

SKEPTICISM 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 prescribed, 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 between such 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 admission and placement exercises cut down the possibility of careful planning and serious revision. 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 single 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

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 cloze 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 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 PTS method of scoring essays and is past chair of CCCC.