Richard Lloyd-Jones

SKEPSICISM ABOUT TEST SCORES

The best tests of writing understate competence. They are subject to tremendous error for particular students. Even the evaluation of writing in real situations for real audiences, that is, in direct performance, is problematical. Not only do we hear several views about the effectiveness of literary writing, but judgments about advertising campaigns, textbooks, business proposals, and news stories are often divergent. That is, the evaluation of writing even in situations where the intended result can be directly observed is often moot. Indirect measures, those with which we are concerned, are built from these disagreements and add more.

Writing samples seem most nearly to approximate the conditions of actual discourse. The problems are numerous, though:

- If a subject is presediss, then most writers must deal with subjects they do not know or want to write much about. In such instances, the essay becomes primarily a test of particular knowledge and experience.
- If no subject is prescribed, comparability is hard to achieve, and many writers become paralyzed trying to imagine what to say. Probably a good exercise fits well between extremes, and it thus has some of the faults of both.
- Time is ordinarily limited. The twenty-five minutes often allotted in National Assessment exercises or the hour or two often given for examination and placement exercises cut down the possibility of careful planning and selection of topic and structure. One must write by a formula or by whatever organizing scheme happens to occur and one has little time to re-think an issue or even re-draft a phrase.
- The writing situation is fictional at best. Often no audience is suggested, as though a meaningful statement could be removed from the situation which requires it. The only exigency for the writer is to guess what a scorer wants, but true rhetorical need depends upon a well understood social relationship. A good test exercise may implicitly or explicitly create a social need for the purpose of the examination, but the result may be heavily influenced by the writer's ability to enter a fiction.
- A corollary is that the writer's role is ambiguous. Although the demands of the exercise may require a role, the situation of the test defines the writer as a person facing a hurdle. Only a test-wise writer is likely to deal with the problem with finesse.
- A simple sample of writing is but one probe into a massive competence. Writers are not equally adept in all situations or for all purposes, so any claim to measure general competence should include several samples.

The list could be extended, but this is long enough to make the point. A writing sample is not real writing. It denies some of the most crucial steps in generating and developing a piece of writing. That may be a minor issue in observing whether a writer can produce on demand a formulaic statement in a familiar situation—say, an answer to a routine letter of inquiry—but it may be a major problem in assessing writing as it relates to the main purposes of education. Even less significant as a measure of writing ability, although probably more consistent, are
TROYKA (Continued)

6. Before essay tests are used, the complexities of such tests shall be carefully considered. Most important, topics shall be designed with great care. Also, readers of the essay tests shall be trained according to principles of statistically reliable holistic and/or analytic reading.

7. The nature and purpose of the test and the various uses of the results shall be clearly explained to all instructors and students in each institution prior to the administration of the test.

8. All possible steps shall be taken to educate the universities and colleges, the public and the legislature, that, though composition faculties have principal responsibility for helping students develop writing skills, maintenance of these skills is a responsibility shared by the entire faculty, administration, and the public.

8. The officers and executive committee of CCC shall make testing a major concern in the immediate future in order to provide information and assess the condition of composition instructors affected by a testing situation.

Commentary: We came of age.

A CCC Committee on Testing now actively facilitates the self-education of our profession. At CCC conventions, it sponsors day-long workshops on essay scoring and other practices, sessions for sharing information about local experiences, and seminars on testing theories and applications. This committee has published an annotated bibliography on testing (See CCC, Composition and Communication, December 1979); it has other projects in progress.

Thus, we have moved from condemning to informing. But more is needed. In the 1970's, our profession was reacting. When commercial test publishers with their narrowly focused multiple-choice writing tests grossly distorted the universe of knowledge that defines an educated mind, we criticized. When college administrators decided on local testing programs without actively engaging teaching faculty in all decisions, we protested. In the 1980's, we are ready to act. The National Testing Network in Writing (NTNW) is a fine demonstration of our profession's potential. The teacher-designed and teacher-supervised Writing Skills Assessment Test at the City University of New York, a writing test that requires students to write, is another example of what we can do.

Undergirding this new spirit of action are, I think, four realities. So that our self-education can continue, I propose that faculty who teach college writing examine these realities and consider what sorts of action are implied by each.

- Writing tests in college are here to stay. Placement tests are almost universal now. Proficiency tests to certify students' writing abilities before graduation are becoming increasingly common.
- We must learn the technology of testing. Humanists traditionally distrust numbers, especially any number that purports to summarize a person's knowledge. Yet, once we turn the vocabulary and operations of testing, we can combine the best of the technology with our comprehensive concepts of both process and product in writing.
- The fact that tests influence curriculum puts us in a double bind. On the one hand, we prefer writing tests that ask students to write. On the other hand, we worry about "teaching to the test," which might restrict instruction to only the mode(s) of discourse required by the test. This bind can cripple us. But it is a thin bind, upon analysis. Curriculum is better off when tests ask for error-hunting or rearranging the prose of others. Curriculum is better off when students have to demonstrate their proficiency by generating language, controlling sentence structure, handling cohesion, and sustaining written discussion—all of which is learned best when instruction provides students with opportunities to write frequently in the various modes of discourse.
- A college education is no longer an interaction between students and teachers only. New partners have arrived on the scene. Parents, college administrators and trustees, business people, and legislators are setting expectations that influence budget decisions on the local, state, and federal level. We who are experts in communication have to communicate with these various publics. In our modern, information-based society, literacy for all is a matter of survival, not egalitarian idealism. We need the partnerships of these publics if we are to succeed. We need to be heard at public forums, not just at professional meetings; we need to become part of the popular press, not just for professional journals; we need to listen carefully to our critics, not only our colleagues.

The 1978 CCC "Resolution" can help faculty begin to set goals for their active participation in the testing of writing abilities. The information gathered and disseminated by NTNW can help all who participate learn much of value from one another. But the rest is up to each teacher of college writing on each campus. A look back over the past ten years tells us that we have come far. The decade ahead holds much potential for our continued growth, especially as we become active leaders in shaping the future quality of all tests that seek to assess writing abilities.

Lynn Quitman Troyka is Professor of Writing at Queensborough Community College, CUNY; she is Immediate Past Chair of the Conference on College Composition and Communication, has conducted research in writing assessment, and is author of several books including Steps in Composition.

LLOYD-JONES (Continued)
evaluations based on observation of some limited element of language separated from actual discourse—the substance of "objective" examinations. A large vocabulary, awareness of conventions of middle class dialects, knowledge of the names of grammatical classes, or even the ability to manipulate sentence structures and complete cloze tests may relate to writing skills—more likely to reading skills—and yet fail short in particular cases of telling much about a writer's competence. Since the mastery of language is so fundamental to any academic endeavor, probably anything reporting some linguistic skill will help in predicting academic success without really telling much about competence in writing.

The dangers are not that the tests will let through some people who write badly, for the tests by definition understate competence, but rather that they eliminate people who might thrive. Tests used more modestly as part of a teaching process of measuring mastery so as to indicate what yet needs to be mastered are still a problem, for good writing represents a blend of skills, not an adding up of separate skills. The more the tests focus on discrete items which can be reliably quantified, the greater the distortion in terms of the whole art.

The conclusion to be drawn from these objections is not that all testing be abandoned, but that interpretations of results all be expressed within limits. Popular accounts will doubtless continue to headline simple interpretations but admissions officers, placement committees and researchers should reject unqualified test scores and in any particular case seek corroborative evidence and opinion.

Richard Lloyd-Jones, Director of the School of Letters, University of Iowa, is a co-designer of the PITS method of scoring essays and is past chair of CCC.
TROYKA (Continued)

5. Before essay tests are used, the complexities of such tests shall be carefully considered. Most important, topics shall be designed with great care. Also, readers of the essay tests shall be trained according to principles of statistically reliable holistic and/or analytic reading.

6. The nature and purpose of the test and the various uses of the results shall be clearly explained to all instructors and students in each institution prior to the administration of the test.

7. All possible steps shall be taken to educate the universities and colleges, the public and the legislature, that, though college composition instructors have principal responsibility for helping students develop writing skills, maintenance of these skills is a responsibility shared by the entire faculty, administration, and the public.

8. The officers and executive committee of CCC shall make testing a major concern in the immediate future in order to provide information and assistance to composition instructors affected by a testing situation.

Commentary: We came of age.

A CCC Committee on Testing now actively facilitates the self-education of our profession. At CCC conventions, it sponsors day-long workshops on essay scoring and other practices, sessions for sharing information about local experiences, and seminars on testing theories and applications. This committee has published an annotated bibliography on testing research: "Composition and Communication, December 1979"; it has other projects in progress.

Thus, we have moved from condemning to informing. But more is needed. In the 1970's, our profession was reacting. When commercial test publishers with their narrowly focused multiple-choice writing tests grossly distorted the universe of knowledge that defines our profession, we criticized. When college administrators decided on local testing programs without actively engaging teaching faculty in all decisions, we protested. In the 1960's, we were ready to act: The National Testing Network in Writing (NTNW) is a fine demonstration of our profession's potential. The teacher-designed and teacher-supervised Writing Skills Assessment Test at the City University of New York, a writing test that requires students to write, is another example of what we can do.

Undergirding this new spirit of action are, I think, four realities. So that our self-education can continue, I propose that faculty who teach college writing examine these realities and consider what sorts of action are implied by each.

- Writing tests in college are here to stay. Placement tests are almost universal now. Proficiency tests to certify students' writing abilities before graduation, are becoming increasingly common.

- We must learn the technology of testing. Humanists traditionally distrust numbers, especially any number that purports to summarize a person's knowledge. Yet, once we learn the vocabulary and operations of testing, we can combine the best of the technology with our comprehensive concepts of both process and product in writing.

- The fact that tests influence curriculum puts us in a double bind. On the one hand, we prefer writing tests that ask students to write. On the other hand, we worry about "teaching to the test," which might restrict instruction to only the mode(s) of discourse required by the test. This bind can cripple us. But it is a thin bind, upon analysis. Curriculum is better off when tests ask for error-hunting or rearranging the prose of others. Curriculum is better off when students have to demonstrate their proficiency by generating language, controlling sentence structure, handling cohesion, and sustaining written discussion — all of which is learned best when instruction provides students with opportunities to write frequently in the various modes of discourse.

- A college education is no longer an interaction between students and teachers only. New partners have arrived on the scene. Parents, college administrators and trustees, business people, and legislators are setting expectations that influence budget decisions on the local, state, and federal level. We who are experts in communication have to communicate with these various publics. In our modern, information-based society, literacy for all is a matter of survival, not egalitarian idealism. We need the partnerships of these publics if we are to succeed. We need to be heard at public forums, not just at professional meetings; we need to convert the popular press, not just for professional journals; we need to listen carefully to our critics, not only our colleagues.

The 1978 CCC "Resolution" can help faculty begin to set goals for their active participation in the testing of writing abilities. The information gathered and disseminated by NTNW can help all who participate learn much of value from one another. But the rest is up to each teacher of college writing on each campus. A look back over the past ten years tells us that we have come far. The decade ahead holds much potential for our continued growth, especially as we become active leaders in shaping the future quality of all tests that seek to assess writing abilities.

Lynn Quitman Troyka is Professor of Writing at Queensborough Community College, CUNY; she is Immediate Past Chair of the Conference on College Composition and Communication, has conducted research in writing assessment, and is author of several books including Steps in Composition.

LLOYD-JONES (Continued)
evaluations based on observation of some limited element of language separated from actual discourse—the substance of "objective" examinations. A large vocabulary, awareness of the conventions of middle class dialects, knowledge of the names of grammatical classes, or even the ability to manipulate sentence structures and complete close tests may relate to writing skills—more likely to reading skills—and yet fall short in particular cases of telling much about a writer's competence. Since the mastery of language is so fundamental to any academic endeavor, probably anything reporting some linguistic skill will help in predicting academic success without really telling much about competence in writing.

The dangers are not that the tests will let through some people who write badly, for the tests by definition understate competence, but rather that they eliminate people who might thrive. Tests used more modestly as part of a teaching process of measuring mastery so as to indicate what yet needs to be mastered are still a problem, for good writing represents a blend of skills, not an adding up of separate skills. The more the tests focus on discrete items which can be reliably quantified, the greater the distortion in terms of the whole art.

The conclusion to be drawn from these objections is not that all testing be abandoned, but that interpretations of results be all be expressed within limits. Popular accounts will doubtless continue to see the conventions of scoring essays and is past chair of CCC.

Richard Lloyd-Jones, Director of the School of Letters, University of Iowa, is a co-designer of the PITS method of scoring essays and is past chair of CCC.