8

E-WRITE AS A MEANS FOR PLACEMENT INTO THREE COMPOSITION COURSES

A Pilot Study

Richard N. Matzen Jr. and Colleen Sorensen

In the fall of 2002 Utah Valley State College (UVSC) began institutional research into placement tests for first-year composition courses: two basic writing courses and a freshman composition course. UVSC researchers had previously presented evidence in the article "Basic Writing Placement with Holistically Scored Essays: Research Evidence" (Matzen and Hoyt 2004) that suggested the college's multiple-choice tests—ACT, ACT COMPASS, and DRP (Degrees of Reading Power)—often misplaced UVSC students in composition courses. As an alternative to placement by these tests, a research team including people from the Department of Basic Composition, the Institutional Research and Management Studies Office, and Testing Services suggested to colleagues and administrators that placement might be more accurately accomplished by timed-essay scores alone or by combining timed-essay scores with reading test scores.

This chapter represents an extension of previous research and is a description of a pilot study regarding the following questions that are relevant to a WPA: will ACT e-Write, an automated essay-scoring program, accurately score UVSC students' timed essays? Will e-Write be a practical technology when used? Will e-Write scores be reported in a timely manner?

PLACEMENT AT UVSC: AN ISSUE OF FAIRNESS AND RETENTION

We saw that a bridge was needed between the current placement system and a possible new placement system.

- Current placement system: placement by composite ACT scores or combined scores of the reading (ACT COMPASS) and editing (DRP) tests.
- Possible new placement system: placement by timed-essay scores alone or by combining timed-essay scores with reading test scores.

Fairness was a significant issue when considering changing the placement system. On the one hand, we thought that the current placement system was unfair to students who were misplaced, but on the other hand, we did not want to replace that system with another system containing unfairness, too. Consequently, the pilot study was not only a dress rehearsal for a new placement system but also a test of e-Write's validity and reliability. In any case, the students in the pilot study had already been placed into composition courses by the current placement system. The bridge to a new placement system was still being formed.

Fairness is a larger issue, too, because of the number of students affected by the current placement system. Each fall semester, the system—based on multiple-choice tests that are not direct measures of writing—affects over seven hundred students who enroll in a basic writing course (ENGH 0890 or ENGH 0990) and affects over nineteen hundred students who enroll in the freshman composition course (ENGL 1010). According to one study, the editing and reading tests at UVSC may accurately place about 62 percent to 65 percent of students into ENGH 0990, while accurate placement into ENGH 0890 is worse (Matzen and Hoyt 2004, 4, 6). At UVSC, consequently, significant numbers of students would be more accurately placed into composition courses if the placement system included timed-essay scores. More UVSC students, in other words, would enroll in and pay for a composition course more aligned with their writing skills as they exist at the beginning of a semester. Another part of defining fairness in this context is that the two basic writing courses do not bear academic credit that can be transferred or applied toward earning a degree, whereas the freshman composition course bears transferable credit as well as credit applicable toward a UVSC degree.

Besides fairness, placement tests for composition courses have implications for retention at UVSC, where retention is a significant problem. Statements such as the following are not unusual to find in institutional research reports about retention at UVSC: "In general, the college has very low retention rates for students. Over half the students drop out of college failing to earn a degree or transfer. The college generally loses 30% to 35% of its students from fall to spring and nearly 60% of its students by the following fall" (Hoyt 1998, 4). Meanwhile, accurate placement into basic composition courses has been linked to improved student retention (Baker and Jolly 1999; Boylan 1999; Cunningham 1983; Glau 1996; Kiefer 1983; McGregor and Attinasi 1996; White 1995). In short, placement by timed essay might improve retention and fairness for each freshman cohort at UVSC.

AN OVERVIEW OF THE PILOT STUDY INTO ACT'S E-WRITE PROGRAM

In the spring of 2003, we wrote and submitted a grant to UVSC Foundation's Exceptional Merit Grants Program. The grant, entitled "ENGH 0890 and 0990 Placement," was approved, funding the e-Write pilot study. We hoped that the grant and e-Write would be the first steps toward having a cost-effective, accurate placement system based on timed-essay scores for three of the courses in the composition sequence: ENGH 0890, ENGH 0990, and ENGL 1010. We had read the ACT marketing pamphlet that announced, *COMPASS e-Write Direct Writing Assessment from ACT*. The pamphlet's first line was this: "COMPASS e-Write™, ACT's exciting new direct writing assessment, can evaluate a student writing sample, score it as reliably as two trained raters, and cost-effectively deliver a score report in seconds" (ACT 2001). This promotional material led us to believe that the program might have promise for UVSC.

Grant funds were spent in two ways. First, prior to the beginning of the fall semester of 2003, Testing Services at UVSC bought three hundred e-Write units (or tests) and installed the e-Write program on thirty computers in a classroom in the testing center. According the all reports, ACT's directions for the installation were thorough and easy to follow.

Thereafter, UVSC students completed the e-Write tests during the first week of classes that fall semester. However, because completion of the tests was like a dress rehearsal for a new placement system, we explained to teachers and students another, more immediate benefit of their participation in the pilot study. Students heard, as the proctor's instructions were read to them, that their teacher would read the e-Write essays to determine the rhetorical and grammatical needs or strengths of individual writers and the entire class. In other words, teachers would read the e-Write essays as diagnostic writing samples. As such, the essays might trigger more or less grammar instruction during the semester or identify students who might be particularly weak or strong writers. In total, approximately three hundred students completed e-Write tests, approximately one hundred from each of the three courses: ENGH 0890, ENGH 0990, and ENGL 1010.

Besides buying e-Write units, grant funds paid human raters for scoring the e-Write essays. The essays were written in response to two of eight possible e-Write prompts. Only two prompts were used because we doubted that an acceptable level of interrater reliability would be achieved in this sampling of three hundred e-Write essays if there were

more than two essay topics. We had previously organized successful norming sessions in which two essay topics defined the anchor and range-finder essays. At the same time, we hoped that better-written timed e-Write essays might result if students were allowed to choose between two prompts.

Interrater reliability was important because a part of the research design was to test another claim in the ACT pamphlet: "A key factor, of course, is how the results of electronic scoring compare to those of trained raters. Our research shows very strong agreement: 100 percent of COMPASS e-Write scores are within one point of each other; approximately 76 percent of scores match exactly" (2001). We understood this statement as a minimum and reliable claim of accuracy and assumed that ACT would not exaggerate its claims, knowing that e-Write administrators might independently test e-Write's accuracy. Moreover, the ACT statement suggested to us that no splits existed when e-Write rated timed essays.

THE DIFFICULTIES OF RECEIVING AND RETRIEVING E-WRITE SCORES AND ESSAYS

Typically, a UVSC student finished his or her e-Write test and saw his or her score on the computer screen while the score sheet was also printing at the testing center's printer. But, immediately after some students completed their e-Write tests, we noticed that some scores were missing. Students with missing scores received this e-Write message on their computer screens: "The response was judged to be unscoreable (e.g., blank response, illegible response, or a response written in a language other than English)." This message confused us because we had watched the students who received it type e-Write essays. In addition, none of us were sure that these *e*-Write essays were preserved in the database.

On the third day of testing, another negative unanticipated event occurred. The first class arrived at the testing center as scheduled at 8:00 a.m., received the proctor's instructions, and began typing timed essays in e-Write. As students finished and submitted their essays, the normal sequence of e-Write messages or screens "froze." Similarly, if a student was word processing, when he or she clicked the "save" button, the computer immediately froze, without any indication whether the document had been saved or not. As each student in the 8:00 a.m. class submitted his or her e-Write essay, a similar event occurred. We directed the students not to give their computers any more commands and called ACT Technical Support for help. The ACT personnel told us that the server, where all the e-Write tests were sent, was down. An ACT

Technical Support person also said that this was the first time such an event had occurred.

We wondered what to do next. ACT Technical Support personnel would not indicate when the e-Write server would be restored and would not answer these questions: were the class's e-Write tests received by the server? If the testing center computers were shut down and restarted, would that mean that the e-Write tests would be permanently lost? At that time, too, the larger problem was whether to administer e-Write tests to the 9:00 and 10:00 classes. We decided to have students in the 8:00 class finish their essays and submit them—only to have the computers, predictably, freeze. After all the e-Write tests were finished and the 8:00 class had left, we shut down and restarted the frozen computers. For the 9:00 and 10:00 classes, we decided to provide students with a paper copy of the two e-Write prompts and to have the students type their timed essays as Word documents to be later typed into the e-Write program by staff. The 9:00 and 10:00 classes received the same proctor's instructions as the previous class but, unlike that class, some students used spell- and grammar-check programs when writing their timed essays.

In general, the freezing incident added to the problem of missing scores and complicated the pilot study. The central problem of the missing scores was that if placement had actually depended on e-Write scores, a significant number of students would not have been able to register for their first composition course at college. Each student would have had to wait for his or her ACT score report and then register for a composition course the following semester.

ACT personnel eventually provided us with explanations as to why some e-Write scores were missing. Apparently, e-Write does not score one of every fifteen essays automatically; some selected e-Write essays are sent to ACT human raters for scoring as a "quality-control measure." Also, according to ACT personnel, some e-Write essays are too short for scoring except by ACT human raters.

Regarding our missing e-Write scores, however, communications with ACT were problematic, and their communications to us were not always received when needed. For example, at least several weeks after the e-Write testing, ACT Technical Support personnel informed us that we had been receiving encoded e-mails regarding the missing scores and missing e-Write essays. Testing center staff did not notice these e-mails because they were not expected and were sent to a generic e-mail address for the testing center. Learning of this, we asked ACT Technical Support personnel to resend them. In response, we were told to look for the encoded e-mails again. When we failed to find them, ACT Technical

Support personnel agreed to resend them. Later we learned that the ACT instructions for decoding the e-mails were incomplete.

By January of 2004, about four months after the e-Write essays had been written, communications between ACT Technical Support personnel and the testing center's technical staff had still failed to resolve the issue of missing test scores and missing e-Write tests. This caused us to contact a representative in the research component of ACT Placement Programs. As a result, by mid-February, or approximately six months after the testing, we had finally received most of the missing e-Write scores. Until then, approximately 17 percent of e-Write scores had been missing. If the college had actually been using e-Write for placement purposes in the fall of 2003, would that have meant that an alternative placement assessment would have had to be devised for at least 17 percent of students?

Other aspects of the situation were relevant when considering the desirability of using e-Write for placement purposes. The exact number of missing scores was subject to interpretation, because during the e-Write testing, some students typed two essays, one or both having a missing score. Sometimes, but not always, students with missing scores experienced this sequence of events: first, they typed their e-Write essays and submitted them for scoring but then received the message, "The response was judged to be unscoreable. . . ." Responding to that message, some students reentered e-Write and typed a second essay which, like the first, may or may not have been scored. After this situation had occurred about twenty times, we realized what was happening and modified the proctor's instructions to lessen the occurrence of the problem. Regarding other practical glitches, we found instances of a student submitting two e-Write essays or exiting the program without finishing one essay. This meant that the testing center paid for more e-Write units than planned. Incidentally, early in 2004, we received the frozen e-Write essays and scores, meaning that the ACT server had received them. Reception was delayed because of encrypted files as well as ACT Technical Support not knowing where the scores and essays were for a time.

ENGLISH TEACHERS SCORE THE E-WRITE ESSAYS

Although paper copies of e-Write essays were normally not available to the testing center administrator, much less an e-Write user, prior to the testing sessions, we received information from ACT Technical Support personnel about how to obtain paper copies of the e-Write essays. With those in hand, we worked with two other English professors to identify anchor and range-finder e-Write essays. In December of 2003, we led a norming session for nine English teachers or raters to ensure the

reliable application of an adapted ACT 8-point scale to rate e-Write essays. To make that scale more meaningful to raters, we also suggested that a score below 4 placed students into ENGH 0890, below 6 into ENGH 0990, and below 8 into ENGL 1010. A score of 8 would exempt students from an entry-level English course. The raters had taught all of these courses and established their interrater reliability as 83 percent, which meant that the two raters' scores agreed for 83 percent of the e-Write essays. A third reader read the 17 percent of e-Write essays that received split scores and assigned a final score to them.

Besides the adapted ACT 8-point scale, shared curricular knowledge helped the raters score the e-Write essays. The ENGH 0890 curriculum is designed for basic writers whose writing suggests written-down speech, the lack of a reading history, or significant problems with common orthographic conventions or with controlling sentences and writing paragraphs. The ENGH 0990 curriculum, which serves about 75 percent of UVSC basic writers, is designed for students who are ready to read academic texts, possess a general knowledge of an essay, and control most of their sentences and paragraphs in terms of rhetorical and mechanical structures. ENGL 1010, Freshman Composition, is designed for high school graduates who are ready to begin writing at a college level.

COMPARING MULTIPLE-CHOICE TEST SCORES, E-WRITE SCORES, AND HUMAN RATERS' SCORES

The office of Institutional Research and Management Studies analyzed multiple-choice test scores, e-Write scores, and human rater scores. Table 1 (next page) suggests that the rater scores have a moderately strong correlation with the student scores on multiple-choice tests.

In contrast, table 2 (next page) suggests that e-Write scores have a much weaker correlation with the same multiple-choice test scores.

In other words, e-Write's validity is weak in terms of its correlations with multiple-choices tests cited in table 2. The correlation between e-Write scores and raters' scores, moreover, is .56, which is below expectations created by ACT. These results mean, first, that the rater scores have greater criterion-related validity, and second, the e-Write scores would have resulted in a substantial misplacement of students.

PLACEMENT AND E-WRITE AT UVSC

Based on this limited experience with e-Write, the Department of Basic Composition and the Department of English and Literature foresee two significant difficulties with e-Write. First, e-Write scores are *not* received

TABLE 1

Correlations between human rater scores and other test scores

Correlation	*Test*	*Correlation*	*Test*
.430	DRP (Degrees of Reading Power)	.431	ACT COMPASS (editing test)
.559	ACT English	.421	ACT Reading
.512	ACT Composite		

TABLE 2

Correlations between e-Write scores and other test scores

Correlation	*Test*	*Correlation*	*Test*
.180	DRP (Degrees of Reading Power)	.267	ACT COMPASS (editing test)
.290	ACT English	.192	ACT Reading
.209	ACT Composite		

in a timely fashion and second, the validity of the e-Write scores is questionable. If the e-Write scores had been used for placement purposes, for example, apparently only 4 of 298 students would have enrolled in the lower-level basic writing course, an outcome that experienced basic writing teachers at UVSC believe is inaccurate.

That said, the e-Write study has had some positive outcomes. Whereas the publication of "Basic Writing Placement with Holistically Scored Essays: Research Evidence" (Matzen and Hoyt 2004) seemed to suggest that only the Department of Basic Composition was concerned about accurate placement, the Exceptional Merit Grant that funded the e-Write pilot study signaled that an active concern for accurate placement is shared by the Department of English and Literature. English faculty in both English departments have agreed to trust human raters and to advocate that timed essays become a part of a placement system for first-year composition courses.