

Hands-on Writing Assessment in an Undergraduate Cohort

Jimmy R. Smith
Interim Director of Assessment/Professor of Mathematics

Emory V. Maiden
Professor of English

Heather H. Langdon
Assessment Coordinator

Katherine A. Truett
Graduate Assistant/Adjunct Instructor

**Institutional Research, Assessment & Planning
Appalachian State University**

Introduction

Appalachian State University is a comprehensive public university (Master's I) that offers degrees at the bachelors, masters, specialist, and doctorate levels. The campus is largely residential, and the undergraduate student population is mainly of traditional college age. Appalachian has been a member of the University of North Carolina System since 1967, and has regularly been recognized by *US News and World Report* as one of the outstanding comprehensive universities of the Southeast.

Since 1994, Appalachian has been conducting cohort-based assessment testing of students. Initially, this testing consisted solely of opinion surveys, whose administration was tied to important events in student's academic lives, such as freshman orientation, sophomore advising, and senior graduation registration. Addition of cognitive testing to Appalachian's assessment efforts began with the 1997 freshman cohort. In Fall of 1998, Appalachian began Focus Day, one day during each semester when all regularly scheduled undergraduate courses are suspended for assessment testing and other events.

Rather than attempting to assess every cohort of students in every cognitive domain, Appalachian's approach has been to select certain cognitive skills to be tested in each cohort, and then perform follow up testing of the same cognitive skill area with the same cohort. This makes it possible to perform before and after measures on assessment data for a given cohort, and also keeps the total testing time for each student during Focus Day under three hours. Examples of skill areas which were tested in the 1997 cohort are critical thinking and math/science reasoning.

Appalachian instituted a new core curriculum program in 1995 that added "designator" course requirements to the standard required courses in mathematics, English, social science, science, history, and the humanities. Under these new requirements, a student must take six courses designated writing intensive ("W"), four designated multicultural ("MC"), and a given number of other courses that are designated numerical data, cross-disciplinary, computing, and/or speaking intensive. Evaluating the way in which completion of designator courses impacts student progress in targeted skill areas is the major focus of Appalachian's general education assessment efforts. This report describes a large-scale

assessment of student writing skills of a cohort of students, both as entering freshmen and as sophomores.

Methods

All entering freshmen were administered a hands-on essay writing test during their attendance at summer orientation in July 1998. Students were given a choice of prompts, in most cases two, each of which asked them to take a stand on a controversial issue and write an essay in support of their position. Prompts were designed to be on general enough topics that students could address them without extensive prior knowledge or research. Students were given an hour to complete this task. The students had some motivation to perform well, since this test was used as a placement exam. Students whose performance was judged to be below the acceptable standard for college writing would be required to take developmental English; those whose writing was judged exceptional could get credit for a required English composition course.

The same cohort of students was tested again on the same task in January of 2000, during Focus Day and two subsequent testing makeup sessions. Students are unable to register at the normal time for their classes for the following Fall semester if they do not participate in assessment, which results in a better than 90% turnout for Focus Day testing and makeup days. Although students are compelled to come, they are aware that the test results will have no immediate foreseeable consequences for them. Thus, their level of motivation and engagement is often not optimal. This attitude was further exacerbated by extreme weather conditions on the day of testing, and test administrators observed that many students complained or were belligerent.

A team of faculty members from the English department scored the essays. Scorers received release time from their regular duties so that they could devote considerable effort to the task of essay scoring. All members of the scoring team had previous experience using the scoring rubric employed. With this rubric, essays are scored holistically on a scale of one to six. Each essay is graded by two faculty members independently, and the final score for each paper is the sum of the two scores given. The issue of reliability of scoring between the initial 1998 testing session and the follow-up test in 2000 was examined by having student workers hand copy 60 essays from the original 1998 assessment, which were resubmitted for blind scoring among the essays from the 2000 assessment.

Only students who took both the pretest and the post test were included in the final analysis. On one day of makeup testing on a Saturday morning, test administrators reported an unusually large number of discipline problems, which appeared to be alcohol-related. The average score for students who tested on this day was significantly lower than for students testing on Focus day (Kruskal-Wallis, $p < 0.000$; Mann-Whitney U as post-hoc, $p < 0.000$; with Bonnferroni correction); the average score for students testing on other makeup days was not significantly different from the average for Focus Day participants. Therefore, students testing on the problematic makeup day ($N = 118$) were not included in the final data set. Additionally, 19 cases were excluded because they wrote essays for the 2000 assessment that were so brief or entirely off-topic that no score could be assigned. The final data set therefore contained 1096 students, all of whom had taken both the 1998 and the 2000 test, and had submitted scoreable essays. Data management and statistical tests were performed using SPSS 10.0.

Results

Reliability of test scoring procedure

The initial scores of the 60 essays from 1998 that were submitted for rescoring in 2000 were compared with the scores they received in 2000. The average score of these tests went down slightly from 6.87 to 6.53, a difference that was not significant. The correlation between initial and second score was 0.469. On examination, it appeared that 35% of the papers received the same score both times, and another 38% differed by only one point. No papers went up more than two points, and only 5% of the rescored papers went down more than 2 points. The majority of papers that went down

more than two points (i.e., each grader graded it one point lower) had been high-scoring papers initially. This suggests that in 2000 graders may have had slightly more stringent criteria of what constituted an exceptional essay, but they tended to be consistent in their evaluation of average essays. We also looked at these papers to see if there was any consistency in lack of agreement between individual scorers (i.e., one scorer giving consistently higher or lower scores than others), but could find no evidence of a pattern of this sort.

Correlation of Fall 1998 Writing Sample Score with SAT Verbal Score

There was a significant correlation between Fall 1998 Writing Sample Score and SAT Verbal score (Spearman, $r = 0.301$, $p < 0.03$ with Bonnferroni correction).

Change in Average Score

Mean score increased significantly from 1998 to 2000 (Wilcoxon T, $p < 0.000$; $x = 6.63 \pm 1.38$ and 6.88 ± 1.40 , respectively).

Effect of number of Writing Intensive Courses on Scores

An initial observation was that the change in score pre and post was to some degree predicted by the first score, i.e., low scoring students tended to go up and high scoring students tended to go down (Spearman, $r = -0.562$, $p < 0.03$ with Bonnferroni correction). To adjust for this, student scores were expressed as the unstandardized residual of the difference between post and pre test scores adjusted for the variable "Fall 1998 score."

Number of writing intensive courses taken beyond freshman level courses was then found to have a small but significant effect on residual score (Kruskal-Wallis, $p = 0.016$ with Bonnferroni correction). Most dramatically, it was clear from looking at the means of residual scores grouped by number of "W" courses taken that a shift in student performance occurred at around five courses. The 12% of students who had taken 5 or more writing intensive courses, thus effectively completing or nearly completing their "W" designator requirements, had a mean increase in residual difference of 0.38 points, versus a slight overall mean decrease of -0.05 for students who took less than 5 "W" courses (Mann-Whitney U, $p = 0.002$ with Bonnferroni correction).

Conclusion

A slight but significant increase in student test score mean was observed between student's freshman and sophomore years. This increase was observed despite the fact that our best evidence suggests that scorers were grading slightly harder and students' behavior demonstrated that they were less motivated on the second test.

Furthermore, number of writing intensive courses taken had a small but significant effect on the degree of improvement in student writing. This effect was most dramatically noticeable where students had taken 5 or more courses. The scores of students who had completed or almost completed the "W" designator requirements for their degree went up, on average, over a third of a point, while those who had taken fewer "W" courses showed virtually no improvement (in fact a slight decrease) in writing skills on this test. This finding supports the present general education requirement of at least 6 writing intensive courses for all undergraduate students.

Paper presented at the Consortium for Assessment & Planning Support Conference, April 2001

[Link to Power Point presentation](#)