High Standards in Writing: The Sagamore Workshop

Our nation's present debate about public education appears to have shifted from questions about goals to questions about implementation. Everyone seems to agree that we need to shift our educational system from an emphasis on basic skills as minimum literacy to higher order skills and maximum literacy for all students. In this context, everyone agrees that all schools need more writing, more inferential problem solving, more interpretive reading. The question is how to achieve and especially how to assess this new standard.

The assessment of this new standard seems to require that we reduce our reliance on machine-scored tests and increase our reliance on the three Ps — performance (writing, problem-solving in math), portfolios (collections of student work over time), and projects (often interdisciplinary and applied, and sometimes collaborative). Although these new assessments have more validity for a thinking curriculum than those that dominate the current practice, they may raise questions about the reliability of scoring and the generalizability and comparability of scores.

Is it possible to achieve reliability in scoring and at the same time have an assessment system that allows for some local and regional variation? And what happens when teachers from one state score papers from students in another state? These were the key questions that the New Standards Project decided to explore last summer with the collaboration of the National Council of Teachers of English.

What is the New Standards Project? It is a partnership of seventeen states and six school districts that have been working together to create a national system of examinations, with local variation, based on a shared set of high standards for all students. The New Standards Project and its partners have been particularly interested in the field of writing because of its place in the school curriculum and its essential role in the literacy of our society. Writing is also the most developed field of performance assessment. Many classroom teachers have had experience developing writing prompts and exploring writing portfolios. Some excellent work has also been done in the scoring of student writing.

The Project's goal of allowing many different examinations to be geared to a single set of shared standards required a pilot exercise in the comparability of state writing assessment standards. In July 1991, we held our first Writing Calibration Workshop at the Sagamore on Lake George, in New York. Ten of the New Standards Project partner states participated in the five-day workshop. The focus was on the comparability of judgments by trained writing teachers from different states on the quality of student writing produced in different state writing assignments. Arizona, California, Colorado, Connecticut, Maine, New York, Oregon, South Carolina, Texas, and Vermont all sent six to ten experienced teachers trained as scorers. Each of these states already had two or more grade-level assessments for elementary, middle, and high school students.

As part of the calibration experiment, participants assigned scores to papers and discussed the meaning of the scores that they had assigned. During the scoring exercises, delegates from one state rated exam papers from other states using their own state's scoring rubric. The goal was to determine the statistical correlation between scores given by raters in one state and scores given by raters from another on the same piece of student work. A high correlation would indicate agreement on appropriate performance standards for judging student writing.

Delegates participated in three scoring sessions, each approximately three hours in length. At the end of these sessions, each student writing sample had been given seven ratings — the original in-state score, and two from each of three other states. A statistical analysis of
the data was performed by Robert Linn, director of the Center for Research Evaluation, Standards, and Student Testing, Carmen Chapman of the Illinois Department of Education, and Paul LeMahieu of the Pittsburgh Public Schools. Preliminary statistical analysis, corroborated by the participants’ own reports, suggests that calibrating several exams to a set of shared standards is indeed feasible.

Participants’ comments on the similarities and differences between the writing assessments scored may be of particular interest. These comments were made in discussions held by state delegations and cross-state groups. In both settings, delegates responded to questions about their own assessments and those that they became familiar with through the scoring exercises.

Delegates were asked whether, in their perception, states appeared to have shared standards for writing. A middle school teacher from Maine wrote: “I worked with writing samples from Connecticut, Texas, and South Carolina. My sense is that we are very close indeed!” A delegate from Texas said: “What is quite evident is that the states represented here seem to be very close in how they assess the attributes of good writing.”

One of the questions that delegates were asked was what their own state valued most in student writing. Delegates from the same state echoed one another when describing these values. Delegates from Colorado, New York, and Texas, for example, all stressed organization, mechanics, and the ability to stay on topic. At the same time, California, Maine, and Vermont privileged creativity, voice, risk-taking, and student ownership.

What is equally interesting to us, however, are participant comments on similarities and differences between the writing assessments scored. Some of the states seemed good companions if not soul-mates. Others stood apart. An Oregon delegate said: “Oregon’s rubric and expectations seem to align most closely with Vermont and Maine. We seemed most different from Texas, South Carolina, and Connecticut in our expectations and purposes.” From the notes of a facilitator leading a cross-state group: “A few states, Vermont and New York, seem to be further from the rest, but most of us have common ground.”

Different assessment purposes often meant different standards for writing. As expressed by a delegate from Maine: “There seem to be two camps — the exit assessments which say ‘you have to do well enough on this to meet our clear criteria’ and those that say ‘show us your best.’” An Arizona delegate argued: “Standards are probably closer than we realize in that we all seem to value the same qualities of good writing. What seems to vary is the purpose and setting for the evaluations. Some seem to have ended the experience check-up mentality while others have the attitude that students need to (and want to) be able to demonstrate what they can do and are like as learners.”

What made possible this convergence of judgment among teachers of writing? Three factors stand out: the harvest of extensive professional work on writing assessment over the last decade, an abundance of experience with large-scale assessment, and an emphasis in classroom instruction on the same kinds of work demanded in performance assessments.

This workshop experiment shows that we have enough agreement among teachers of English and English language arts to consider developing a system of common standards that does not depend upon a single examination. Such a system seems essential if we are to develop appropriate assessments that reflect our goal of maximum literacy for all.

This workshop experiment shows that we have enough agreement among teachers of English ... to consider developing a system of common standards that does not depend upon a single examination.

It is important to note, however, that writing assessment has not solved all of its problems. We still do not have agreement on how to respond to topic-to-topic variation in writing by the same students. Our writing assessments tend to treat writing as a generic activity and yet students can score at significantly different levels on different kinds of writing assignments. Similarly, while common standards of judgment among trained scorers can produce a high reliability in the scoring process, this does not mean that there is a common standard for the content of a writing assessment. Not all the participating states agreed on what should be assessed even when they agreed on how to judge individual pieces of student work. We have reason to hope for progress, but much work remains yet to be done.

Daniel P. Resnick is director of assessment development for the New Standards Project. Miles Myers is executive director of the National Council of Teachers of English.