
WHAT ASSESSMENT REVEALS ABOUT ASSESSMENT

Suellynn Duffey, The Ohio State University

At Ohio State, we are studying the limitations of holistic scoring as a tool for measuring growth in literacy. I described my efforts to measure the effects of a curricular change in Ohio State's basic writing program. Recently, a new sequence of courses was added as part of a retention program, a sequence based on the model in David Bartholomae and Anthony Petrosky's *Facts, Artifacts and Counterfacts: Theory and Method for a Reading and Writing Course*; the courses aim to teach advanced literacy by integrating reading, writing, and thinking on a single topic. Because of the visibility of the retention efforts, begun in conjunction with the university's switch to a selective admissions policy, there was pressure to account for results.

One way in which we attempted to measure the effects of the new courses was by comparing the writing of students in them to a comparable group, students who had the same English placement level but who enrolled in the typical course sequence, not the new one. Both groups wrote essays on a test prompt designed to elicit the complex thinking and literacy that the new courses attempted to teach. The prompt asked students to analyze a change or invention that had unforeseen consequences. The essays were scored holistically using a six-point scale. Graduate students with prior scoring experience who were trained and calibrated served as readers. Even though this assessment was closely linked to instruction, the project was not able to demonstrate the literacy learning that was evident under the classroom conditions. The scores of both groups were roughly the same, even though the classroom performance of the groups was quite different.

Furthermore, the scoring sessions presented special problems for the readers because the essays of good students—ones who had learned to marshal evidence, interpret it, reach conclusions, and focus the movement of an essay upon a single and complex point—were still unconventional because the students did not have complete mastery of the forms of academic discourse. It

was not simply that surface errors abounded, although sometimes that was the case; but more importantly, the arrangement of ideas and the signals of coherence were unconventional and subverted the readers' expectations.

Consequently, readers found the essays very difficult to score, as a follow-up study indicated. For example, the readers indicated they would have changed one-third of the scores they originally gave, a sign of weak or inexperienced readers; but they were neither.