Kagoshima University, Kanazawa University, Kita Kyushu City University, Kobe Foreign Studies University, Kobe Commercial College, Kochi University, Kumamoto University, Kyoto Teachers' College, Kyushu University, Miyazaki Teachers' College, Nagasaki University, Nagasaki Prefectural University, Nagoya City University, Niigata University, Oita University, Osaka University, Osaka City University, Osaka Teachers' College, Osaka University of Foreign Studies, Osaka Women's College, Ryukyu University, Saga University, Saitama University, Shiga University, Shiga Prefectural University, Shimane University, Shimonoseki City University, Shinshu University, Shizuoka University, Shizuoka Prefectural University, Tohoku University, Tokyo University, Tokyo Gakugei University, Tokyo University of Foreign Studies, Tokyo Prefectural University, Tokyo Institute of Technology, Tokyo Science & Technology University, Tokyo Teacher's College, Tsukuba University, Tsuru Bunka University, Utsunomiya University, Yamagata University, Yamaguchi University, Yamanashi University, Yokohama University, Yokohama City University

Figure 2: List of Japanese private universities which have or have had FEC questions on their JUEEE as of 2002:

Note: Due to the hundreds of private universities in Japan, and difficulty getting published materials on all of their exams, this list may not be comprehensive.

> Aoyama Gakuin, Chuo University, Hamamatsu Medical University, Kansai University, Kansei Gakuin, Keio University, Konan University, Ritsumeikan University, Seishin Women's University, Sophia University, Waseda University (Law)

Figure 3:

Instructions from Fukui Prefectural University FEC questions from 1996-2001 (the 2002 exam has not been published yet): "Write an essay (not a list) in English on the following topic. Your essay should be well organized. You will be given marks for expressing your ideas clearly. Communicating your thoughts is even more important than using correct grammar. Write one full page (200-250 words)." (emphasis in the original)

Instructions from Sophia University FEC questions (2002): "Answer only one of the following questions. Write at least three paragraphs. You may write more than three if you want…Give reasons for your opinion."

Figure 4:

Yoyogi Zeminar question for Osaka practice exam (held October, 2002): (translated from the original Japanese by author):

> "What do you think is the most important reason why students go to university?
> Write about 70 words. Total points: 25."

Grading criteria disclosed by Yoyogi Zeminar employee on December 2nd, 2002:

> If a student writes 35 – 55 or 85 – 105 words, deduct 10 points
> If a student writes under 35 or over 105 words, deduct 25 points (score zero)

Figure 5:

Student model answer for the Yoyogi Zeminar practice exam question listed in Figure 4 (transcribed without any changes from original):

> I think the most important **things** in college life is what we learn in college. And the point if very often said recently.
>
> First today college student is often said to go to college for no purpose. I think the system in Japan society caused this situation. For example, people regarded people who graduate from college as a valuable man. People are apt to think that only to graduate from college is good.

Second, what we learn in college is the most important for job in future. For example, we can't be a doctor, a dentist, a lawer and so on without learning in college.

In conclusion, I think if we have a purpose to go to college surely, we can spent a college days as valuable time.

(128 words, the exam grader, "Oda", wrote "123 words" and crossed out the student's entire response after underlining 'things' in the first line).

Figure 6:

These Z-KAI questions are modeled after prestigious national universities such as Tokyo, Osaka, Kyoto and Kobe University entrance exams (original questions in Japanese, translated by the author):

2001: Starting in April of 2002, Saturdays in all public schools will be abolished. What do you think about this? Start your answer with "I agree with this system.." or "I disagree with this system…" and write 40 – 50 words in English.

2002: Every January, the coming-of-age ceremony is held. Recently however, some say this custom should be abolished. What do you think? Start your answer with "I agree with the idea of abolishing this ceremony.." or "I disagree with the idea of abolishing this ceremony..". Write 50 – 60 words in English. Do not count commas (,) or periods (.) as one word. Similarly, compound words like coming-of-age should be counted as one word.

NOTE: When contacted, a Z-KAI representative stated that this year, students who wrote more than 86 words had 10 points deducted and that those who wrote over 105 words received a score of zero (of a possible 20).