Fear and Trembling, Shock and Awe: A Social History of Writing Assessment
Norbert Elliot
Professor of English
New Jersey Institute of Technology
Overview for the ETS Presentation

NJIT

- A World Where Something Was A-Doing, 1874-1917
- Promiscuous Intermingling, 1917-1937
- Irresistible Colored Beads, 1937-1947
- Off the Diving Board, 1947-1966
- Lone Wolves, 1966-2004
- Now and Then: An Analysis
- Lagniappe
A World Where Something Was A-Doing

1874-1917
Charles William Eliot, President of Harvard University, 1869-1907. "The President of the University is Primarily an Executive Officer….The President's first duty is that of supervision." Inaugural Address, October 19, 1869
A short English composition is required, correct in spelling, punctuation, grammar, and expression. Thirty lines will be sufficient. Make at least two paragraphs.

**Subject:**

- The story of the Caskets, in the Merchant of Venice;
- Or, The story of Shakespeare's Tempest;
- Or, The story of Rebecca, in Scott's Ivanhoe.

*English Composition.* (Each candidate will be required to write a short English Composition, correct in spelling, punctuation, grammar, and expression, the subject to be taken from such works of standard authors as shall be announced from time to time. The subject for 1874 will be taken from one of the following works: Shakespeare's Tempest, Julius Caesar, and Merchant of Venice; Goldsmith's Vicar of Wakefield; Scott's Ivanhoe, and Lay of the Last Minstrel. (The Harvard University Catalogue, 1873-74)*
"For he who scrawls ribaldry, just a truly as he who writes for all time, does that most wonderful of things,--gives a material body to some reality which till that moment was immaterial, executes, all unconscious of the power for which divine is not too grand a word, a lasting act of creative imagination.

English Composition: Eight Lecturers Given at the Lowell Institute (1891)
“The preparation of the papers should be in the hands of capable men, under such supervision and so restricted that the absurd travesties now sometimes inflicted upon us would be impossible.”

The Reform of College Entrance Requirements. Inaugural Address (1895)
"The work of the Board will promptly elevate the secondary-school work in English, in history, and in the natural sciences to a new plane of importance and of effectiveness. It will control the examination system in the interest of education, and resist the tendency to make it a mere machine-like performance. It will declare and enforce standards of attainment which represent, not the labors of a zealous individual, however wise, but the mature judgment of a group of mature scholars of different training and points of view." Nicholas Murray Butler, First Annual Report of the Secretary, 1901.
1874-1917

"We conquer the facts of nature when we observe and experiment upon them. When we measure them we have made them our servants." Educational Psychology (1903).
1874-1917

EUGENICS is the science which deals with all influences that improve the inborn qualities of a race; also with those that develop them to the utmost advantage. The improvement of the inborn qualities, or stock, of some one human population will alone be discussed here.

<table>
<thead>
<tr>
<th>Activity</th>
<th>Correlation with Gen. Intell.</th>
<th>Ratio of the common factor to the specific factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classics</td>
<td>0.99</td>
<td>99 to 1</td>
</tr>
<tr>
<td>Common Sense</td>
<td>0.98</td>
<td>96 to 4</td>
</tr>
<tr>
<td>Pitch Dis.</td>
<td>0.94</td>
<td>89 to 11</td>
</tr>
<tr>
<td>French</td>
<td>0.92</td>
<td>84 to 16</td>
</tr>
<tr>
<td>Cleverness,^4</td>
<td>0.90</td>
<td>81 to 19</td>
</tr>
<tr>
<td>English</td>
<td>0.90</td>
<td>81 to 19</td>
</tr>
<tr>
<td>Mathematics,^4</td>
<td>0.86</td>
<td>74 to 26</td>
</tr>
<tr>
<td>Pitch Dis. among the uncultured,^6</td>
<td>0.72</td>
<td>52 to 48</td>
</tr>
<tr>
<td>Music</td>
<td>0.70</td>
<td>49 to 51</td>
</tr>
<tr>
<td>Light Dis.,^5</td>
<td>0.57</td>
<td>32 to 68</td>
</tr>
<tr>
<td>Weight Dis.,^5</td>
<td>0.44</td>
<td>19 to 81</td>
</tr>
</tbody>
</table>
Quality 0: "Dear Sir: I write to say that it ain’t a square deal. Schools is I say they is I went to a school. red and gree green and brown aint it hito but I say he don’t know his business not today nor yeaterday and you know it and I want Jennie to get me out." Thorndike, "Scale for Merit in English Writing by Young People," Journal of Educational Psychology 2 (1911): 367-368.

Quality 95: “If we finally ask what are the limitations of Aristotle, we find none save the limitations of the age and city in which he lived. He lived in a city-state where thirty thousand full male citizens, with some seventy thousand women and children dependent upon them, were supported by the labor of some hundred thousand slaves. The rights of man, whether native or alien, male or female, free or slave, had not yet been affirmed. That crowning proclamation of universal emancipation was reserved for Christianity three centuries and a half later. Without this Christian element no principle of personality is complete. Not until the city-state of Plato and Aristotle is widened to include the humblest man, the lowliest woman, the most defenseless little child, does their doctrine become final and universal. Yet with this single limitation of its range, the form of Aristotle’s teaching is complete and ultimate. Deeper, saner, stronger, wiser statement of the principles of personality the world has never heard.” Thorndike, "Scale for Merit in English Writing by Young People," Journal of Educational Psychology 2 (1911): 367-368.
Promiscuous Intermingling

1917-1937
"Often it seemed that my foremost duty and obligation—one foe which I usually felt myself peculiarly unsuited—was to vanquish seemingly insuperable difficulties by overcoming the passive resistance of ignorance and the active opposition of jealously, misinformation, and honest disagreement."
Devins Literacy Test, Form 1. A score below 6 indicated illiteracy. Scores from 6 to 20 were to be associated with primary literacy. Scores from 21 to 25 were to be associated with a level of literacy appropriate to grammar school. Scores from 26 to 30 were to be associated with junior high school, and from 31 to 35 with senior high school. Scores ranging from 36 to 42 were at a college level of literacy.

"In the case of white recruits a high correlation has been found to exist between psychological rating and military value. The report from Camp Upton [in Yaphank, Long Island] indicates that the same is true in the case of negroes, and that just as they are inferior to the whites in intelligence, so also are they inferior in their value to the military service…The question of the military value of the negro fighter when he gets actually into action may well be quite a separate one and it is at all events one which our present data do not enable us to discuss" (742).
Thomas Scott Fiske: “It should be the purpose of the College Entrance Examination Board not only to ascertain whether the candidates have acquired the information and methods of thought necessary for successful work in college, but also to determine whether they possess certain important intellectual qualities which are sometimes described as alertness, strength, and endurance, although these terms would seem to indicate excellences of the body rather than the mind. The Board's endeavors in this direction are exhibited most conspicuously in the comprehensive examinations. If these examinations continue to respond to the demands of the times they should eventually become the best possible tests combining the necessary elements of an informational and of a mental, or psychological, character.” (Nineteenth Annual Report of the Secretary [New York: College Entrance Examination Board, 1919] 2)
"A movement launched by Francis Galton has eventually grown to encompass every description of man by man."

- The Reading of the Comprehensive Examination in English: An Analysis of the Procedures Followed During the Five Reading Periods from 1929-1933 (Princeton: Princeton UP, 1934)
For all candidates.

6. From one of the following suggestions select a subject and write upon it a composition of about three hundred and fifty words. Follow an orderly, consecutive plan. Prefix an appropriate title. Give the letter of the suggestion followed.

a) Should your school include in the last year of English study more contemporary literature?

b) “Have nothing in your house that you do not know to be useful, or believe to be beautiful.”—William Morris, The Beauty of Life.

c) Highways are flattened walls.

d) When you have a radio and a victrola in your home, is it worth while to learn to play a musical instrument?

e) Drama close-packed with life

f) It is impossible to study any subject without studying English; it is almost equally impossible to study English without studying something else.

g) “Our prevalent notion [about freedom] is that it is a most happy and important thing for a man to be able to do as he likes. On what he is to do when he is free to do as he likes, we do not lay so much stress.”—Matthew Arnold, Doing as One Likes.

h) “Oh, a day in the city-square, there is no such pleasure in life.”

i) Mountain, mist, and valley

j) Waste that we pay for when we buy

k) Working my way through the depression

l) If science has injured mankind, it is not because science has gone too far but because it has not as yet gone far enough.
The inter-reader correlation for the essay was recorded as .49. Brigham concluded that “the factors making for unreliability are now so great that the examination should be discontinued unless progress can be made in getting a single stable score which is descriptive of the candidate.”
Brigham on Conant’s support of consolidation of testing agencies: “Such a suggestion brings to the fore the fundamental purposes of the proposed organization. A new organization solely for the dissemination of present knowledge concerning tests and the promotion of testing programs would be difficult to justify; an organization which advances knowledge, however, is to be desired. It is the writer’s belief that the present testing movement carries the germs of its own destruction and that unless the proposed organization is set up to develop a cure for these afflictions it will retard rather than advance education.”

Irresistible Colored Beads

1937-1947
"If care is taken to formulate essay-test questions in such a way that a restricted form of answer is required and if readers will judge papers on the basis of specific predetermined criteria which trial has shown can be consistently evaluated, there is no reason why reliability of reading cannot be measurably improved"

“The purpose of this bulletin is to describe the methods of preparing, holding, and reading the comprehensive English examination of June 1937. Teachers in the colleges and the secondary schools are rightfully curious about the procedures of the College Entrance Examination Board, and the Board, in turn, is anxious to have its procedures known by all persons interested.”

Directions: Write a composition of about 350 words, entitled “the Ruling Passion.” In your composition develop and illustrate ideas found in the following passage. As illustrations choose three characters from your reading in history or biography, or three from your reading in other literature, or three from contemporary life.

“In every life there is a ruling passion. People are ruled by such passions as the love of honor, money, children, friends, art, or nature; by the love of a man for a woman; or if a woman for a man; by ambition or by pride; by a sense of duty or a sense of loyalty; by the desire for revenge, or the desire to stir up strife. With some people this single passion seems to rule throughout life; with others the rule is shorter lived. Some people recognize that their lives are controlled by a ruling passion; some do not. When circumstances operate to prevent a person from finding expression of, or satisfaction in, his ruling passion, then character emerges . . . .”

—Henry Van Dyke, The Ruling Passion (adapted)
“...an important step in the effort to attain reasonable uniformity of judgments in the answers to the essay question...”
1937-1947

NJIT

Obviously, this test cannot be directly one of literary knowledge or background, although the Examiners and Readers alike would probably agree that thorough training in reading is one of the best fundamental preparations for it. Nor is it a test of ability in so-called creative writing, or of writing the belles-lettres tradition. The artist can scarcely be required to create on demand, and in a brief time. In college, students must constantly express in writing their ideas and the results of their study, not in English courses alone, but in all courses. The English composition test, therefore, should provide candidates with an opportunity to demonstrate their normal skill in written expression, and should measure that skill.

A problem at once arose regarding the form which the test should take. A theme question, for which one hour had been allowed, had traditionally formed Part III of the previous comprehensive English examinations [Question 6]. Hence it was natural for the Examiners, faced with a mandate to provide a one-hour test in composition, to select the theme or essay type of test. Since previous experience had shown the impossibility of grading themes on different topics according to a common standard, one topic only was given”


Directions: The following paragraphs state three criticisms of the radio.

In a theme of from 400 to 500 words discuss each of these criticisms. In your discussion of each criticism you may support or attack it, or adopt a middle ground. In all three cases, however, illustrate by description of a specific program the position you adopt and the point you are making.

Your theme will be graded on

– Material: your use of specific and pertinent evidence;
– Organization: of the theme as a whole and of the separate paragraphs;
– Style: the clarity and accuracy of your writing.

Note: You will be given a special answer booklet in which to write your theme.

Critics of the radio allege that the tendency of radio programs is to lull listeners into a state of unquestioning receptivity, thereby preventing them from using their reasoning powers. The kind of entertainment offered by the radio must make its appeal to thousands; it is therefore pitched at a relatively low level and, proving successful, is monotonously repeated. It is fatally easy to sit back and listen uncritically to such programs. The state of mind encouraged by this attitude is undiscriminating; the taste of the audience becomes standardized.

Further, these critics maintain that under the ceaseless barrage of advertising broadcasts the American way of life is becoming stereotyped. People are coming more and more to eat, drink, wear, buy, believe, think, and live alike; they thus become standardized as consumers of goods as well as of entertainment.

Finally, the critics charge that the picture of life presented by radio fiction in serials or drama is hackneyed and sentimental. Its stock of characters and situations, in its over-simplified solutions for the problems with which it deals, is like bad fiction of any kind, but more dangerous, since it reaches a public larger than the reading public.” (1943)
Stalnaker in 1937: Empirical evidence of consistent reading should “dispel the notion that essay papers can not be read reliably.”

Stalnaker in 1943: “The type of test so highly valued by teachers of English, which requires the candidate to write a theme or essay, is not a worthwhile testing device. Whether or not the writing of essays as a means of teaching writing deserves the place it has in the secondary school classroom may be equally questioned. Eventually, it is hoped, sufficient evidence will be accumulated to outlaw forever the ‘write-a-theme-on . . .’ type of examination.”
Qualifying Test for Civilians: “...........is an opinion held in opposition to the established ........:: (1) belief .. doctrine (2) stubbornness .. doctrine (3) heresy .. doctrine (4) heresy .. government (5) fascism .. government”

Qualifying Test for Civilians: WATER is to sponge as INK is to ............ (1) pen (2) bottle (3) write (4) blotter (5) desk
My ‘predictions’ are not probably ‘valid.’ But everywhere I go, I find increasing interest and new enterprises directed toward the tailoring of education to fit the individual. And I think that this increasing emphasis on the individual student, in the fact of growing numbers of students, must mean that testing will play a larger and larger role in college admissions and guidance programs. Not, perhaps, testing as we have known it in the past—testing that measures only isolated hues or qualities of a student’s talents—but testing which displays the full spectrum of his personality. Keeping pace with the developments in education, no less a development must take place in testing itself. Henry Chauncey, “The College Entrance Examination Board: Origins and Current Trends,” The College Board Review 1.5 (1948) 67.
Off the Diving Board

1947-1966
“Since the President’s announcement, there has been a great deal of editorial comment on the program, much of it sharply critical. So far, ETS has pretty well escaped criticism, which has been directed toward the policies of Selective Service rather than toward the test or toward ETS as the organization responsible for its administration. Happily, the first outburst of negative criticism regarding the program seems to have been followed by expressions of opinion which have been on the whole somewhat more favorable. The college deferment program was vulnerable because of the apparent inequality of treatment of individuals in and out of college. And since in general students in college are from families of higher incomes, the plan seems to be undemocratic in favoring the sons of the rich.” Memorandum for the Board of Trustees, 20 April 1951, Chauncey Merger File (Princeton: ETS Archives) 1.
Edith Huddleston (seated fourth from right, with William E. Coffman on her left). "The investigation," Huddleston concluded in her 1954 study," points to the conclusion that in the light of present knowledge, measurable 'ability to write' is no more than verbal ability."

Earl G. Eley (standing at left, discusses the General Composition Test with 58 readers at Hun School, Princeton, before scoring 4,122 essays) "Too often, we have achieved precision in the measurement of human behavior by mechanizing the elements of that behavior. Such a procedure is effective only part of the time: the creative or dynamic aspects of the human personality resist atomization. They must be assessed as wholes or not at all."

"Each paper is read by at least two readers. The essays are read quickly for an over-all impression. You and the other readers are asked to react immediately to a paper and to score it on the basis of that first impression. You are asked not to reread papers or to analyze specific points of view. The basic assumptions of the holistic reading are that each of the factors involved in writing skill is related to all the others and that no one factor can be separated from the others."

To Edward Noyes, who wrote the Introduction to The Measurement of Writing Ability, it must have seemed that the end to a very long road had been reached. The measurement of a student’s ability to write, he proclaimed, “has at long last been solved, insofar as the current English Composition Tests are concerned.”
Lone Wolves

1966-2004
"As a test of writing ability, no test is as convincing to teachers of English, to teachers in other departments, to prospective employers, and to the public as actual samples of each student's writing, especially if the writing is done under test conditions in which one can be sure that each sample is the student's own unaided work."

(From Measuring Growth in English [Urbana, Ill: NCTE, 1974] 1)
“If a measure of measurement scheme is reliable, it is fair to writers, permitting them to demonstrate what they can really do. For purposes of a quick, tentative screening or placement decision, we need only ask a writer for one piece produced under the best conditions we can arrange. However, from previous research we know that either a comprehensive description of writing performance or a reliable rank ordering of a group of writers can be achieved only by asking for more than one piece of writing on more than one occasion and then involving two or more people in describing or rating each piece. Even then, we have reliable measurement of writing ability in only one mode. To test ability to write in different modes (personal narrative, explanation, dramatic dialogue, reportage, prose fiction) or to write for different purposes (explanation, persuasion, expression), we need multiple pieces on multiple occasions. As an example, we know from previous research that syntactic patterns vary from mode to mode, and further, that the best writers display most variation.” Charles R. Cooper and Lee Odell, eds., Evaluating Writing: Describing, Measuring, Judging (Urbana, Il: NCTE, 1977) xi.
“American students are writing no better in 1984 than they were ten years earlier,” Applebee and his colleagues report, and “performance of these 9-13- and 17-year-old-students is distressingly poor.”


What can tests tell, or what can they not tell, teachers or administrators about students’ writing problems?

What should a teacher—or parent, school administrator, school board member, or legislator—know about testing—especially about the implications and limitations of test results?

What have been the effects of local and centralized minimum competency programs on a) test designs, b) procedures, c) curriculum, and d) writing program development?

What are the most effective ways to respond to a mandated test of minimum competency in writing?

On the basis of your experience, what mistakes or pitfalls can those interested in testing writing avoid in the future?

What are the major concerns that the classroom teacher should have about a college-wide testing effort, and how should the classroom teacher participate in test development and use?

Based on your conversations and/or research, how do students perceive writing tests—their purpose and use?

What are the unresolved issues in testing writing? In what direction should the field be heading? What research should be conducted?

What should a teacher—or parent, school administrator, school member, or legislator—read to be the most informed or current issues in testing writing?

"The modern era of writing assessment began during the fall of 1971 when the English departments of the California State University (CSU) system, then consisting of 19 separate campuses, were able to defeat an administrative move to institute an external multiple-choice test for first-year English equivalency for the entire system." Edward M. White, Comparison and Contrast: The 1973 California State University and Colleges Freshman English Equivalency Examination ED 114825; College English 63.3 (2001): 306-320.

A. To provide a comprehensive measure of writing ability that would allow secondary school students to demonstrate a wide range of writing experiences.

To provide a focus for the improvement of secondary school writing curricula by demonstrating the value of writing in assessment, by formulating a set of writing tasks commonly recognized among teachers of writing, and by enhancing the professional development of teachers of writing; to provide the instructional and descriptive materials necessary to the integration of the portfolio into existing writing programs and a forum for discussion of writing and writing instruction in which can be developed 1) common standards for the evaluation of writing, 2) awareness of strengths and weaknesses exhibited by student writers and 3) a context for curriculum improvement.

To facilitate, eventually, the transition from secondary to postsecondary institutions by providing information less subject to distortion than that provided by the current application process and by improving communication between secondary and postsecondary institutions.

“We had no real, focused or concentrated space to initiate, continue, or contribute to an ongoing dialogue about writing assessment. This lack of space was made all the more poignant by our experience at Miami University because we were pleasantly aware of the wealth and depth of scholarship going on just about portfolios. At one of the social activities during the evening, the talk turned to the obvious lack of space for talking about writing assessment on a regular basis. There were many voices in that conversation. I remember Pat Belanoff’s, Peter Elbow’s, and Liz Hamp-Lyons’ (all members of the editorial board) as prominent in the discussion that eventually prompted Kathy Yancey and me to exchange e-mail addresses for the purpose of establishing a journal in writing assessment.”

1994: E-Rater

- Surface features
- Structural features
- Rhetorical structure
- Content analysis
- Argumentative ability

Think carefully about the issue presented in the following quotations and the assignment below:

1. While secrecy can be destructive, some of it is indispensable in human lives. Some control over secrecy and openness is needed in order to protect identity. Such control may be needed to guard privacy, intimacy, and friendship.

   Adapted from Sillela Bok, "The Need for Secrecy"

2. Secrecy and a free, democratic government, President Harry Truman once said, don't mix. An open exchange of information is vital to the kind of informed citizenry essential to healthy democracy.

   Editorial, "Overzealous Secrecy Threatens Democracy"

Assignment: Do people need to keep secrets or is secrecy harmful. Plan and write an essay in which you develop your point of view on this issue. Support your position with reasoning and examples taken from your reading, studies, experience, or observation.

Now and Then: An Analysis

- Enrollment Trends
- The Ontology of Accountability
- Epistemology: Knowing About Writing Ability
- Validity: 1929-1999
  - Validity in Writing Assessment: From Scales to Portfolios
- Reliability: 1929-1999
  - Reliability in Writing Assessment: From Specimens to Portfolios
- (Re)Articulating Writing Assessment
Historiography: The Dangers of Hegel, the Seduction of Judgment

Perspective: Documenting the Scenes

The Nature of Association

The Impulse for Accountability: The Rhetoric of Disenfranchisement

The Construction of Literacy: Implementing Variance

And So....