DOCUMENT RESUME

ED 158 310 CS 204 294

AUTHOR Newkirk, Thomas
TITLE The Mass Testing of Writing: How Well Is It Being Done?
PUB DATE [77]
NOTE 33p.; Report prepared at the University of New Hampshire.

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS Basic Skills; *Composition Skills (Literary); *Standardized Tests; Test Bias; Test Interpretation; *Test Validity; Timed Tests; *Writing Skills
IDENTIFIERS *Educational Testing Service; *National Assessment of Educational Progress

ABSTRACT

The validity of current standardized competency tests for writing is in doubt as is the need for such testing at all. Some tests, especially those requiring little writing, may not be testing what they purport to test (content validity). Instructional validity (testing what has actually been taught) raises the issue that what is being tested is precisely the teaching which brought on the demand for testing in the first place. Many states are using tests (such as those produced by the National Assessment of Educational Progress and by the Educational Testing Service) which provide inadequate time for the student to write, provide inadequate motivation for students to write their best, or require students to write on subjects about which they may have little information. Other tests are so simple as not to ensure even marginal literacy or they focus on mechanical skills exclusively. The data which brought about the demand for testing are open to question. Cultural bias on tests, inadequate understanding of test results, and irresponsible use of test data create confusion about literacy levels and make the term "functional illiterate" meaningless. The demand for competency testing comes when a major breakthrough in teaching reading may preview a similar development in teaching writing. The school system in England and the Bay Area Writing Project may provide useable models for teaching and testing writing skills. (TJ)

The Mass Testing of Writing: How Well Is It Being Done?

Thomas Newkirk
Assistant Professor
English Department
University of New
Hampshire

When the U.S. government set out to innoculate the country for swine flu, the decision seemed both reasonable and noncontroversial. The disease was a serious one, and national health officials assured the president that a mass innoculation program could be set up. Vaccine was rushed into production, and vaccination centers were opened. But a few weeks into the program a number of those who had been vaccinated contracted a frightening paralyzing disease, and the vaccination program came to an abrupt halt. The government was faced with over fifty million dollars in lawsuits, and most likely, will be reluctant to initiate any kind of mass innoculation program in the near future.

There is a parallel between the swine flu innoculation program and the competency testing programs that have been mandated in thirty-three states. There is the seeming simplicity of the problem; there is the assurance by testmakers (most of who stand to make big profits by selling tests to states) that such testing is feasible; there is the haste in setting up the programs. There is also the possibility that such testing will work to the severe disadvantage of some individuals and groups. And it is very likely that once the complexity and costs of mass testing become apparent, state legislatures will have second thoughts about continuing these programs.

For now, however, the competency testing bandwagon is a very comfortable place for politicians to be. A recent Harris Poll found that two-thirds of these polled favored the idea

of nationwide standards for high school students. Evidence such as this has prompted Representative Mottl of Ohio to introduce a bill calling for the nationwide testing of basic skills. Despite such public pressure, a panel, appointed by the Department of Health Education and Welfare, rejected the notion of nationwide or even statewide competency testing:

> Any testing of minimal competency standards for awarding the high school diploma -- however understandable the public clamor which produced the current movement and expectation -- is basically unworkable, exceeds the present measurement arts of the teaching profession, and will create more social problems than it can conceivably solve.[1]

A statement such as this is bound to be met by skepticism by legislators and laymen. It must seem incredible that a society so preoccupied with measurement and evaluation lacks the technology to test basic skills. Educators are surely trying to hide their failures, or, as an editorial in the Pensacola News argued:

> One of the best reasons we can think of for requiring the passage of functional tests before students are granted a high school diploma is that so many people seem to fear them, students and school officials alike.[2]

Educators, if their protests are to be taken as anything more than defensiveness, must clearly explain the inadequacy of the tests now being used. In this paper I will focus on the sections of these tests which purport to test writing ability.

Any discussion of testing requires an explanation of some of the jargon of testing. One key term is validity.

particularly content validity . A test has content validity if it truly tests that which it claims to test. For example, a test which claimed to test writing and only required students to spell a list of words would lack content validity because it tests only one aspect of writing.

One of the major questions concerning content validity is whether a writing test can have such validity if it requires little or no writing. The Educational Testing Service* has long claimed that an actual writing sample is not necessary, although due to pressure from English teachers, they will include a twenty-minute writing sample in the December, 1978 English Composition Test. The ETS justifies the predominance of multiple-choice questions by claiming that such questions allow the test to sample a wider range of the student's knowledge, and they minimize the danger that the student might be severely penalized for misinterpreting an essay question or for having nothing to say on the assigned topic. The ETS also cites research that demonstrates a high correlation between the multiple-choice questions they use and the actual writing the students do.[3]

Critics of this approach to the testing of writing argue that it runs counter to a basic concept of fairness. If tests are to measure writing ability, they must evaluate the writing of the student. Once we begin to enter the world of correlations, we enter a slippery world indeed. On the New Jersey Basic Skills Test, for example, there is a substantial section which tests a student's ability to complete analogies. No doubt such items correlate highly with language

* Hereafter referred to as the ETS.

language ability, and are good predictors of future performance. Yet it can hardly be argued that completing an analogy is a basic skill. It has little direct relationship to the actual language activities students must perform. Is it fair to test basic skills by testing non-basic skills?

Critics also argue that the ETS justification rests heavily on the claim of expediency. It is expedient to construct a one hour test, most of which can be machine-scored. It is expedient to rely on measures that correlate well with writing ability rather than going to the trouble of evaluating a fair sample of the student's work. Such a sample, according to John Mellon, would be 800-1,000 words, drawn from four or five essays.[4] Can states, on the one hand, attempt to make decisions on the proficiency of individual students, and on the other rely on measures that bear a largely statistical relationship to this proficiency?

A new type of validity has been proposed in reference to the competency movement -- instructional validity. A test has instructional validity if the objectives tested correspond with the objectives of the instruction that the student has received. In other words, it is unfair to test students on material or skills they have not been taught. Merle McClung, attorney for the Center for Law and Education in Cambridge, has argued that the Florida tests lacked this kind of validity, and that the concept of instructional validity may be the basis for legal cases brought by students.[5]

At first glance this concept of validity would seem self-evident. But it follows that if states are to administer

competency tests, the objectives of the tests must coincide with the curricula in the schools, many of which have widely different curricula. Competency testing presupposes a uniformity of objectives that most likely does not exist.

The concept of instructional validity also leads to an inescapable dilemma. To set objectives that do not coincide with the school curricula is to test competencies that students may not have been taught, but to test objective that do coincide with school objectives is to, in effect, condone the educational practices that produced the "crisis" in the first place. To take writing instruction as an example, the National Council of Teachers of English has consistently argued that students often fail to learn how to write because of the preoccupation with grammar and correctness in the schools. But if a statewide test is to have instructional validity, it must focus on what is taught, grammar and correctness.

The acceptable validity of any test depends on the type of decision that the test will help determine, the more important the decision, the greater the level of validity required. If a teacher wants a rough idea of those students who might have basic writing problems, a short essay on the first day of class may be valid enough. Such a decision is reversible; the teacher can modify an initial judgment. Decisions, based on competency tests, about who is and is not to graduate are far less reversible, in part, because legislators and parents seem to distrust the reversible information that a teacher might provide. While the student may be given a second and even a third chance, the judgment of his competency is based on no more than a few afternoon's work.

Given the significance of the decision to graduate or not to graduate, and given the limited reversibility of the decision, it would follow that the method of testing be as valid as possible. As Terry TenBrick, a specialist in educational measurement has stated," The information needed to make these kinds of decisions must be as accurate as possible, no matter what the cost."[6]

It is the decision to exclude teachers from the determination of competency that reduces the validity of competency tests. Tests alone, so the legislators state, must determine competency. Yet most textbooks on educational measurement argue that important decisions about pupils should not be made without confirmatory evidence from parents and teachers, and the opportunities to change such decisions.[7] The need for confirmatory evidence is particularly acute in judging writing ability, for writing performance can vary greatly from topic to topic.[8] It is extraordinarily difficult to find a single topic that will elicit the best work from all students tested.

Some Writing Exams: The Trivial

According to Gary Hart, author of the California Pupil Proficiency Bill, one of the purposes of his bill was to restore meaning to the high school diploma and thus improve the public's attitude toward their schools.[9] This rationale is echoed by most proponents of competency testing. Many of the tests devised, however, are either so simple, or so marginal to the basic skills of writing, that it is unlikely that they would meet Mr. Hart's objective.

The items devised for the New York State Competency Tests seem to be so easy that it is doubtful that the public would be assured of the literacy of a student scoring above the 65% cut-off point. One of the sample questions in mechanics is:

Somebody _____ the window.
has broke
bust
has broken
bursted [10]

The questions in the sentence writing section are even easier. Students are asked, for example, to use "will eat" in a sentence. According to the rules for scoring, all students must do is to complete the sentence with no mechanical errors, so presumably the students could correctly respond with the sentence-- "I will eat."[11]

One of the proficiency tests that is generally taken as a model was developed by the Denver, Colorado school system. Students have to do no writing on these exams, but are given a "Language Proficiency Exam" which asks them to proofread 50 lines that "might have been written by a student."[12] Surely "language proficiency" entails more than proofreading. Another proficiency test, proclaimed by its developers as a success, was developed by the Westlake, Nebraska school system. The writing exam requires that the student write three related paragraphs on a topic agreed on with the teacher and that the writing contain no more than five grammatical erors.[13]

Such attempts at measuring writing competency are inadequate, not because the mechanics of writing are unimportant, but because "proficient" writing is more than "cor-

rect" writing. Imagine the student navigating his way through the three required paragraphs at Westlake, aware that his primary task is to avoid errors. He may avoid syntactic errors by using short sentences; he avoids difficult spelling words by using simple words or circumlocutions; he avoids contractions, and, in general, he tries to be as timid and safe as possible. Is this a measure of competence? Or does it reinforce the the negative image that many Americans have about writing? Don Graves, who surveyed the teaching of language arts in the United States, described this attitude:

> In America, writing is basically a form of etiquette in which words are put on paper, not necessarily with clarity, but free of mechanical errors. The American does not feel he belongs in writing. He is similar to the person who has been reluctantly invited to a party of distinguished guests. Being a person of modest station he attends with great discomfort. He has but one thing on his mind -- to be properly attired, demonstrate the finest manners, say nothing, and leave quickly.[14]

Tests such as those mentioned above have prompted some educators to compare the claims of the testmakers to the deceptive "bait and switch" advertising tactic.[15] The merchant advertises an almost unbelievably good deal (the bait), but when the customer arrives at the store that product, unfortunately, is sold, but would the customer like to see something almost as good (the switch). Those who design the competency programs make claims similar to that made by the ETS that competency programs will "certify that the student (granted a high school diploma) possesses the basic skills needed to be a productive adult citizen capable of functioning in a complex society." The tests then proceed to test marginalia.

In fairness to the testmakers, many of them are in a no-win situation for they must deal with the "Groucho Marx Complex." Groucho once said that he could not respect any club that would have him as a member. The only way a club could keep his respect would be for it to exclude him. If almost all students pass competency tests, the public will view them as too easy, too unselective. Only if a substantial number of students fail will tests be perceived as hard enough. But such a failure rate will obligate the state to provide remedial education at a substantial cost. And all evidence indicates that many of these failures will be minority and poor students who generally do poorly on standardized tests. As Stephen Bailey, a member of the HEW panel on competency testing, stated:

> If success on tests -- for the purposes of graduation or promotion -- is achieved by four-fifths of a suburban school system but only one third of a central city's system, the consequences could be serious for domestic tranquility as well as social equity in a world where a high school diploma, regardless of intrinsic meaning, is frequently a ticket to particular jobs.[16]

Given this ominous possibility, it is not difficult to see why some states and communities have selected easily-achieved levels of competence.

Some Writing Exams: The Insufficient

Most of the states which are developing competency tests have chosen to evaluate writing for more than mechanical correctness. Many states, rather than developing their own writing exercises, have used exercises developed by the National Assessment of Educational Progress* or they have chosen

* Hereafter referred to as the NAEP

tests developed by commerdial publishers such as the ETS or the Measurement Research Center. There appear to be two reasons for this dependence. First, many state legislatures have set deadlines that do not allow for the time-consuming process of field testing and consultation. Second, state departments may feel secure in selecting exercises that supposedly have been validated through field tests.

But this dependence creates other problems. As noted earlier, if competency tests are to have instructional validity, school curricula must stress the objectives of the tests. Even if states determine objectives, the testmakers will specify what types of writing will meet the objectives. In effect, major curricular decisions may be made by independent and semi-independent organizations. Clearly, the types of exercises that are being passed on to the states must be carefully examined.

Any competency test is artificial to a degree. A real test of competency would take place in a situation where the student actually had to use a skill to get what he wanted. Here motivation would not be a problem. A writing competency test, artificial at best, can easily become so artificial as to become an inadequate test of writing. To go back to the traditional rhetorical triangle (writer--audience--subject), there is an inevitable problem with audience. The student, even if he is told to write for a friend, or teacher, is ultimately writing for an unknown examiner. What a test can assure is that the student not be asked to write about an unknown subject, for if he must write about a topic he knows little about to someone he knows nothing about, it is un-

likely, to say the least, that he will produce his best work.

Despite the readiness of states to use NAEP exercises, the record of the NAEP has not been particularly good when it comes to stimulating writing. John Mellon, in his analysis of the first round results (1969-70), claimed the NAEP failed to adequately motivate the students to write. By handing out topics and expecting students to do their best work, the NAEP seemed to copy the technique that fails so regularly in the classroom.[17] Mellon found the second round assesment to be little better. The average length of pieces written by 17-year-olds was only 137 words, the length of an average paragraph.[18] Unless a writing assignment can stimulate (and allow time for) the student to do his best work, the most sophisticated evaluation techniques can tell us little.

Time limits. Most of the tests which require students to write put limits on the time allowed. The ETS has generally been the most restrictive, usually allowing twenty minutes. The assumption seems to be that a good writer writes faster than a poorer one, and that anyone who found the time limit unduly restrictive would probably be writing too slowly to be an effective writer. While there is some evidence to suggest that better readers read faster than poorer readers,[19] what does research have to say about the writing behavior of good and poor writers?

Charles Stallard compared the writing behavior of a group of good 12th grade writers with that of a group of

randomly selected 12th grader. He found that while students in the "good" group wrote more words on the average (356 words compared to 304), they took considerably longer to complete their work (45 minutes compared to 23 minutes). In other words, the better writers wrote considerably slower than the randomly selected writers (8.7 words/minute compared to 13.5 words/minute).[20]

One of the most significant studies, both for writing researchers and for competency testers, was carried out among community college students by Sara Sanders and John Littlefield.[21] The experimenters pretested 25 students in two ways. They were given an impromptu essay, and they were allowed to choose and research a topic and then write on that topic. In each case students were given two class periods to write. After five weeks of instruction which emphasized the writing process--pre-writing, writing, revising -- the students were given a similar dual post-test. While there were no significant differences between the pre- and post-instruction impromptu essays, the researched papers done after the course were significantly better than those done before the course.

Sanders and Littlefield question the traditional use of the impromptu essay for the evaluation of writing. They argue that to require such an essay is to put the student in an artificial situation, and the impromptu essay does not fit the type of instruction that emphasizes the writing process:

> Unfortunately, the rigidly controlled essay test surely represents the ultimate in an artificial writing situation; as such it has been shunned in many modern composition courses. While the

> test essay is precisely timed, in many courses the student is encouraged to think, research, write, rewrite, perhaps solicit the advice and reaction of others, and rewrite again and again. The student has no chance to go through this process in writing an impromptu essay. In addition, the timed situation, which places a premium on verbal fluency and the speed of composition, may particularly work to the disadvantage of the slower student who needs more time to demonstrate what he has learned.[22]

Other researchers, while not rejecting the use of impromptu essays, urge generous time limits. The most respected guide for the conduct of writing research is the NCTE monograph, Research in Written Composition. The authors claim that "20-30 minutes seems ridiculously brief for a high school or college student to write anything thoughtful."[23] They suggest that junior high students be given 50-70 minutes and high school students be given 70-90 minutes.

Don Murray, who has made an exhaustive study of the writing habits of published writers, argues that delay is an essential part of the writing process. Writers need a period of what Wordsworth called "wise passiveness" before beginning to write. Murray is contemptuous of the practice of evaluating students on the basis of short impromptu essays.

> [illegible] when our "bad" students hesitate, stare out the window, dawdle over blank paper, give up and say, "I can't write," while the "good" students smugly pass their papers in before the end of the period.
>
> When publishing writers visit such classrooms, however, they are astonished at students who can write on command, ejaculating correct little essays without thought, for writers have to write before writing.
>
> The writers were the students who dawdled, stared out the windows, and, more likely than [illegible]

like to admit, didn't do well in English -- or in school.[24]

Even young children need a period of time to rehearse what they are going to write. Graves, in his study of the composing process of seven-year-olds, noted that many had to draw and talk aloud before they could write. This activity was indispensable preparation for such students.[25]

What then are we to make of the claims for the validity of writing tests. The ETS, for example, claim in their manual for the Basic Skills Assessment that:

> While no process can be said to guarantee the existence of content validity, the steps carried out in developing the specifications for the BSA certainly resulted in a high probability of measures with content validity. ... a comprehensive review of research, other relevant materials from testing programs, and information from local school districts and professional organizations provided the program with a large data base on the assessment of basic skills which could be used to develop test specifications.[26]

To claim, as the ETS does, that a large body of research suggests that their measures are valid is to ignore the fact that another body of research and informed opinion rejects the validity of the 20-minute writing sample as a measure of writing proficiency. Critics of the ETS would argue that educators who claim to assess the writing ability of individual students have an ethical obligation to allow the student to write under the best possible conditions, including time to plan, write, and revise. A writing sample is not like a blood sample, it cannot be extracted in a few minutes.

Creative writing. The first round NAEP exercises were criticized for their narrow focus. Imaginative or "creative" writing was not called for. Subsequent exercises illustrate the general confusion about creative writing. As with many changes in English Education, the movement toward creative writing, was a reaction against the dominant mode of instruction, in this case the exclusively analytic and functional approach to writing in many U.S. schools. Those who in the 1960's reacted against this dominance, Ken Macrorie, James Moffett, Don Murray, and others, argued for the importance of the perceptions, experiences, and voice of the individual student, the importance of the "I" that was so often prohibited by high school teachers.

These rebels did not advocate fictional writing where all trees are "leafy" and all brooks "babble." Nor did they advocate the shapeless type of writing where students were simply encouraged "to write what they feel." Few liked the term "creative writing" because the term implies that some writing (usually fiction) is creative and other writing (usually, non fiction) is uncreative. But in many classrooms creative writing turned into a type of fanciful writing where the stu- [illegible]

[illegible] to call [illegible]

[illegible]

the NAEP and [illegible] for its writing competency test [illegible] the purpose [illegible]

cise, according to the Ohio manual, is for the student to "demonstrate ability in writing to reveal personal feelings and ideas through free expression:"

> Sometimes people write just for the fun of it. This is a chance for you to have fun writing.
>
> Pretend that you are a pair of tennis shoes. You've done all kinds of things with your owner in all kinds of weather. Now you are being picked up again by your owner. Tell what you, as the tennis shoes, think about what is going to happen to you. Tell how you feel about your owner.[28]

Surely their are better ways to assess the ability of the student to express personal feelings than to ask him to imagine himself a pair of tennis shoes.

Another topic developed by the NAEP illustrates the same danger. Students are shown a picture of five children playing on an overturned dory. The dory is on a dock which reaches into a small inlet where sailboats are moored. Those taking the test were given the following instructions:

> Look carefully at the picture.
>
> These kids are having fun jumping on the overturned boat. Imagine you are one of the children in the picture. Or if you wish, imagine you are someone standing nearby watching the children. Tell what is going on as he or she would tell it. Write as if you are telling this to a good friend, in a way that expresses strong feelings. Help your friend feel the experience too.[29]

Such an exercise violates the most basic rules for a writing assignment. In the first place, the children in the picture seem to be about 8 years old, and few 13 and 17 year old students (who were given the assignment) would be motivated to write about children so much younger than they are. Although I distrust the automatic criticism of socio-economic bias, it does seem likely that the setting would be

unfamiliar to students who have spent no time at seaside or lakeside resorts.

But the biggest problem with such an assignment is that it seems to invite a highly sophisticated phoniness. It forces the student into falseness. The following piece was given a top rating by NAEP judges on this assignment. This piece, according the the NAEP, domonstrates an "imaginative entry into experience:"

> Jumping and running on the boat's very enjoyable. Up we jump and down we float. I feel as if I could sail the boat around the world and back. The salty air blows through my nostrils. My body is engulfed in this salty concoction. The wind beats against my cheeks.
>
> The white glistening enamel underside of the boat feels like wilk(?) to the touch. The trees are alive, pulsating, watching our childish games.
>
> I feel like I could play forever. No concept of time, no stresses encourage my exuberance.
>
> My body has separated from my spirit. I am no longer encaged in a prison of bones and skin. There are no barriers now. I can do whatever I want, whenever I want. (age 17)[30]

Surely this is a skilled writer performing an awkward task. The writer has been pushed into a difficult, and I believe dishonest, stance by a topic that on the one hand asks for personal feeling and on the other takes the writer away from the [illegible] which he feels strongly about.

Some other exercises give [illegible] difficult to see how the writing students do can be evaluated. Washington state used the following exercise on its eighth grade writing test:

> Music does different things to different people. Perhaps it makes you have one feeling or another. Perhaps it reminds you of some place or something happening.

> Now listen to this piece of music and write about what things this piece of music does to you. Start writing anytime you wish.[31]

while such an exercise may be a good one for free writing, it seems unlikely that it is specific enough to give the student an idea of what is expected.

Writing and information. Writing must be about something. It is not some set of skills that can be assessed apart from the act of communication about a subject. The greatest and most consistent weakness of the NAEP writing exercises is their failure to ask students to write on subjects they know something about. They move the student into areas where his knowledge is severely limited. He often finds himself in some hypothetical territory where he must be "creative," or he is pushed into what Don Graves has called "extended territory": national and international events and personalities. When the student must write about this extended territory, he often writes with limited information. An example:

> Everyone knows of something that is worth talking about. Maybe you know something about a famous building like the Empire State Building in New York City or something like the Golden Gate Bridge in San Francisco. Or you might know a lot about the Mormon Tabernacle in Salt Lake City or the new sports stadiums in Atlanta or St. Louis. Or you may be familiar with something from nature like Niagara Falls, a gigantic wheat field, a grove of orange trees, or a part of a wide muddy river like the Mississippi. Choose something you know about. It may be something you have seen while travelling, or something you have studied in school. Think about it for a while and then write a description of what it looks like so that it could be recognized by someone who read your description.[32]

while ostensibly this is an open assignment, the implication is that the student should write about something

"important," something famous, and not something as insignificant as the student's home or school or favorite park. The writing samples that such exercises elicit invariably deal with extended territory and illustrate the difficulty students have with such assignments. The following piece was written for a similar NAEP assignment where students were asked to write about a person they admired. As usual, the possibilities suggested were all national figures, Mickey Mantle, Winston Churchill, Martin Luther King, Jr. This piece was rated in the 87th %ile of the work done by 17-year-olds:

> Dr. Christain Bernard: I believe he is a person worth looking up to. He has tried to make our life longer for us through research about operations on heart transplants. His determination to help mankind is recognized even though public opinion is very much against the practice of transferring one person's heart to another person. This determination shows how a true doctor or any kind of scientist would work for the betterment of of mankind, both for a longer life and easier world to live in. Dr. Bernard shows the dedication of a true doctor, to help man when he is sick or dying, as are many of the people on whom he operates. His determination makes him try to show the world that the heart is not a sacred organ of the body, and is just like any other part of the body when it needs to be repaired. His determination to find a better method, new drugs to help after surgery, is to be admired. This is one of the reasons I admire him.[33]

I find this writing dismal, what Ken Macrorie has called "English." The student lacks anything specific to say about Dr. Bernard and must resort to abstractions and platitudes.

Even some of the more practical types of writing create problems. One exercise requires students to write a report about the moon:

> In the box below are some facts about the moon which you can use for your report. You may add other facts that you can remember about the moon

from your reading and classwork, from television, or from listening to people.

...Be sure to report the facts in an order that will make sense to your classmates.

Facts about the Moon

Made of rock
Mountainous, contains craters
Covered with dust
No air or water
No plant or animal life[34]

Unless the student has some knowledge of the moon (and those who received the highest marks seemed to) the exercise must have been puzzling. With the exception of showing a causal relationship between the fourth and fifth fact, the facts are clear in themselves. Such an assignment would also seem to be a questionable test of report writing, denying as it does the possibility of conducting research.

The problem of inadequate information is also evident in the letter writing exercises. One of the NAEP exercises shows a notice of a summer job opening in a clothing store. The student is asked to do the following:

> Chris Jones lives at 3600 March Street in New York, New York 10004. Chris has finished the junior year at high school and has been looking for a summer job. Chris spotted the advertisement in the New York Times and decided to apply for the job. Write Chris' letter of application to Mr. Fried.[35]

The student is again put in a curious position. He is to write Chris' letter although he knows nothing about Chris except his or her address and year in school. The criteria used by the NAEP in evaluating student letters clearly show that to be awarded the higher marks the letter should included references, a statement of qualifications, and a way to be contacted. While these are valid criteria for such

a letter, the instructions do not make clear the necessity of the student inventing such information.

Conclusions and Discussion

1. While the multiple-choice test of writing is cheaper and more convenient to administer than a test which requires extensive writing, the multiple choice test lacks content validity. Any claim for the validity of such tests rests on the correlation between them and actual measures of writing performance.

2. Competency tests raise the issue of instructional validity, the correspondence between the objectives measured and the instruction the student has received. A legal argument can be made that it is unfair to test students on material they have not been taught. Yet to focus on the objectives taught in the schools is both difficult, due to the lack of uniformity from school to school, and profoundly conservative, in a sense endorsing the practices that led to the current dissatisfaction.

3. The acceptable level of validity of a test depends on the type of decision that is to be made, the more important the decision the higher the level of validity required. If a test is to determine promotion or graduation, it must be as valid as possible, no matter what the cost.

4. Many states rather than developing their own writing exercises are adopting those of the National Assessment of Educational Progress, the Educational Testing Service, or other national testmakers.

5. Some of those exercises that have been developed by states and local communities are either so simple that they do not ensure even marginal literacy, or they focus exclusively on mechanical skills. The emphasis on mechanics reinforces the myth that "proficient" writing is "correct" writing.

6. Many of the tests provided by national organizations, particularly the Educational Testing Service, may well fail to give students adequate time to demonstrate their writing ability. There is evidence to suggest that better writers may be slower writers and that the impromptu essay may not be an adequate guage of writing ability.

7. While it is extremely difficult to devise writing assignments that are stimulating for all students, those used on competency tests exhibit some glaring weaknesses.

the most serious being the tendency to require students to write on topics about which they have little information.

Much of the demand for writing competency tests and other tests of language arts has been caused by repeated claims that a large percentage of high school graduates, indeed a large percentage of adults, are "functionally illiterate." Illiteracy is proclaimed a national disease. A study by researchers at the University of Texas found close to 20% of adult Americans to be "functionally incompetent."[36] The New York Times reported the results of a Canadian study which found 37% of Canadian adults to be functionally illiterate.[37] Newspapers also reported a Swedish study which found 15% of Swedish 12th graders to be functionally illiterate.[38] Do we have an international literacy crisis on our hands?

Closer examination of these studies reveals that some are inadequate and some are just silly. The silliest is the Canadian study. The researchers defined "functional illiterate" as anyone 15 years of age or older with less than than eight years of formal education. Using this standard they found that among Canadians over 65 the level of "illiteracy" was very high, 61%. Apparently no attempt was made to determine whether those classified as "functional illiterates" could read or write.

The Swedish study did attempt to determine the reading ability of their sample. The researchers found that 15% of Swedish 12th graders were reading at the 6th grade level. Newspapers interpreted this to mean that 15% were functionally

illiterate. Such labelling is irresponsible because no attempt was made to determine whether the students reading at the 6th grade level can function.

Studies such as this one also illustrate a common misconception about grade levels. If a headline would read -- "Fifty Percent of U.S. Students Read Below Grade Level" -- some politician is sure to proclaim a literacy crisis. But it is inevitable that the headline is true. The mythical average 12th grader reads on grade level, 50% read above this level and 50% read below it. Grade levels are statistical averages, not minimum acceptable levels. Laymen should also learn to be skeptical about some of the extremely low scores reported. A standardized test does not measure well deviant student achievement; it is pegged to the level to be tested. Therefore, the best students will find it too easy and the weakest students will find it too hard to be an adequate test. I have also found that students for whom a test is very difficult may become so frustrated that they give up, and scores of functional (but below average) readers do not differ significantly from those of non readers.

The Texas study is one of the few that makes an attempt to define the competencies needed by an adult and then to test these competencies. But Merle McClung has claimed that the tests developed by the Texas researchers are culturally biased.[39] The examples he presents raise questions not so much about the cultural bias of the tests, but about the concept of adult competence to begin with. One of the questions on the "Government and Law" section asks the student to

"discuss the concept of party politics including why the two-party system has been successful." It is ironic that such a task was developed by researchers at the university in a state that has not had a Republican governor in this century.

One would hope that there might be a moratorium on the use of the term "functional illiterate," since no one is quite sure what skills such a person would lack. Paper and pencil tests of functional illiteracy, while claiming to test survival skills, do so in an artificial context. Placed in a situation where survival were at stake, some of those who might do poorly on paper and pencil tests might do considerably better. I was always amazed, when I taught in an inner city Boston school, how many of the students "reading at the second grade level" could pass their driver's test.

One would also hope that that states, before committing themselves to costly testing programs that will create new levels of bureaucracy, will make their own attempts to determine the extent of the literacy crisis. They should not be guided by reports of mass illiteracy and isolated legal suits.

As with many educational innovations, the push for minimal competencies is not new. Arthur Applebee, in his history of the teaching of English in the United States, shows how a concern for competencies grew out of the progressive movement of the 30's and 40's. At its most creative the progressive movement saw democracy as evolving, and it saw

education as a vital force in the reconstruction of society. But near the end of the progressive movement, many educators were lowering their sights and were stressing "life adjustment." What had once been challenging and individualistic became mediocre and conformist. Applebee, in his evaluation of the progressive years, views this shift as clearly a negative one:

> In their concern for general education for the general student, they adopted a condescending position that removed virtually all "striving" and challenge from the activities suggested, especially for non-college-bound students. They allowed their empiricism and pragmatism to narrow their definitions of needs to the point where they were trivial and dull.[40]

Many of those who favor minimal competency testing claim that these minimal standards should not unduly influence the curriculum. Proponents usually follow such claims with the argument that one of the virtues of competency testing is that it will hold teachers and schools accountable. If school and teacher performance is to be judged by student performance on minimal competency tests, it is naive to assume that these competencies will not become the focus of instruction. And, as happened before, the curriculum for the general student may be reduced to the trivial and the dull. As Gordon Cawelti has put it," ...there is a real likelihood of further trivializing secondary education with a curriculum that is already regarded as irrelevant by many students."[41]

Such minimalism, according to British critic R.L. Blackmur, is itself a form of illiteracy. Writing in the early 50's Blackmur's criticism seems relevant today:

> This crisis of our culture rises from the false belief that our society requires only enough mind to create and tend the machines together with enough of the new illiteracy for other machines -- those of the mass media -- to exploit. This is perhaps the the form of society most expensive and wasteful of human talent mankind has yet thrown off.[42]

A final irony in the competency testing movement is the timing. One of the greatest educational advances in recent years has been the use of the informal reading inventory in the teaching of reading. Frank Guszak, who has worked extensively with reading teachers, has claimed that " the most powerful means of determining individual achievement levels are informal tests."[43] Such tests have the virtue of flexibility; they can be used on different days with different reading material. As the name indicates they are given informally and avoid the authoritarian atmosphere of the standardized test. They are easily scored. And if done well, they can provide a great deal of diagnostic information.

As usual, advances in writing instruction are slower. But organizations like the Bay Area Writing Project have helped teachers in diagnosing student writing problems and in aiding students in revising. In England, where all students must pass exams to receive certificates of education, a good deal of work is done with the continual assessment of writing. Rather than basing an evaluation on a single product produced under test conditions, students are evaluated throughout their 5th year (age 16). Their final evaluation is then based on on their 10 best pieces of work.

Procedures like these are pschometrically messy; they lack the rigid time and topic controls that have been a tradition

in writing assessment. But their overwhelming virtue is that they involve the classroom teacher. Those who seek to improve literacy in this way work <u>with</u> the teacher. As Seymour Sarason, who studied the process of change in schools, has noted, those who feel that they can legislate change often fail to realize that they are asking teachers to learn and unlearn.[44]

* * * * * * *

I spent one morning going through the NAEP writing exercises and the elaborate means developed for assessing writing. The more I read, the more depressed I became, not because all the exercises were bad (although some were very bad), but because there seemed to be something limiting in setting any task, then devising a set of criteria that would apply to all the writing produced. I began to feel like I was in an airless room.

I would wander to the kitchen to fix coffee and listen to my nine-month-old daughter make "b" sounds; "bbbbbBBBBBB," she would increase volume and smile, delighted with the sounds she made. I came back to my study, to minimal competencies and primary trait scoring, reluctantly. I began to read some of the student work and came to one that seemed to dispell the gloom. It was written in response to probably the most interesting NAEP topic:

> Imagine you are taking care of a neighbor's children for an afternoon. You send one of the children to the corner store to buy some peaches for a snack.

The store owner, Mr. Jones, whom you have known and liked for several years, apparently took advantage of the child. The peaches are rotten. You want to send the child back with the peaches and a note to clear up the situation.

Write a note to the grocer that expresses your displeasure and proposes what Mr. Jones should do about the situation.[45]

After exercises requiring the writing of phone messages and letters of request, here we finally had something -- melodrama. The following piece clearly baffled the evaluators who classified it "generally factual" and "not personally abusive." I like to think of the writer as someone who, as an infant, delighted in making "b" sounds.

Dear Mr. Jones;

I am writing you in regard to the peaches that were purchased by a child I was keeping. I wanted some peaches for a snack but as I bit into one I found to my horror and disgust that they were rotten. Fortunately, I kept cool. I tried so hard to forgive and forget, but the child I was keeping abviously couldn't. After eating 12 of your rotten peaches she regurgitated all over the carpet but I tried to endure it because I had a brainstorm. I could feed the remaining peaches to the dog. But as luck would have it upon eating the peaches the dog's hair fell out. However, Mr. Jones even though these terrible things happened to me I am not mad, I am merely writing to tell you I realized something about you and extend you all my sympathy. Because after this event I now realize why you have no teeth or hair, because you've eaten your own rotten peaches.

Your friend,

Lee Smith[46]

Footnotes

1. Quoted in "Minimum Competency Movement Under Attack," Council-Grams, 40; 3 (May, 1978), p. 1.

2. Pensacola News, September 29, 1977.

3. College Entrance and Examination Board, "The English Composition Test with Essay," (a pamphlet explaining the test), 1978, p.2.

4. John Mellon, "Round Two of the National Writing Assessment -- Interpreting the Apparent Decline in Writing Ability," Research in the Teaching of English, 10; 1 (Spring, 1976), p. 67.

5. Merle McClung, "Competency Testing: Potential for Discrimination," Clearing House Review, (September, 1977), pp. 446-447.

6. Terry TenBrick, Evaluation: a Practical Guide for Teachers. New York: McGraw-Hill, 1974, p. 69.

7. Frederick Davis, Educational Measurements and Their Interpretations. Belmont, California: Wadsworth Publishing Company, 1964, p. 182.

8. Gerald Kincaid, "Some Factors Affecting Variations in the Quality of Students' Writing." Unpublished D.Ed. dissertation summarized in Richard Braddock, Richard Lloyd-Jones, and Lowell Schoer, Research in Written Composition. Urbana, Illinois: National Council of Teachers of English, 1963, p. 83.

9. Gary Hart," The California Pupil Proficiency Law as Viewed by its Author," Phi Delta Kappan, 59;9 (May, 1978), pp. 592-595.

10. "The New York State Basic Competency Testing Program: A Brief Overview," The New York State Bureau of Pupil Testing and Advisory Services, 1977, p. 9.

11. Ibid.

12. Donald Henderson, "Denver, Colorado: A 17-Year-Old Minimum Competency Testing Program," Phi Delta Kappan, 59;9 (May, 1978), pp. 610-611.

13. Jim Findley, "Westlake's Minimum Competency Graduation Requirements: A Program that Works," Phi Delta Kappan, 59;9 (May, 1978), pp. 614-618.

14. Don Graves, "Balance the Basics: Let Them Write," an Information Paper for the Ford Foundation, 1977, pp. 10-11.

15. Joe Nathan and Wayne Jennings, "Educational Bait and Switch," Phi Delta Kappan, 59;9 (May, 1978), 621-624.

16. "Minimum Competency Movement Under Attack," Council-Grams, 40;3 (May, 1978), p. 2.

17. John Mellon, National Assessment and the Teaching of English. Urbana Illinois: The National Council of Teachers of English, 1975, pp. 33-34.

18. John Mellon , "Round Two of the National Writing Assessment." See footnote 4.

19. Frank Smith, Understanding Reading. New York: Holt, Rinehart, and Winston, 1971.

20. Charles Stallard, "An Analysis of the Writing Behavior of Good Writers," Research in the Teaching of English, 8;2 (Fall, 1974), pp. 206-218.

21. Sara Sanders and John Littlefield, "Perhaps Test Essays Can Reflect Significant Improvement in Freshman Composition," Research in the Teaching of English, 9:2 (Fall, 1975), p. 145-153.

22. Ibid., p. 147.

23. Richard Braddock, Richard Lloyd-Jones, and Lowell Schoer, Research in Written Composition. Urbana, Illinois: National Council of Teachers of English, 1963, p.9.

24. Don Murray, "Write Before Writing," a paper presented at the 1977 NCTE Convention in New York City.

25. Donald Graves, "An Examination of the Writing Processes of Seven-Year-Old Children," Research in the Teaching of English, 9:3 (Winter, 1975), pp. 227-241.

26. "Basic Skills Assessment Technical Manual," Fall, 1977, Educational Testing Service, pp. 38-39.

27. Mellon, National Assessment and the Teaching of English, p. 15.

28. "Writing: Released Exercises from the 1973-73 Assessment," # 101013-N, National Assessment of Educational Progress (Denver), 1976, p. 23.

29. Ibid., p. 33.

30. "Expressive Writing; Selected Results from the Second National Assessment of Writing," National Assessment of Educational Progress (Denver). ERIC # ED 130-312

31. Washington Statewide Educational Assessment, "Mathematics/ Writing," Eighth Grade Test administered in the fall of 1976.

32. NAEP Released Writing Exercise # R 203012-NT, p. 197.

33. "Group Results A, Science and Writing: Report 5 of the 1969-1970 Assessment," National Assessment of Educational Progress (Denver), 1971, p. B-46.

34. NAEP Released Writing Exercise # 203008-NT, pp. 180-181.

35. NAEP Released Exercise # 202013-S, p. 163.

36. Discussed in McClung, op. cit., pp. 441, 443-444.

37. "37% of Canada's People Termed Functional Illiterates," New York Times, December 4, 1977, p. 39.

38. Hans Grulim, "The Reading and Writing Abilities of Swedish Pupils; A Survey of the Development from Grade 1 to Grade 12." ERIC # 134-977.

39. McClung, op. cit., pp. 442-444.

40. Arthur Applebee, Tradition and Reform in the Teaching of English. Urbana, Illinois: National Council of Teachers of English, 1974, p. 175.

41. Gordon Cawelti, "Requiring Competencies for Graduation --Some Curricular Issues," Educational Leadership, November, 1977, p. 89.

42. R. L. Blackmur, The Lion and the Honeycomb. New York: Harcourt, Brace, and Company, 1955, p. 25.

43. Frank J. Guszak, Diagnostic Reading in the Elementary School. New York: Harper and Row, 1972, p. 109.

44. Seymour Sarason, The Culture of the School and the Problem of Change. Boston: Allyn and Bacon, 1971, p. 193.

45. NAEP Released Exercise #R 201008-S, p. 89.

46. Ibid., p. 95.