

## MEASURING IMPROVEMENT IN WRITING: SOME CAUTIONS...

*Editor's Note:* This essay was abstracted from "Neglected Areas in Evaluating Writing Performance," a paper presented at the AERA annual meeting, San Francisco, 1981. Fiche available from ERIC Clearinghouse, 1111 Kenyon Road, Urbana, Ill., 61801 (ERIC ED 202888).

The single-sample, holistically scored, impromptu writing assignment, developed by ETS for college entrance tests and widely used in local school assessments,<sup>1</sup> serves well for screening, placement, or credentialing. But as researchers and evaluators, we generally make three mistakes in adopting this measure to assess program effects:

- We fail to view writing as a multiple construct.
- We fail to treat writing as a process.
- We fail to extract information about development, especially when progress is accompanied by flaws in the writing.

As a consequence, many annual assessments and pre-post studies produce disappointing or inconclusive results, obscuring real gains brought about by instruction.

### Writing as a Multiple Construct

Since Aristotle, discourse theorists have argued that different kinds of writing demand different composing strategies, beyond the basic literacy that underlies all writing.<sup>2</sup> Hence, some test questions, like "Write about an experience from which you learned a lesson," elicit narrative strategies, while others, like "Tell about a favorite person," require descriptive strategies. These tasks are excellent measures of general fluency, use of details, personal voice. But I have seen a tenth grader's holistic score drop from her ninth grade performance because she attempted to apply newly learned expository strategies to a narrative task. To uncover her gains in the sophomore year, a different kind of prompt was needed, such as "Not all inventions have been good for humanity. Write about one invention we might be better off without, and explain why."<sup>3</sup>

When a writing program has narrow goals, the task can be closely tailored to test the kind of writing being taught. For school-wide assessments, however, adequate evaluation of program effects requires several kinds of writing tasks—personal, informational, argumentative. A single task will not reveal improvement in other kinds of writing and may narrow program goals as teachers teach to the test.

### Writing as a Process

The best writing programs are likely to change writers' processes, especially the way they use writing time as they begin sooner, think more, revise at a deep rather than surface level, proofread more carefully. Although the short impromptu essay test is generally adequate to measure the abilities of young writers, or those of older students who always dash off a single draft in 30 minutes for better or worse, this kind of test does not reveal changes in what students can do when they learn to invest more time in their writing.

The short timed test produces short, easily scored samples, allowing quick ranking of individuals. But program evaluators are not asking which students can do the task best and fastest. Rather they need to know how many students can do the task at all, compared to performance earlier in the year, or compared to students in other programs. They need to know *how much* better each writer can do the task given the strategies and

(Continued on page 21)

## KEECH (continued)

resources they gained from the writing program.

Evaluators can simulate conditions for multi-draft writing without sacrificing test controls by providing two- or three-day test sequences, with discussion and exploratory writing on day one; sharing and drafting on day two; final drafting and editing on day three. At the beginning of the term this sequence provides for diagnosis based not only on a finished product but on the whole composing process. At the end, the test is able to reveal which students have learned to use extended composing time to good advantage, in addition to how well students at each ability level have learned to write under conditions more like those allowed for school assignments or self-initiated writing.

Evaluation of portfolios containing both impromptu and multi-draft samples as well as a variety of discourse types provides the ideal portrait of the individual writer and the best data for assessing effects of instruction. But work is only just beginning on quantifying the levels of performance or amount of growth revealed in these rich and complex data collections.

### Flaws that Signal Advances

Holistic scoring has been criticized for failing to reveal particular strengths and weaknesses of the writer. Some evaluators prefer the information available from mechanical error counts, but these are grossly misleading if errors are not analyzed for their effects on readability or for the level of attainment they signal. Writers are actually unable to make some errors before they reach relatively high levels of performance; dangling modifiers cannot appear until the student begins to use participial phrases; failure to mark non-restrictive clauses cannot appear in the papers of students who do not use relative clauses at all. The absence of error common in papers of students who merely play it safe, using simple syntax to avoid run-ons, writing only words that are easy to spell, cannot be regarded as signalling advanced competency.

Holistic scores are superior to error counts in assessing how the balance of strengths and weaknesses affects a reader, but they respond erratically to the uneven nature of development. A pre-test sample, produced when a student is easily able to do all that he attempts, may score high, having few errors and a fluent, easy style, while a post-test sample, produced when he is applying new skills not fully assimilated or is attempting a more complex intellectual task, may score low, showing a fresh rash of errors or awkwardness resulting from a shift in the area of concentration during writing. <sup>4</sup>

Post-holistic analysis of the papers <sup>5</sup> can break out performance on any number of features which may have affected the holistic score, providing important information for diagnosing individual writers or understanding program effects, in some cases revealing changes that were masked by the holistic scores. In addition, work is underway to measure increases in the cognitive complexity of what a student undertakes which may cause unexpected drops in holistic scores. <sup>6</sup> Research examining student texts at the micro-level may help identify which characteristics of student writing commonly regarded as rhetorical flaws may be associated with advances to higher levels of communicative competence.

Meantime, program evaluators need radical new techniques for recognizing and measuring positive