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 NERA annual meeting, San Francisco, 1981. Piche available from ERIC Clearinghouse, 1111 Kenyon Road, Urbana, IL, 61801 (ERIC ED 220268).

The single-sample, holistically scored, impromptu writing assignment, developed by 8 for college entrance tests and widely used in local school assessments, 1 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 leaps 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. 2 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." 3

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)

Charles R. Chew

A COMPETENCY TEST IN WRITING: AN IMPETUS FOR CHANGE

New York State has a mandated Regents competency program in three basic skills areas: reading, writing, and mathematics. Students must meet competency requirements in these areas in order to receive a high school diploma. The focus of the material which follows will be on the writing test part of the competency program. A Preliminary Competency Test in Writing (PCT) is administered to students in grades 8 or 9, depending upon the administrative set-up of the school district. This test is a direct measure of students' writing ability and requires students to write three specific types of writing: a business letter of order, a report from given data, and a persuasive essay. The PCT's are evaluated holistically at each school by three teachers. The ratings are then averaged. Students scoring below the statewide standard of 65 must receive remedial instruction, directions which are detailed in the publication, New York State Preliminary Competency Test Writing: Manual for Administrators and Teachers.

Students scoring between 50 and 64% on the PCT's receive individual additional instruction in writing as part of the regular English curriculum. These instructional strategies include:

- using students' own writing as a basis for instruction rather than assigning more drill in unrelated material;
- grouping students within the regular English classroom according to their compositional needs;
- using individual instruction and differentiated assignments;
- using such exercises as sentence combining to develop syntactic fluency;
- using interviewing techniques with students;
- encouraging peer teaching;
- providing practice in editing and revising.

Students scoring 49% or below on the PCT's receive intensive instruction in writing. Such students seem to benefit most from a tutorial approach during all stages of the composing process. Special attention is given to the prewriting stage of the composing process in order for students to gain verbal power, to learn to observe and describe, and to develop a positive attitude toward themselves and their work. Remediation must be documented and continued until the student has proved to the satisfaction of the school district that weaknesses have been overcome—or the student passes the Regents Competency Test in Writing (RCT).

The RCT in Writing is first administered in January of the eleventh year. This test consists of three writing tasks—a business letter of complaint, a report from given data, and a persuasive essay. A different version of this test is offered three times a year. Students who do not meet the statewide reference point of 65 or above must also receive remedial instruction. Remedial help at this stage should, of course, include all of the aspects detailed above, but may also need to be more test specific at this point. Students may continue to take the test, if failed, until they are 21.

The Regents Competency Test in Writing has had a direct and powerful influence on classroom instruction, curriculum development, and in-service programs for teachers, and on the approximately 200,000 high school

(Continued on page 14)
KEECH (continued)
resources they gained from the writing program.

Broderick (continued)
were in fact teaching writing.

In saying this, I should also add that these students
said nothing about some materials specially designed
to help them. When students pick up the packets of reading
a month before the exam, they also get a Fact Sheet, a
Sample Packet for Examinations, and a twenty-two page manual,
Nine Elements of Proficiency in Writing, which explains
the criteria used in grading the exam. I remember that in
1979 the manual was highly praised by students
preparing for the first exam even though the exam was,
as I have said, generally resented. Ironically, now that
the exam itself has become almost routine and the manual is
cited in many Core courses, no student even mentioned
it. More about that later.

In any case, the twenty bluebooks showed something
else, that students did not view the exam as testing a
special skill or talent. "The exam requires nothing,"
said one who was thankful for the month allowed for "study
and consideration" of the readings. Another found that a
main theme in exams and grades in Liberal Arts was a
complaint about students' inability to "think critically,
their "inability to reason;" she decided that the exam
provided a valid way of testing that ability to think
critically. Just so, the manual stresses thought; the first
five of its admonitions are: 1. Answer the question
directly; 2. Support your answer with clear, developed
arguments; 3. Show that you have read the source
materials accurately; 4. Define key terms where
necessary; 5. Focus on the intellectual issues in the
questions and avoid irrelevant digression into personal
reminiscences and anecdotes. Only the last four of the
manual’s admonitions deal with language, paragraphs,
sentences, grammar, and spelling. All of the bluebooks
showed that the students had learned that the exam was
designed to test their ability to think in writing.

The few students who expressed their disapproval of the
exam also trimmed their opinions. One who frankly
said, "I am not crazy about it; so I don’t want it," also said
that "the impulse to get rid of it might be an unthinking,
"bureaucratic" response to a bureaucratic requirement.
Another student saw the exam as part of an attempt to
make the University "selective and elitist;" for him and for
other students the exam was a "psychological obstacle" to
"our pursuit of a liberal education." He wanted to
eliminate all exams and grades in favor of tutorials and
senior exams; but he expressed, too, his worry about
students' need to compete in an increasingly
technological society. All in all, the criticism of the exam
in the bluebooks was limited and, allowing for the fact
that bluebooks inhibit protest, comments were thought-
ful—certainly a long way from the formulaic sit-in claim
that it was discriminatory, elitist, and punitive.

Thus, while the June bluebooks showed some
skepticism about the exam, students overwhelmingly
accepted it as a part of a sensible and desirable college-
wide emphasis on writing. Of that elaborate, overlapping,
and, to use my son's basketball term, "massive" effort to
provide instruction in writing, the students were
conscientious and properly appreciative. But they were
indifferent, naturally enough I suppose, to the design of
the testing program. They ignored its commitment to a
"subject matter" that minimized cultural and social
disadvantages of some students, that is, the test's
commitment to a topic adequately presented in readings
distributed to everyone who planned to take the test. The

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.

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 can simulate conditions for multi-draft
writing without sacrificing test controls by providing two
or three-draft test sequences, with discussion and
exploratory writing on day one; sharing and drafting on
day two; and following according 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 pre-compose,
re-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.

Holistic scoring has been criticized for failing to reveal
particular strengths and weaknesses of the writer. Some
evaluators can simulate conditions for multi-draft
writing without sacrificing test controls by providing two
or three-draft test sequences, with discussion and
exploratory writing on day one; sharing and drafting on
day two; and following according 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 pre-compose,
re-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 can simulate conditions for multi-draft
writing without sacrificing test controls by providing two
or three-draft test sequences, with discussion and
exploratory writing on day one; sharing and drafting on
day two; and following according 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 pre-compose,
re-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.

Holistic scoring has been criticized for failing to reveal
particular strengths and weaknesses of the writer. Some
evaluators can simulate conditions for multi-draft
writing without sacrificing test controls by providing two
or three-draft test sequences, with discussion and
exploratory writing on day one; sharing and drafting on
day two; and following according 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 pre-compose,
re-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.

Holistic scoring has been criticized for failing to reveal
particular strengths and weaknesses of the writer. Some
evaluators can simulate conditions for multi-draft
writing without sacrificing test controls by providing two
or three-draft test sequences, with discussion and
exploratory writing on day one; sharing and drafting on
day two; and following according 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 pre-compose,
re-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.

Holistic scoring has been criticized for failing to reveal
particular strengths and weaknesses of the writer. Some
evaluators can simulate conditions for multi-draft
writing without sacrificing test controls by providing two
or three-draft test sequences, with discussion and
exploratory writing on day one; sharing and drafting on
day two; and following according 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 pre-compose,
re-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.

Holistic scoring has been criticized for failing to reveal
particular strengths and weaknesses of the writer. Some
evaluators can simulate conditions for multi-draft
writing without sacrificing test controls by providing two
or three-draft test sequences, with discussion and
exploratory writing on day one; sharing and drafting on
day two; and following according 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 pre-compose,
re-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.

Holistic scoring has been criticized for failing to reveal
particular strengths and weaknesses of the writer. Some
evaluators can simulate conditions for multi-draft
writing without sacrificing test controls by providing two
or three-draft test sequences, with discussion and
exploratory writing on day one; sharing and drafting on
day two; and following according 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 pre-compose,
re-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.

Holistic scoring has been criticized for failing to reveal
particular strengths and weaknesses of the writer. Some
evaluators can simulate conditions for multi-draft
writing without sacrificing test controls by providing two
or three-draft test sequences, with discussion and
exploratory writing on day one; sharing and drafting on
day two; and following according 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 pre-compose,
re-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.