

Exploring the Potential of a Web-Based Writing Instruction Program and AES: An Empirical Study Using My Access

Sangmin Lee (Kyung Hee University)

Lee, Sangmin (2008). Exploring the potential of a web-based writing instruction program and AES: An empirical study using My Access, *Multimedia-Assisted Language Learning*, 11(2), 103-125.

This study aims to comprehensively investigate the potential benefits of My Access, a web-based writing instruction program, in EFL writing classrooms, focusing specifically on the Automated Essay Scoring (AES) system and qualitative feedback. To examine the reliability of AES in My Access, the holistic and analytic scores of 6 middle school student essay samples rated by My Access were compared with the scores given by 5 human raters. As an extension of writing evaluation, My Access provided qualitative feedback on the student essays and gave various revision goals for each aspect of writing based on process-oriented writing. The types and strategies of the feedback utilized by My Access were carefully examined and compared with those given by the human raters. The results of the quantitative analysis using AES showed that the quality of the AES embedded in My Access was high enough to be used in writing classrooms. The qualitative analysis provided by My Access revealed that it has the potential to facilitate student writing by guiding students through the processes of writing and motivating them to revise their essays. The study concluded that My Access could be beneficial for writing teachers and student writers when used creatively and effectively.

I. INTRODUCTION

As the Internet has rapidly expanded the opportunity for communication and writing, writing ability has become more important than ever. In cyberspace, people's writing ability represents their ideas, knowledge, emotions, and even their own selves. It may not be too much to say that writing ability is necessary for thriving in cyberspace as well as the real world. This emphasizes the importance of writing in EFL classrooms. However, writing has been considered one of the most difficult classroom skills to convey (if not the most difficult) by both teachers and students. For teachers, evaluating writing and providing feedback to a large body of students is a great burden. Moreover, evaluating student essays requires trained readers (or raters), but most Korean English teachers are not fully prepared due to a lack of experience in teaching EFL writing. On the student side, the current Korean classroom situation (large size, test-focused environment) does not allow many opportunities for students to explore writing in English.

To mitigate the teachers' burden and support student writing, Seoul Metropolitan Office of Education offers an online writing instruction program, My Access, to elementary and secondary school students for free. Currently, 250 students (50 elementary school students, 100 middle school students, and 100 high school students) are enrolled in the writing courses on this site. These writing courses are offered as part of an advanced English curriculum; hence, the language proficiency level of the students in these courses is much higher than the average level of their peers. To manage the courses and support student writing, 25 middle school and high school English teachers work in tandem with My Access.

As a writing instruction program, My Access is grounded in process-oriented writing and guides students from planning to revision. Since My Access is equipped with an Automated Essay Scoring (AES) system, it can diagnose and assess a large number of essays in a short time. In addition, its scoring criteria are valid and pedagogically sound as compared to other widely used writing assessment standards; thus, it provides valuable guidelines for novice writing teachers. In addition to providing scores, My Access gives qualitative feedback about the writing, which is intended to motivate students to revise their essays. However, despite the potential benefits for writing classes, it is necessary to review the program's functions and verify its effectiveness in teaching writing and in terms of educational outcomes because online writing programs and AES are still in the rudimentary stages in Korea. This paper proposes two research questions as follows:

- 1) To what degree does My Access, as an AES system, provide reliable scores of

student essays as compared to the scores of trained human raters?

2) What types of qualitative feedback are available on My Access?

In answering these questions, this paper also intends to offer insight into the potential benefits and limitations of My Access as well as its educational implications in EFL writing classrooms.

II. LITERATURE REVIEW

As the role of writing in English education increases, there is a great demand for valid and reliable ways to measure student writing ability. Assessing writing should be included in the student learning cycle; hence, "scores, grades, and evaluative feedback should consistently contribute to writer's learning processes and to the improvement of their measurable writing skill" (p. 300, Ferris). Writing assessment, when performed in responsible and appropriate ways, can serve a pedagogically valuable function and increase student writers' proficiency.

The quality of writing assessment depends heavily on the reliability and validity of the assessment. Reliability in writing assessment refers to the consistency of the essays' scores after multiple ratings by trained raters (Henning, 1991). Validity reflects whether an instrument actually measures what it purports to measure. Face validity, often termed "surface credibility," allows students to perceive the instrument as measuring what it is intended to measure. Content validity permits students to activate knowledge domains, skills, or strategies and bring them together to constitute the composing process (Wiegler, 2003). Although a group of scholars claim that reliability is a prerequisite to validity, as Ferris (2005) and Hamp-Lyons and Henning (1991) point out, it is fair to say that they are interdependent. Writing assessment can have a positive washback effect (the influence - both beneficial and harmful - of the test on classrooms or curricula) when it is reliable and valid (Bachman & Palmer, 1996; Bennet, 1999; Powers, Chodorow, Fowles, & Kukich, 2002; Wiegler, 2003).

Approach to scoring is another prominent issue in writing assessment (Choi, Cho, Shin, & O, 2005). There are three general approaches to scoring: holistic, analytic, and trait-based scoring. Each has unique advantages and challenges. In this section, only holistic and analytic scoring, which are relevant to the issue of the present paper, will be discussed. On the one hand, holistic scales emphasize the general impression of the essay and its overall quality, focusing mainly on the analytic and rhetorical qualities of the

essay. Analytic rating scales, on the other hand, are concerned with discrete components of the essay such as content, style, organization, mechanics, convention, and focus, which are closely tied to communicative effectiveness and accuracy of the essay (Rudner & Gagne, 2001). While holistic scoring demonstrates high validity, it does not reveal much about student writing (Penny, Johnson & Gordon, 2000; Wiegler, 2002). Additionally, when the student receives the score, it may often be too vague for him/her to understand the result because it does not identify his/her weaknesses and strengths in writing. In contrast, analytic scoring can provide more information on student writing and has high reliability, yet the validity is "of questionable because it isolates text features from context" (Barkaoui, 2007). Moreover, this scoring method cannot avoid the criticism that the parts cannot make the whole.

Another noteworthy issue when assessing writing in recent years is Automated Essay Scoring (AES), since computers are becoming useful tools to facilitate the evaluation of student essays. Assessing writing is time-consuming. In addition to fatigue, human raters can be subjective, even if they use guidelines or rubrics in assessing student essays (Attali & Burnstein, 2006; Hout, 1996). In contrast to human raters, AES is cost-effective, highly objective, consistent, and impartial (Attali & Burnstein, 2006; Huff & Sireci, 2001). However, skepticism and criticism about AES have, of course, accompanied this appraisal. Unlike human raters, AES systems are incapable of reading sophisticated emotions and humor, or distinguishing between creative and inspirational essays (Powers et al., 2002). This criticism is ascribed to the fact that AES often relies heavily on the surface features of writing, ignoring overall writing qualities and other important aspects of writing such as content, meaning, and audience rapport. One of the oldest AES systems, Project Essay Grade (PEG), does not, in fact, consider the meaning or content of writing, but rather uses only surface features. PEG, a regression model, uses approximations (proxes) for intrinsic variables (trins) that are usually indicators of the syntax and conventions of the essays. However, as technology evolves, AES also improved and became capable of measuring many different features of writing. For instance, Intelligent Essay Assessor (IEA) is a good tool to measure content of essays. Latent Semantic Analysis (LSA) is also able to take into account content of the essays, but is not a good tool in terms of measuring writing styles and rhetorics (Hout, 1996). E-rater, developed by the Educational Testing Service (ETS), by taking the advantages of PEG and IEA, can analyze the syntactic structure (surface features), the rhetorical structure, and content (topical analysis). These AES systems have been proved reliable through numerous inter-rater reliability (consistency of scores among different raters) tests. Reliability of PEG marked

as high as .87, and E-rater V.2. reached over .90, closely to the apparent reliability of the target judge groups (Powers et al, 2002).

In addition to AES, many writing instruction applications are also available. One of the best examples, Criterion, is a web-based writing instruction service also developed by ETS that evaluates a student's writing skills and instantly provides a score along with diagnostic feedback. Embedded in Criterion are writing analysis tools, which identify discourse structures and errors in grammar, usage, and mechanics, and provide feedback about the errors and discourse elements present or absent in the essay. Providing qualitative feedback to students is particularly important to EFL student writers since a number of studies have concluded that students' errors in writing essays, including syntactic and lexical errors, were not disappearing without explicit or implicit corrections "as a natural consequence of focus on students' ideas or writing processes" (p.-43, Ferris, 2003).

In fact, feedback on student writing is critical to its improvement. Numerous writing scholars maintain the importance of writing as a process rather than a single, fixed, final product. This approach to writing specifically emphasizes multiple drafting cycles, teacher feedback, and substantive revisions rather than a single submission of a final product. Nonetheless, in practices like evaluating and assessing writing, providing teacher feedback for writing requires teachers to invest significant time and effort. According to previous studies, 1) students consider teacher commentary extremely helpful for their writing development (Hyland & Hyland, 2006); 2) students value teacher feedback on a variety of issues, not just language error corrections or ideas; 3) correcting every single error in the student essay does not help student writing development; 4) both constructive criticism and encouragement through feedback are necessary (Hamps-Lyons & Henning, 1991); 5) teachers should treat their students as individuals and consider their feedback as a part of the writing processes and a conversation between them; 6) not only written feedback, but also face-to-face oral feedback through student-teacher conferences can greatly promote student writing development (Hamps-Lyons, 2003); and 7) feedback should focus not only on language-related problems, but also include meaning-related problems (Weigle, 2002).

Because error correction is crucial to EFL student writing development, a number of researchers have compared different types of error feedback. According to Ferris (2005), one key dichotomy is the distinction between direct and indirect error feedback. Direct feedback is defined as the explicit correction of linguistic forms by the teacher to the student. This includes writing the correct expressions or words, deleting unnecessary words or morphemes, and inserting missing morphemes or words. Indirect feedback, on

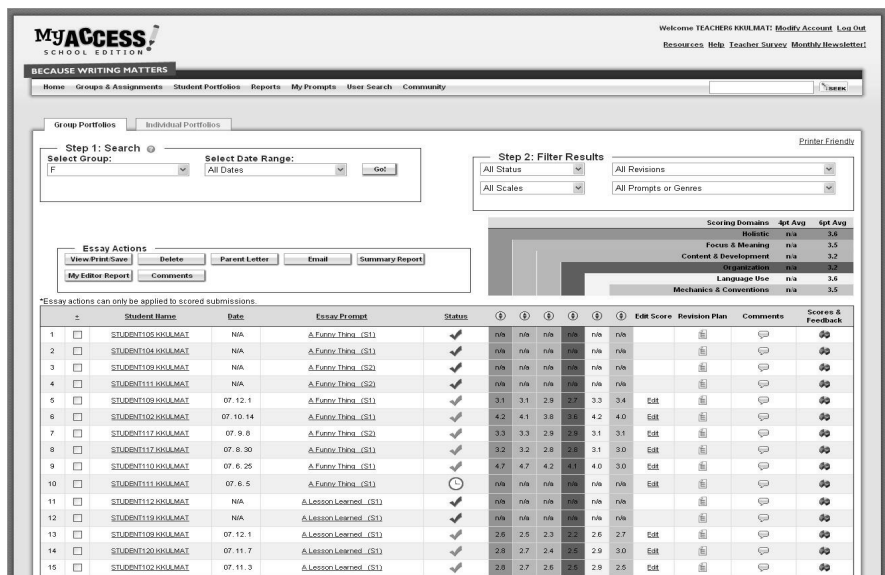
the other hand, occurs when the teacher implicitly indicates the errors without explicitly correcting them. Indirect feedback encompasses marking the errors (with or without further guidance for correction). However, because of limited linguistic resources and the resulting need to focus on language rather than content, EFL writing in English is often hampered. Thus, keeping a balance between direct and indirect feedback is key to responding to student writing.

III. METHODOLOGY

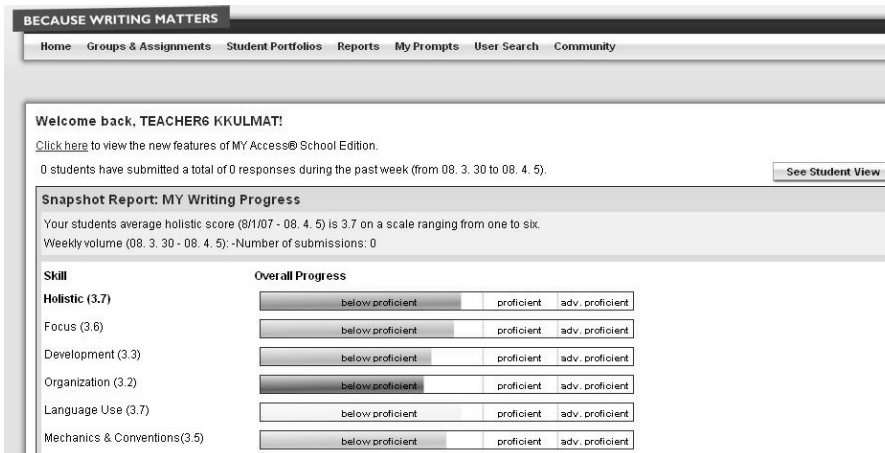
1. Program Description

My Access is a web-based writing program that is installed with the AES system, Intellimetric. This AES evaluates a large number of student essays holistically and analytically in a short time. Both the holistic and analytic scoring rubrics comprise six bands (or scales). The analytic scoring scale has five components: content and development; focus and meaning; organization; language use, voice, and style; and mechanics and conventions. Similar to E-rater, My Access can assess surface features, rhetorical qualities, and content. Its reliability and validity will be discussed in a later section of this paper.

As a writing instruction application, My Access (which is based in process-writing) offers several distinctive features with which teachers can customize writing tasks for their own classes. First, a teacher can register her students individually or as groups. Then, she can create a writing prompt or select one from the archive on the program. She can set up the due date and the standards (or use the rubrics that My Access has already established) for the task. When the task is published online, students can view the prompt and start the task. After students upload their essays, My Access automatically assesses them. Students' grades are marked on both the holistic and analytic scales. The teacher then reviews each student's grade and, if necessary, adjusts the grade given by My Access and adds comments. She can also view a comparative group report to get information about the whole class and compare individual students. Figure 1 shows the portfolios of the students of the particular teacher, and Figure 2 gives the averages for each aspect of essays written by her students. This page allows the teacher to view the holistic and analytic scores of the individual essays and move to the next action, such as editing scores, viewing students' revision plans, or adding comments on the essays.



[Figure 1] The Overall View of My Access



[Figure 2] Average Scores of Each Aspect of Writing

With regard to the students, once an essay prompt is given, My Access encourages the students to focus on the writing process through various planning activities. For example, the students can visualize their thoughts about the topic using a Venn diagram or organize them into tables and charts, all provided on My Access. After the pre-writing stage, the students write essays and submit them. The students can later check their grades on the My Tutor page and their feedback on the My Editor page. The My Editor page shows the overall grades (holistic and analytic) and the interpretation of the grade

(e.g., what the score indicates about the student's proficiency level in comparison with the whole group). The My Editor page further explains the student's errors in details. The My Tutor page suggests revision goals, revision samples for each category, and gives a writing sample.

2. Data Collection

1) My Access - The Program

My Access was one of the primary data sources for this research. It was examined from two perspectives: as a writing instruction program and as an AES. As a writing instruction program, the entire cycle of the writing process and its pedagogical design, detailed features, feedback and comments provided by My Access were reviewed. In particular, feedback was carefully examined, categorized, and compared to the feedback given by the human raters. To verify it as an AES, reliability and validity checks were conducted in comparison to human ratings and previous studies. SPSS 13 was used to test reliability.

2) Student Essay Samples

The researcher reviewed 176 essays written by 41 middle school students who enrolled in the writing classes of three teachers for the fall semester of 2007. The mean number of essays written by each student was 4.2, where the maximum number of essays written by a student was 7, and the minimum number of essays was 2. Among these essays, 34 essays from 8 students were randomly selected for in-depth analysis. Each essay was carefully examined and scored holistically and analytically by five trained human raters.

The scores given automatically by My Access and the scores rated by human raters were then compared for reliability. The qualitative feedback offered by My Access and by human raters was also reviewed and compared to check whether there were any significant differences between them. In addition to one-time evaluation of the essays, four students were randomly selected and all of their essays produced over the semester were reviewed and further analyzed to investigate writing development and the influences of feedback on student writing.

3) Human Raters

The human raters who participated in this research included a professor in English education and four Ph.D. candidate English teachers. None of them were directly involved

in the My Access writing classes offered by the Seoul Metropolitan Office of Education. The three middle school teachers who participated in My Access classes were not included in this research for the reason that they were already tuned to the scoring standards of My Access; thus, it was unlikely that they would be useful for obtaining reliable results on inter-rater reliability tests.

The raters used the rating rubric developed by My Access to verify the reliability of the AES of the program. The raters used the rubrics provided by My Access in Korean for a better understanding of the rubrics. During rating, the raters did not consult with one another and submitted their results individually. Like My Access, the raters assessed the essays holistically and analytically, corrected the errors, and provided overall comments on content, development, and organization, and specific comments on the errors. It took 10 days for the raters to finish scoring the 34 essays.

3. Data Analysis

1) Inter-rater Reliability

Table 1 shows the descriptive statistics of the scores from My Access and the five raters for each category. The mean number of words in the samples was 348.06, and the mean readability level was between 4th and 5th grades (readability is calculated based on American English textbooks).

[Table 1] Descriptive Statistics of the Scores

| | N | Minimum | Maximum | Mean | Std. Deviation |
|------------|----|---------|---------|--------|----------------|
| Word Count | 34 | 121 | 880 | 348.01 | 218.9 |
| My Access | 34 | 2.0 | 5.9 | 3.7 | 1.198 |
| Rater 1 | 34 | 2.5 | 6.0 | 3.8 | 1.296 |
| Rater 2 | 34 | 1.5 | 5.4 | 3.1 | 1.046 |
| Rater 3 | 34 | 2.0 | 5.5 | 3.4 | 1.093 |
| Rater 4 | 34 | 1.0 | 5.5 | 3.6 | 1.326 |
| Rater 5 | 34 | 1.0 | 5.0 | 3.3 | 1.263 |

Prior to verifying the reliability of the AES of My Access, inter-rater reliability (Pearson's correlation coefficient) among the five human raters was calculated. Table 2 indicates that inter-reliability among the human raters was significantly high. Three out of six cases were significant at the 0.01 level (2-tailed) and two were significant at the 0.05 level (2-tailed).

[Table 2] Correlation Coefficients Among the Human Raters

| | | Rater1 | Rater2 | Rater3 | Rater4 | Rater5 |
|--------|---------------------|----------|----------|----------|----------|----------|
| rater1 | Pearson Correlation | 1 | .453 | .891(**) | .722(**) | .553(*) |
| | Sig. (2-tailed) | . | .068 | .000 | .001 | .021 |
| | N | 34 | 34 | 34 | 34 | 34 |
| rater2 | Pearson Correlation | .453 | 1 | .527(*) | .442 | .746(**) |
| | Sig. (2-tailed) | .068 | . | .030 | .076 | .001 |
| | N | 34 | 34 | 34 | 34 | 34 |
| rater3 | Pearson Correlation | .891(**) | .527(*) | 1 | .848(**) | .722(**) |
| | Sig. (2-tailed) | .000 | .030 | . | .000 | .001 |
| | N | 34 | 34 | 34 | 34 | 34 |
| rater4 | Pearson Correlation | .722(**) | .442 | .848(**) | 1 | .757(**) |
| | Sig. (2-tailed) | .001 | .076 | .000 | . | .000 |
| | N | 34 | 34 | 34 | 34 | 34 |
| rater5 | Pearson Correlation | .553(*) | .746(**) | .722(**) | .757(**) | 1 |
| | Sig. (2-tailed) | .021 | .001 | .001 | .000 | . |
| | N | 34 | 34 | 34 | 34 | 34 |

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

Next, inter-reliability between My Access and the human raters was calculated. Table 3 displays that the scores from My Access were highly correlated with four of the human raters and had a weak correlation with the remaining human rater (Rater 2). Table 4 provides a closer look at the correlation coefficients of each aspect for My Access and Rater 1. This result reveals that all aspects except Focus & Meaning showed strong correlations between My Access and Rater 1. The analysis on correlations with the other human raters in each aspect reports similar results.

[Table 3] Correlation Coefficients Between My Access and Human Raters

| | Rater 1 | Rater 2 | Rater 3 | Rater 4 | Rater 5 |
|-----------|---------|---------|---------|---------|---------|
| My Access | .778** | .530* | .871** | .816** | .817** |

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

[Table 4] Correlation Coefficients Between My Access and Rater1 in Aspects of Writing

| Dimension Coefficient | Holistic | Focus & Meaning | Content & Development | Organization | Language Use, Voice & Style | Mechanics & Conventions |
|-----------------------|----------|-----------------|-----------------------|--------------|-----------------------------|-------------------------|
| | .778** | .586* | .681** | .780** | .634** | .651** |

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

3) Qualitative Data Analysis

In brief, My Editor focuses more on the surface features of the essay and adopts a quantitative approach to direct error corrections, whereas My Tutor is concerned with promoting writing strategies and revision skills, with a qualitative approach to writing and implicit feedback.

(1) My Editor Feedback

As shown in Figure 3, error details of the essay were shown on My Editor Feedback. The comments on each error were embedded in the essay, but those comments were confined to mechanics and language usage; in other words, surface features of the essay. The categories of errors that My Access detected included spelling errors, missing articles, punctuation errors, capitalization, pronoun errors, clause errors, and sentence formatting errors. The comments were highlighted in blue, and each comment was hyperlinked to lead the student to view the explanation and an example of a corrected one. For example, in the sentence "Anyway this is why I am crazy about this man. A <Clause errors> great scientist, thinker <Punctuation errors> and peace keeper" My Access pointed to two errors of the surface features. When the student clicks on "Clause errors," he/she can view why this is not a correct sentence (because it is an incomplete sentence) and how he/she should correct it (My Access recommends revising it to "He is a great scientist, thinker, and peace keeper.").



Error Detail

A<Clause errors> Person who I<Formatting errors> Admire

A lot of time<Missing articles> has passed since man started knocking stones together. Now the world is full of<Similar words> high technology and innovations. The reason we could live in this modern world is because we had people who studied the wonders and mysteries of our world. People like Archimedes<Spelling errors>, Socrates, Newton, Dalton, Einstein, Curie and many others helped in mankind's development. Without them we would be back in the stone age still grunting about how to create fire. So who<Pronoun errors> do I admire the most? I admire the great man Albert Einstein. Why do I admire him? Well<Punctuation errors> let's find out shall we?

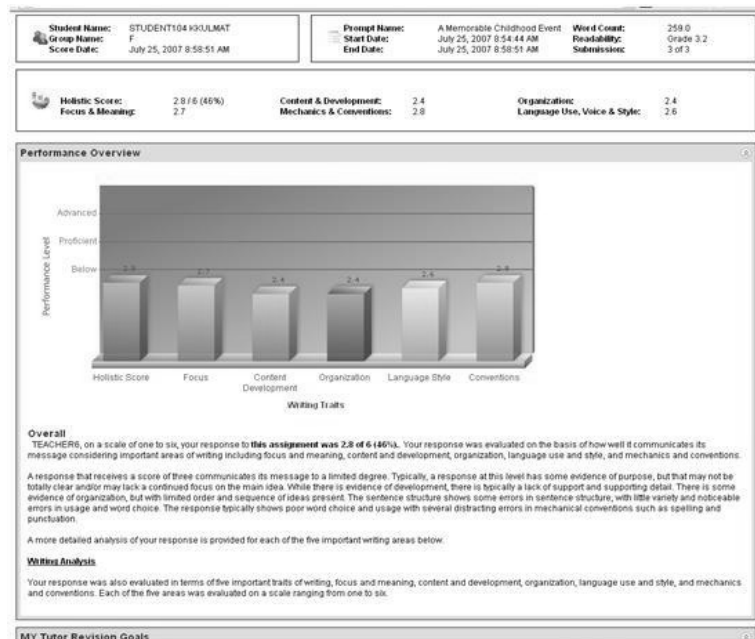
<An excerpt from Figure 3>

[Figure 3] View of My Editor Feedback

(2) My Tutor Feedback

On the My Tutor Feedback page, My Access provided a performance overview, overall comments on the essay, and revision goals for the five aspects on the analytic scoring scale. For a better idea of how My Tutor Feedback works, it maybe helpful to view a student's writing sample rated by My Access (Figure 4). In this example, the student received a 2.6 out of 6 for the holistic scoring, a 2.7 for focus and meaning, a 2.4 for content & development, a 2.4 for organization, a 2.6 for language use, voice, and style, and a 2.6 for mechanics and conventions. My Access then explained what these scores meant, as follows:

"... on a scale of one to six, the response to this assignment was 2.8 of 6 (46%)... A response that receives a score of three communicates its message to a limited degree. Typically, a response at this level has some evidence of purpose, but that may not be totally clear and/or may lack a continued focus on the main idea. While there is evidence of development, there is typically a lack of support and supporting detail. There is some evidence of organization, but with limited order and sequence of ideas present. The sentence structure shows some errors in sentence structure, with little variety and noticeable errors in usage and word choice. The response typically shows poor word choice and usage with several distracting errors in mechanical conventions such as spelling and punctuation." (excerpt from Figure 4.)



[Figure 4] View of My Tutor Feedback

After explaining the scores, My Access suggested revision goals in each domain. In this specific example, in the domain of Focus & Meaning, My Access first advised the student to understand the purpose, audience, and task. My Access also suggested narrowing the focus and using details to support the purpose of the day. Under each goal, My Access provided a guide for revision in greater detail, as is discussed for Examples 1 and 2.

The student was also advised with regard to Content and Development to create a detailed setting, develop realistic characters, and build a well-developed plot. The revision goals for Organization were "to put events in order and use transitions to show your reader how events and ideas are connected". In Language Use, Voice, & Style, My Access suggested that the student should choose words carefully, write short and long sentences, and make the beginnings of the sentences more interesting. Finally, in Mechanics & Conventions, revision goals included eliminating errors in grammar, mechanics, punctuation, and spelling.

For further analysis of how My Access provided feedback, it is necessary to look more closely at the specific guides under each revision goal in the following examples:

[Example 1]

Revision Goal 2: Narrow your focus.

1. Highlight the main event in yellow.
2. Underline events and details that are NOT about the main event.
3. Move or remove the events and details that are NOT about the main event.

Example:

Before Revision: I will always remember the day I sang a song in the talent show. Jake and I watched a scary movie that night.

Frank's Strategy: The topic of my story is the talent show; watching a scary movie is not. I need to delete details that are not related to the main event.

After Revision: I will always remember the day I sang a song in the talent show.

Frank's Reflection: I removed "Jake and I watched a scary movie that night" because it has nothing to do with the main event, the talent show.

[Example 2]

Revision Goal 2: Use transitions to show your reader how events and ideas are connected.

1. Highlight, in orange, the transitions (words such as first, second, last, next, yesterday, today, in addition) in your story.
2. Add more transitions to show how your events go together. Use the word bank to help you!

Example:

Before Revision: First, Sarah got out of bed. Then, she ate some cereal. She heard her bus coming down the street. She ran out the door.

Oscar's Strategy: I need to add transitional words or phrases to help my reader follow the events in my story.

After Revision: First, Sarah got out of bed. Then she ate some cereal. A few minutes later, she heard her bus coming down the street. She ran out the door.

Oscar's Reflection: I added the transitional phrase, "A few minutes later" to show when the events took place and to make the sequence of events flow together.

Example 1 is an excerpt from the revision goal in Focus & Meaning, and Example 2 is from Organization. Based on its analysis, My Access suggested only two or three revision

goals for each category, no matter how many errors students made in their essay. Consequently, it highlighted the errors regarded as the most prominent and urgent problems in the essay. For the essay shown in the examples, My Access suggested three revision goals in Focus & Meaning and two goals in Organization. Example 2 followed "Revision 1: Put your events in order" which prepared for Revision 2. The revision goals were usually interconnected, and were often sequential within a category.

Of particular interest on this page is the fact that My Access adopted five human names for each domain when suggesting revision goals: Frank for Focus & Meaning; Anne for Content & Development; Oscar for Organization; Larry for Language Use, Voice & Style; and Marcy for Mechanics & Conventions. The educational influence of this strategy is discussed in the following section.

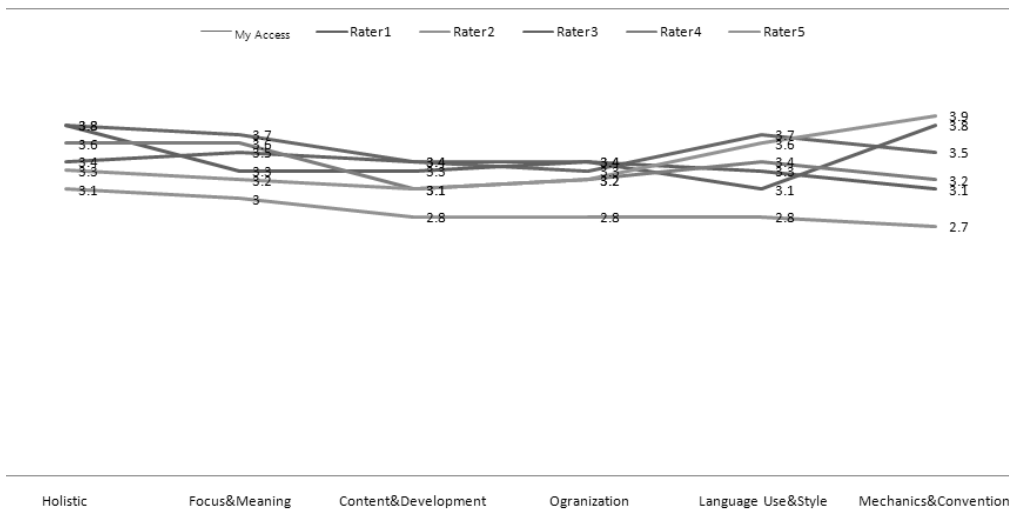
IV. FINDINGS AND DISCUSSION

1. My Access as an AES System

Before discussing the reliability of My Access as an AES system, it is useful to examine the validity of the grading criteria used on My Access and in this research. Based on literature and scoring rubrics developed previously by other scholars and institutes, the validity of the rubric was examined. The holistic rubric of My Access, like the TOEFL Test Integrated Writing (TIW) Rubrics (scoring standards), which is the most well-known example of a holistic scoring rubric, contains descriptors of the syntactic and rhetorical qualities of six levels of writing proficiency. The descriptors at each level of the rubrics of My Access are in fact almost identical to those of the TOEFL TIW rubrics; thus, this similarity can refute doubts cast on the validity of the holistic scoring rubrics of My Access.

The analytic scoring rubrics of My Access also turned out to be valid when compared to other widely accepted analytic writing rubrics, such as Jacob, Wormuth, Hartfel, & Hughey,(1981) scoring profile or the TEEP attribute writing scales (Weir, 1990). The analytic scoring scale of My Access has the same five aspects that Jacob et al. adopted. The descriptors of each score for each aspect also share many similarities between the two rubrics, with little variation. The only difference is that each aspect of My Access has the same weight, whereas the aspects of Jacob et al.'s rubric are differentially weighted depending on their significance in writing. A careful examination and comparison of the rubrics of My Access to Jacob et al.'s profile confirmed the validity of My Access.

The Pearson correlation coefficient implied that My Access achieved high reliability as an AES system. The inter-rater reliabilities between My Access and each human rater were high (all over 0.79, with the exception of Rater 2, whose inter-rater reliability was 0.45). The inter-rater reliabilities between Rater 2 and the other raters were comparatively low (two correlation coefficients out of four of under 0.5), while inter-rater reliabilities among the other raters proved to have very strong correlations (between 0.78 and 0.82).

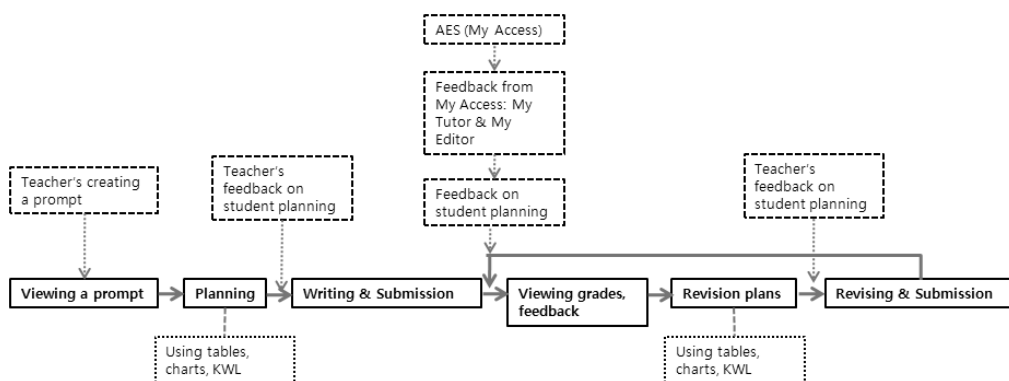


[Figure 5] Averages of Scores Rated by My Access and the Human Raters

The strong correlations between the AES system of My Access and at least four of the human raters offered evidence that My Access has a reliable AES program. The low inter-rater reliability with Rater 2 was, after careful examination of the scores, written feedback, and comments made by My Access and the other human raters and an interview with Rater 2, concluded to be ascribed to a lack of experience on the part of Rater 2. In fact, a number of articles maintain that AES systems are more reliable because they are more objective, strictly following the grading criteria. Conversely, human raters cannot be completely objective due to fatigue, errors, misunderstanding of the prompts or criteria, or different perspectives on writing. In addition, raters may fall back on their individual rating styles, even when they rate essays based on the criteria of a rubric (Wiegler, 2002). Therefore, if this research had been conducted with a large group of raters, Rater 2 would have been eliminated due to his deviation from a normal distribution. Human raters are not always reliable, and My Access, which turned out to be a reliable AES system, can be a useful and reliable tool that is complementary to writing teachers.

2. My Access as a Writing Instructional Program

This paper has examined the AES system of My Access to date. This section will discuss My Access as a writing instructional program, focusing specifically on feedback and revision guides. Obviously, the pedagogy underlying the instructional design and features of My Access is process-oriented writing. My Access guides student writers (and teachers) to focus on the processes of writing, not just outcomes or grades. As shown in Figure 6, My Access helps teachers create or select writing prompts from its archive. My Access then leads students to be engaged in planning before writing. My Access provided qualitative feedback and comments on student essays that were expected to be reflected in the students' revisions, as well as grades.



[Figure 6] The Entire Writing Process on My Access

This paper has already pointed out the importance of assessment in the cyclical process of writing. To gain a positive washback effect in writing, as noted earlier, it is essential that the assessment not only be reliable and valid, but also that feedback should be included in the assessment. My Access utilized both direct and indirect feedback on My Editor Feedback, focusing mostly on grammatical and mechanical errors in Mechanics and Language Use. As mentioned in the literature section, direct feedback refers to explicit corrections of students' errors, while indirect feedback leads students to think about their errors and revise them without explicit error corrections. My Editor Feedback first marked student errors by indicating the category of the errors (e.g., Archimedes <spelling errors>); only when students click on the feedback does it show students explicit correction of the errors (e.g., Archimedes => Archimedes). This is particularly helpful because it provides students with the opportunity to revise the errors themselves. Direct feedback can also support to EFL writing, as EFL students, especially those at this

level, are often not capable of making the right decisions when correcting their errors if they rely only on indirect feedback, as teachers intend.

In fact, some of the feedback on the My Editor page appeared to be beyond the current level of the student language proficiency. For instance, whereas 'spelling errors', 'punctuation errors', and 'missing articles' are easy to grasp, 'similar words', which suggest that the student use an alternative expression that works better than the present one, seemed beyond the student's current language proficiency. Knoblauch and Brannon (1981) summarized previous studies on the effects of teacher feedback as follows: "1) students often do not comprehend teacher responses to their writing; 2) even when they do, they do not always use those responses and may not know how to use them; 3) when they use them, they do not necessarily write more effectively as a result" (p.1). Their study supports the idea that since the language used on My Editor Feedback is not suitable for the English proficiency level of Korean middle school students, it may not be able to support student revisions and, even worse, that it may demotivate students with regard to English writing.

Furthermore, considering that making corrections on every error in the essay is not the most effective feedback strategy, it is important to prioritize the order of corrections and create a hierarchy of importance of issues for revisions. From this perspective, 'similar words' or 'missing articles' are not the most urgent revisions at this level of writing. Rather, some errors that My Editor overlooked should be prioritized over others. For example, awkward lexical usage or expressions (e.g., "knocking stones") and obvious grammatical errors, such as tense (e.g., "the reason we could lived in this modern world") as shown in the example in Figure 6, can hinder readers' understanding of the essay and decrease the quality of essay.

My Tutor Feedback, on the other hand, employs indirect feedback only. While My Editor Feedback intends to increase the accuracy of student essays, My Tutor Feedback tries to enhance writing fluency. Rather than focusing on a micro-level of writing as My Editor Feedback does, My Tutor Feedback emphasizes a more macro-level approach to writing revision strategies. My Access draws the students' attentions to writing strategies instead of simply correcting the errors. The examples that My Access used to demonstrate the revision goals were not excerpts from the student's essay, but rather examples drawn from My Access' archive. My Access does not use the student's essay for the examples to illustrate revision goals because this may be still technically too complicated and, therefore, would not be cost-effective. Moreover, showing similar examples rather than directly correcting the student's essays can be more effective in

advancing the student's writing skill, by allowing him/her to revise it by himself/herself.

Additionally, as discovered in previous studies, certain types of comments appeared to be more effective for student revisions than other types. As mentioned earlier in this paper, My Tutor Feedback used a first person point of view when responding to student essays. These human names functioned as writing teachers, or pedagogical agents, to make the feedback and the relationship between the student and the grader (My Access) more human. This can make the comments seem more individualized and friendly, so students can feel more comfortable, as if they were hearing the comments from their human teachers. Yet, while My Tutor Feedback was limited only to statements, human raters responded to the students' essays in more diverse ways, using statements, questions, praise, imperatives, and requests. Another difference between My Tutor Feedback and the human raters was that the human raters appeared to use both positive and negative comments, while positive comments or praise were not used by My Tutor Feedback. According to Straub (1996, 1997), students prefer positive comments and diverse modes of teacher feedback. When students appreciate the teacher's comments, they are more likely to use the feedback in their revisions.

Among the 34 students involved in this research, very few made revisions on the same prompt after being graded and receiving feedback from My Access. The researcher tracked those students' revisions to examine to what degree the feedback given by My Access was correctly or appropriately incorporated into their revisions. Most students corrected the errors that were directly marked by My Editor, primarily regional errors, but were not successful in global revision, such as 'narrowing your focus,' or 'considering your readers'.

Empirical studies on feedback demonstrate that how teacher comments are constructed is as important as on what teacher feedback is focused (Ferris, 2005). The studies also point out that teachers should utilize different feedback strategies depending on the student's level and stage of the writing process. As clearly shown in Examples 1 and 2, however, My Access failed to utilize diverse modes and strategies of feedback or adjust its language level to accommodate the students' levels. My Tutor Feedback was often too vague for the middle school EFL students to understand as the program was originally designed for native English-speaking students; furthermore, as on My Editor Feedback, the language used in the feedback did not match the language proficiency level of the middle school EFL students. It is also questionable how much the students can apply the writing strategies and examples provided by My Access when revising their essays or writing other essays. To discover this, the researcher investigated different essays

produced by the same students on a variety of prompts over the semester. There was not much improvement over the semester, at least in terms of the scores rated by My Access. This may be because the students simply did not acquire writing strategies or increase their writing proficiency through this program. It may also be because the students produced essays on different discourse modes and genres; hence, it was more difficult to transfer writing strategies to the other essays. It is hard to conclude what led to this result without a longitudinal study and student interviews.

V. CONCLUSION

My Access, anchored in process-oriented writing, offers many valuable features to promote student writing, including a potentially effective (and efficient) instructional writing program. However, neither teachers nor students fully utilized the program; the students did not use the pre-writing planning page or revision planning, nor did they submit revised versions after being graded and receiving comments. In addition, the teachers did not make any effort to make further comments on student essays that complemented the automated feedback. Although these features were supported by My Access, the teachers did not use them because they may not have realized the importance of process-oriented writing or did not have enough time to make extra comments on student essays. Since Korean middle school students do not have many opportunities to learn writing, and their English is not yet fluent enough to accomplish writing tasks without support, teacher feedback and comments on their essays would be useful and helpful for facilitating student writing. This research discovered that My Tutor Feedback was not successfully reflected in the students' revisions, presumably because the students could not fully understand what to do in their revisions. To maximize the effect of My Access and further support student writing, two things are recommended. First, writing teachers should provide extra feedback in accordance with the students' levels and give explanations of the feedback provided by My Access to help student understanding. Second, customizing the program or developing a new program would certainly be beneficial, particularly when accommodating Korean students' needs.

The results of this study confirm that the reliability of My Access as an AES system is high enough to be used in writing classrooms. This means that when effectively used, it can share the workload of writing teachers. This research also showed that although AES is more efficient, cost-effective, and, in a way, more objective as compared to human

raters, such as AES systems still cannot fully replace human teachers. Previous studies on the effects of teacher feedback on student writers find that students greatly appreciate and value teacher feedback and consider it extremely helpful for their writing development. Studies also show that student writers value emotional encouragement as well as criticism, feedback on language errors as well as comments on content, face-to-face conferences as well as written comments, and learner control as well as teacher guides. In EFL writing classrooms, teachers have the double burden of promoting student writing skills while enhancing language proficiency. Even though automated writing instruction programs are not yet perfect, if teachers can creatively and effectively use them, they can be a very useful tool for both teachers and students.

The main limitation of this study is that the program was evaluated only from the graders' viewpoint, not from the students'. To confirm long-term outcomes of My Access, a student-user analysis is needed. Classroom observations in a blended learning situation, where the teacher and students participate in both online and offline classes, would provide insight into the nature of automated writing instruction programs and their effects on student writing. To measure the maximum effect of My Access on student writing, it is recommended that a study be conducted in which the teacher and students utilize all the features of My Access.

REFERENCES

- Attali, Y., & Burnstein, J. (2006). Automated Essay Scoring with e-rater V.2. *The Journal of Technology, Learning, and Assessment*, 4(3), 3-30.
- Bachman, L., & Palmer, A. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Bakaoui, K. (2007). Rating scale impact on EFL essay marking: A mixed-method study. *Assessing Writing*, 12, 86-107.
- Bennet, R. (1999). Using new technology to improve assessment. *Educational Measurement: Issues and Practice*, 18(2), 5-12.
- Choi, S. H., Cho, D. W., Shin, D. I., & O, K. M. (2005). Validating scoring processes and rating criteria a college-level web-based writing assessment. *Multimedia-Assisted Language Learning*, 8(1), 34-52.
- Ferris, D. (2003). *Response to student writing: Implications for second language students*. Mahwah, NJ: Lawrence Erlbaum Associates.

- Ferris, D. (2005). *Teaching ESL composition: Purpose, process, and practice*. Mahwah, NJ: Lawrence Earlbaum Associates.
- Hamps-Lyons, L., & Henning, G. (1991). Communicative writing profiles: An investigation of the transferability of a multiple trait scoring instrument across ESL writing assessment contexts. *Language Learning, 41*, 337-373.
- Hamps-Lyons, L. (2003). Writing teachers as assessors of writing. In B. Kroll (Ed.), *Exploring the dynamics of second language writing* (pp. 162-189). Cambridge: Cambridge University Press.
- Henning, G. (1991). Validating an item bank in computer-assisted or computer-adaptive test. In P. Dunkel (Ed.), *Computer-assisted language learning and testing: Research issues and practice* (pp. 209-222). NY: Newbury House.
- Hout, B. (1996). Computers and Assessment: Understanding two technologies. *Computers and Composition, 13*, 231-243.
- Huff, K., & Sireci, S. (2001). Validity issues in computer-based testing. *Educational Measurement: Issues and Practice, 20*(3), 16-25.
- Hyland, K. & Hyland, F. (2006). *Feedback in second language writing: Content and issues*. Cambridge: Cambridge University Press.
- Jacobs, H., Zingraf, S. Wormuth, D., Hartfel, V., & Hughey, J. (1981). *Testing ESL composition: A practical approach*. Rowley, MA: Newbury House.
- Knoblauch, C. H., & Brannon, L. (1981). Teacher commentary on student writing: The state of the art. *Freshman English News, 10*, 1-4.
- Penny, J., Johnson, R., & Gordon, B. (2001). The effect of rating augmentation on inter-rater reliability: An empirical study of a holistic rubric. *Assessing Writing, 7*, 143-164.
- Powers, D. Burnstein, J. Chodorow, M. Fowles, M., & Kukich, K. (2002). Stumping e-rater: Challenging the validity of automated essay scoring. *Computers in Human Behavior, 18*, 103-134.
- Rudner, L., & Gagne, P. (2001). An overview of three approaches to scoring written essays by computer. *Practical Assessment: Research & Evaluation, 7*. Retrieved March, 2008, from the World Wide Web: <http://pareonline.net/getvn.asp?v=7&n=26>.
- Straub, R. (1996). The concept of control in teacher response: Defining the varieties of directive and facilitative commentary. *College Composition and Communication, 47*, 223-251.
- Straub, R. (1997). Students' reactions to teacher comments: An exploratory study. *Research in the Teaching of English, 31*, 91-119.
- Weir, C. J. (1990). *Communicative language testing*. Mahwah, NJ: Prentice Hall Regents.

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

Keywords: web-based writing program, EFL, Automated Essay Scoring, reliability

Application Levels: elementary and secondary education

Author: Lee, Sangmin (Kyung Hee University); sangminlee@khu.ac.kr

Received: May 15, 2008

Reviewed: July 30, 2008