# The Phenomenon of Impact:
# The CUNY Writing Assessment Test

Lynn Quitman Troyka

In "Do You Agree or Disagree: The Epistemology of the CUNY Writing Assessment Test," Judith Fishman objects to the test's topics, which end with "Do you agree or disagree?" She misses the point, I think. Test topics can be changed. Given the proper research rigor, investigation of alternate topics for an in-place test is simple to undertake. I will discuss the CUNY writing topics presently, but first I want to set straight a record I think Fishman misunderstands.

The Writing Skills Assessment Test (WAT) created and used at CUNY has had an impact far more positive than Fishman implies by her statement "In short, the system 'works.'" The WAT is process as much as product. Its existence is both symbol and fact. Moreover, the quality of the CUNY Freshman Skills Assessment Program, of which the WAT is part, surely must have influenced FIPSE in its recent decision to award a three-year, $363,000 grant to the National Testing Network on Writing, now part of the CUNY Office of Academic Affairs, for a project on "College Assessment Program Evaluation."

The phenomenon of impact is real. Sure, CUNY deserves to be proud of the excellent consistency with which it runs a complex, multi-campus assessment program, but the stakes are far higher than that. The WAT is raising consciousness in the CUNY feeder high schools about the need for a curriculum that stresses actual writing; it is alerting CUNY students and faculty to the centrality of writing as a means of learning; it is nudging CUNY curricula toward greater emphasis on large amounts of frequent writing; it is making the public aware that writing is important. At the core of any testing program lie what Stephen Witte and Lester Faigley call "unintended effects" (62).

I

I was among the faculty back in 1977 who thought it could not be done. I did not think CUNY, with its yearly incoming freshman class of about 33,000 students in the fall and 12,000 students in the spring,[1] could manage to assess students' writing abilities with an essay test. I did not think that a sufficient number of topics could be properly developed and kept secure; that CUNY English faculty would agree to take on the added

task of scoring then of thousands of essays each year; that so huge a collection of essays could be read reliably across-CUNY at each of the separate 17 CUNY campuses; and that the testing schedules would work out easily for the students. I was wrong.

I was delighted to be wrong. Who could object to learning from the likes of Mina Shaughnessy, who chose the Task Force members charged to select a writing test suitable to the CUNY Board of Trustees' mandate for a check on skills? from the likes of Ken Bruffee, who chaired that Task Force? from the likes of Marie Jean Lederman, Robert Lyons, and Harvey Wiener, who were among the members of that Task Force? Most of all, who could object to discovering that in-house proficiency assessment with essays was possible on a colossal scale?

"First-generation concerns," as I call them, occupied that first Task Force on Writing in 1977. The group, comprised of only highly experienced teachers of writing,[2] met approximately 24 times during six or seven months of what must have been exhausting but exhilarating activity. The Task Force's first major decision proved to be landmark: to use an essay test exclusively.[3] They had three reasons, I'm told:

1. multiple-choice tests, with their attention to error hunting and revising others' prose, do not represent the universe of knowledge called "writing ability,"

2. reinforcing the need for as much actual writing as possible has a salubrious effect on teaching and curricula, and

3. an assessment combining multiple-choice test and essay is too vulnerable to budget crunches which can too easily eliminate the expensive-to-administer essay test in favor of the inexpensive machine-scorable test.

Of course, the Task Force had to develop topics. Preferring topics of appeal to the unique student body attending CUNY, the Task Force wanted topics that were timely and rooted in the real world. They wanted the tasks to be straight-forward and concrete, and thus reasonably accessible to the students. The Task Force decided on a choice of topics for each test, because one topic seemed too limiting and more than two topics seemed intimidating within the 50-minute constraint. Finally, the Task Force wanted a format that might help to propel the students to "get started." The closing words of the WAT prompts, "agree or disagree," were spawned by that concern: to give the students a gentle push into the flow. (Fishman asks why the earliest tests read "agree or disagree with this statement" instead of the current "agree or disagree." Simply this: The early WAT forms, produced under the pressure, had a vague referent for "this." Embarrassing, but true.) A typical WAT test with the

complete prompt and two sample topics no longer part of the "secure" pool of approximately 50 CUNY topics appears in Appendix I. (All topic-statements are more fully developed than Fishman was able to show, due to test security, in her opening list.) Topic development, by the way, goes on yearly. All faculty get a memo inviting them to submit new topics. Then a sub-group of the Task Force on Writing develops, pilots, and finally selects the best of the new ideas. I know not if Fishman has ever submitted topic suggestions to the CUNY committee, but I do know that every idea is welcome because of ongoing need for fresh topics.

"Second-generation concerns" occupied the Task Force about the time I began to serve in 1980.[4] All was in place, and it was time to take a hard look at what we had. Our major priority was research. For example, we instituted a yearly central audit of randomly selected papers from all campuses so that we could assure that standards for the test were being applied consistenly. More important, the Office of Academic Affairs freely offered research specialists to Task Force members and any other interested faculty who would write proposals. We were urged to ask hard questions. I wish that Fishman had come forward. She would have liked what we were doing, I think. She would have found, I know, a forum for sharing her ideas and the means to be guided into some very interesting research. In the early 1980's, four research monographs on the WAT were published by the Office of Academic Affairs.[5]

I had the privilege of conducting one of the studies (*An A Posteriori Examiniation*) to review the WAT descriptive criteria, written *a priori* by the first Task Force. We wanted to see if the criteria might be made fuller and more performance-based, thereby helping students and test scores. The project, involving the collaboration of over 100 CUNY and non-CUNY faculty, was complex and costly for the Office of Academic Affairs. As I worked, I found a spirit of honest inquiry without a trace of "why bother if things are going well?" And when refined criteria were worked out, the administration vigorously undertook to disseminate the new information. Those refined criteria, given in Appendix II, do not depend in the slightest on an "agree or disagree" format, just as the first set of criteria did not.

Fishman raises what I would call a "third-generation concern." Can the topics be improved? I cannot imagine why not. Nothing is perfect. Although the existence of the WAT seems assured, the particulars of the WAT have never been cast in concrete. The spirit of the entire WAT enterprise has been colleagues consulting with colleagues to improve what they can offer to students.

## II

Despite the potential of Fishman's concern, I am troubled by some of the separate points she makes. Her discussion and suggestions seem to me to suffer from inexperience in reading about, or conducting, unbiased rsearch. Let's take the story of the Radcliffe student. If one of the topics troubled her, why did she not choose the alternative topic? How typical is she? The WAT has been shown to measure well the ability of talented students. For example, two years ago a selective upstate New York college, a college not unlike Radcliffe in its rigorous standards, asked CUNY to include in an annual WAT audit a few hundred of their college's initial and re-test essays, without being singled out to the scorers. In that select group, 85% passed the initial test and 90% passed the re-test.

If we want to go on anecdotal evidence, which I might enjoy talking about but cannot put much stock in for decisions affecting large nmbers of highly diverse students, I can bear witness to excellent student reaction to the WAT. Until the WAT was in place, the placement test at the CUNY college where I taught was a multiple-choice test with laughable face validity but bulls-eye accuracy for sorting students into the various levels of composition we offered. During those years, I spent much of my time trying to justify the test to angry students who correctly saw no connection between their actual writing and the nit-picking items on the placement test. Once the WAT was in place, the tension ended. In fact, I always urged my students to look up their placement essays after about six weeks of class. On returning from the college's testing office where the essays are kept on file, dozens of students each semester would bounce into class chattering about the fact that now they could "see" and evaluate their own writing more clearly. They were beginning to take control of their own educations.

Fishman's suggested alternate topic illustrates the important difference between a classroom topic given as exercise and a university-wide topic. I wince at the idea of telling an uncle how I am going to spend a million dollars because the end of the world is near. The possibility of the end of the world is no joke to today's students who, according to surveys and studies, sincerely and deeply fear a nuclear holocaust. What kind of values are suggested by telling the students to think about going on a spending spree and chattering about it to an uncle? What about the materialism, even nihilism implied by that so-called "dire situation"? Most of the students I know think about crisis in terms of yearning for the comfort of family, friends and, in some cases, religion.

Fishman is again exclusively anecdotal when she reports opinions about the WAT held by some of her colleagues at her CUNY college. She does not also refer her readers to the widely circulated research report by Karen Greenberg, *CUNY Writing Faculty: Practices and Perceptions* (1983).

Greenberg presents the results of a scientifically developed anonymous survey of CUNY writing faculty, which asks opinions about the WAT, among other matters. Of the 433 faculty from all CUNY colleges who responded, 75% thought that the WAT was as effective as, or more effective than, the placement system used previously by the individual college. Only 15% of the faculty said that they "taught to the test," and half of this group consider the WAT a benefit to their curricula. One teacher said:

> I teach directly to the test—I give frequent, full-length, in-class essay assignments. Because of this we spend more time on writing rather than on grammar or mechanics. Also, my students are improving their ability [sic] to provide logical reasons for assertions thanks to WAT practice. (47)

Well, I did not use WAT-type practice in my own teaching, but I admit that were I forced to choose between a teacher who uses grammar drills to teach writing and one who uses essay practice, even in a WAT format, to teach writing, I would choose the latter. I would not much like the choice, but I would make it knowing that grammar drill is a not uncommon focus in too many writing classes.

Off and on, I have heard about teachers who "teach to the test." I wonder about those teachers. I wonder what they were doing when the test was a multiple-choice test. And is drill in WAT-type topics all they can think of doing? I wonder what kind of writers they are. And readers. But then, in all fairness, I began to wonder about a possible explanation for some of our teachers' behavior: many CUNY colleges require, on their own, that re-tests be given after only a quarter or a semester of college work. That is an error, in my opinion. No test, essay or multiple-choice, can be statistically sensitive enough to measure growth over the short haul. Writing ability grows with exposure to good teaching over an extended period of time. The Office of Academic Affairs mandates *only* that the WAT be taken for placement and, if failed, be taken again before the completion of 60 credits. No central mandate requires re-testing after one course in basic writing. Some CUNY colleges, however, have made that decision on their own. Colleges that re-test students after only ten to fifteen weeks of life in college are bound to be disappointed with the results, and the teachers whose final grades are determined by one test are bound to be frustrated.

## III

The research I have done over the years has always turned out to be more difficult and fascinating than I had anticipated. I would like to offer, therefore, a few basic guidelines for Writing Program Administrators and other English faculty who wish to undertake thorough, unbiased topic research for the purpose of placement or proficiency assessment with essays.

1. If the test is going to be used for determining requirements that affect the time and money a student must spend in college, the test has to hold up to close scrutiny, perhaps even in the courts.

2. An essay test can have a wide variety of topics, but each one has had to be put through a rigorous screening process.
    a. A "jury" of faculty has to meet, read and discuss the topics. It must eliminate any topics that "on the face of it" (face validity) do not seem appropriate.
    b. Each topic has to be administered in a test-like situation to a group of students as much like the ones who will take the test as possible. The students should be from as varied backgrounds as the ones who will take the test (within reason, of course). The minimum number of pilot tests per topic is 300 for a testing program the size of CUNY's.
    c. To end up with two topics that can be used, pilot ten. (This is conservative; professional testing companies—which write test topics for national audiences—expect a much higher rate of eventual rejection, but if a college knows itself, the 2:10 ratio will probably work).
    d. The pilot tests have to be read by a group of faculty, preferably some of those on the original "jury" and some new people, to check for ambiguity of topic and to see if *the topic is biased against any one subgroup of the pilot test group.* For example, a topic about rock music obviously would be biased against people who do not listen to it; other biases are less obvious and often surprising on reading pilot test essays.

3. The topics have to be "equivalent," but not exactly alike. This means that no student is penalized for selecting a topic more difficult than another. This factor has to hold "over time," which means that once a test is in place, new topics have to be of equivalent difficulty. A student taking the test in 1984 cannot have a benefit (due to topic variance) that was not available to someone who took the test in 1978. This matter is less restricting than it might seem. A good consultant in research design knows how to set up comparative studies so that statistical equivalence can be objectively measured.

4. The topics should not be random but should have an undergirding rationale. But the rationales should not control the test unduly. The students have to be able to do their best under time pressure (assuming there is a time limit as exists at CUNY, with its tests in writing, reading, and math). For example, some faculty today might be attracted to topics that make explicit audience, purpose and other details of the rhetorical situation. Fine, but check it out with pilot testing. James Hoetker in "Essay Examination Topics and Students' Writing" reminds us of Brossell's work which "concludes that too detailed a specification of rhetorical context 'can be counterproductive to writers by inducing them to repeat needlessly the information in the topic and thus waste time.'" (388)

Carefully collected data are, after all, the best assurance we have that our decisions will not be derived from one controlling personality or philosophy.

## Notes

[1] Some of Judith Fishman's facts are a bit off. She says that 33,000 accounts for "all incoming freshmen," but that accounts for only the fall semester; the total for an academic year is over 43,000. She says that "the inception of the WAT" was in 1976, but she is early by two years. 1976 was the year the mandate for skills assessment was given by the CUNY Board of Trustees; in 1977, Task Forces in the various skill areas met; the Writing Assessment Test (WAT) was instituted in the spring of 1978 to be binding on students entering the fall 1978 semester.

[2] To the credit of the CUNY administration, especially during those pressured months of quick decision-making about mandated tests, only faculty were appointed to decide on matters that would long affect CUNY faculty and curricula.

[3] As I was researching WAT history, I was impressed to discover, that the Task Force on Writing was *not* the first CUNY committee to recommend an essay test for writing assessment. During 1976 and early 1977, the Chancellor's Committee on Movement from Lower to Upper Division, a cross-curriculum group of faculty, deliberated about all of CUNY's skills-assessment needs. In March 1977 their final report recommended an essay for writing assessment. The Task Force on Writing, formed soon after, had that recommendation at hand—and it was only a recommendation, not a mandate—when they began their work. (For the full text of the recommendation, see "Freshman Skills Assessment: A Proposal," Office of the Vice-Chancellor for Academic Affairs, The City University of New York, August 1977).

[4]Membership on the Writing Task Force rotates regularly, to keep ideas about test improvement fresh, and to ensure representation from all 17 CUNY colleges. The chair is an experienced member of the Task Force who also serves on a CUNY Policy Advisory Board which coordinates the entire Freshman Skills Assessment Program in writing, reading, and math.

[5]The complete list of these monographs and research monographs on other topics including computers in the English class (along with cost information— around $2.00 each) can be obtained from the Instructional Resource Center, CUNY, 535 East 80th Street, New York, New York 10021.

## Works Cited

Greenberg, Karen. *CUNY Writing Faculty: Practices and Perceptions.* New York: The City University of New York, 1983.

Hoetker, James. "Essay Examination Topics and Students' Writing." *College Composition and Communication* 33 (1982): 377-92.

Troyka, Lynn Quitman. *An A Posteriori Examination of the Evaluation Scale of the Writing Skills Assessment Test of the City University of New York.* New York: The City University of New York, 1982.

Witte, Stephen, and Lester Faigley. *Evaluating College Writing Programs.* Carbondale: Southern Illinois University Press, 1983.

# Appendix I

**The City University of New York**
**Writing Skills Assessment Test**

*Directions*

You will have fifty minutes to plan and write the essay assigned below. (You may wish to use your fifty minutes in the following way: 10 minutes planning what you are going to write; 30 minutes writing; 10 minutes rereading and correcting what you have written.)

You should express your thoughts clearly and organize your ideas so that they will make sense to a reader. Correct grammar and sentence structure are important.

Write your essay on the lined pages of your booklet. You may use the inside of the front cover of the booklet for preliminary notes.

You must write your essay on **one** of the following assignments. Read each one carefully and then choose either **A** or **B**.

---

**A.** It always strikes me as a terrible shame to see young people spending so much of their time staring at television. If we could unplug all the TV sets in America, our children would grow up to be healthier, better educated, and more independent human beings.

Do you agree or disagree? Explain and illustrate your answer from your own experience, your observations of others, or your reading.

**B.** Older people bring to their work a lifetime of knowledge and experience. They should not be forced to retire, even if keeping them on the job cuts down on the opportunites for young people to find work.

Do you agree or disagree? Explain and illustrate your answer from your own experience, your observations of others, or your reading.

# Appendix II

**The City University of New York**
**Freshman Skills Assessment Program**

*Writing Skills Assessment Test Evaluation Scale* (refined 1981-82)

**6**
The essay provides a well-organized response to the topic and maintains a central focus. The ideas are expressed in appropriate language. A sense of pattern of development is present from begining to end. The writer supports assertions with explanation or illustration, and the vocabulary is well suited to the context. Sentences reflect a command of syntax within the ordinary range of standard written English. Grammar, punctuation, and spelling are almost always correct.

**5**
The essay provides an organized response to the topic. The ideas are expressed in clear language most of the time. The writer develops ideas and generally signals relationships within and between paragraphs. The writer uses vocabulary that is appropriate for the essay topic and avoids oversimplifications or distortions. Sentences generally are correct grammatically, although some errors may be present when sentence structure is particularly complex. With few exceptions, grammar, punctuation, and spelling are correct.

**4**

The essay shows a basic understanding of the demands of essay organization, although there might be occasional digressions. The development of ideas is sometimes incomplete or rudimentary, but a basic logical structure can be discerned. Vocabulary generally is appropriate for the essay topic but at times is oversimplified. Sentences reflect a sufficient command of standard written English to ensure reasonable clarity of expression. Common forms of agreement and grammatical inflection are usually, although not always, correct. The writer generally demonstrates through punctuation an understanding of the boundaries of the sentences. The writer spells common words, except perhaps so-called "demons," with a reasonable degree of accuracy.

**3**

The essay provides a response to the topic but generally has no overall pattern of organization. Ideas are often repeated or undeveloped, although occasionally a paragraph within the essay does have some structure. The writer uses informal language occasionally and records conversational speech when appropriate written prose is needed. Vocabulary often is limited. The writer generally does not signal relationships within and between paragraphs. Syntax is often rudimentary and lacking in variety. The essay has recurrent grammatical problems, or because of an extremely narrow range of syntactic choices, only occasional grammatical problems appear. The writer does not demonstrate a firm understanding of the boundaries of the sentence. The writer occasionally misspells common words of the language.

**2**

The essay begins with a response to the topic but does not develop that response. Ideas are repeated frequently, or are presented randomly, or both. The writer uses informal language frequently and does little more than record conversational speech. Words are often misused, and vocabulary is limited. Syntax is often tangled and is not sufficiently stable to ensure reasonable clarity of expression. Errors in grammar, punctuation, and spelling occur often.

**1**

The essay suffers from general incoherence and has no discernible pattern of organization. It displays a high frequency of error in the regular features of standard written English. Lapses in punctuation, spelling, and grammar often frustrate the reader. Or, the essay is so brief that any reasonably accurate judgment of the writer's competence is impossible.