

**Testimony before the Commission on No Child Left Behind/The Aspen Institute**

**Dr. Betty J. Sternberg**  
**Connecticut Commissioner of Education**  
**May 9, 2006**

First, let me thank you for inviting me to testify to this Commission. I am honored by your invitation.

I believe that I am in a unique position to comment on today's topic – assessment – because, while I have been Connecticut's education commissioner since November 2003, I have more than two decades of front-line experience with student assessment. I co-directed the development of the Connecticut Mastery Test (CMT), which began testing students in Grades 4, 6 and 8 in reading, writing and mathematics in 1984, and the Connecticut Academic Performance Test (CAPT), which began testing Grade 10 students in the same subjects plus science in 1994.

As you know, this year the full testing requirements of the federal No Child Left Behind (NCLB) Act have begun. Nationwide, approximately *45 million tests* will be given to students in Grades 3-8 and one high school grade (Thomas Toch, *Margins of Error: The Education Testing Industry in the No Child Left Behind Era*, 2006). Roughly 6.4 billion test pages – 38.4 billion test items in reading and mathematics – will be produced and graded by a handful of testing companies.

In Connecticut, approximately 315,000 students spent from 8 to 8.5 hours, depending on grade level, taking tests spread out over many days in March. For example, third grade students – children between the ages of 7.5 and 9 – worked through three

booklets: a 47-page booklet for reading, a 59-page booklet for math and a 29-page booklet for writing. Direct instruction in test-taking strategies, anxiety-prevention coaching, morale building and the like preceded the actual administration of the tests. Moreover, throughout the school year before administration of the tests, schools focused instruction heavily in the areas of reading, mathematics and writing – giving students problems like the ones they would see in those test booklets in March. Many teachers and parents criticized the minimizing of instruction in the visual arts, music, theater, dance, history, civics, science, health and safety, physical education and world languages because of these activities.

For those charged with administering the tests, it meant receiving boxes upon boxes of test booklets for students – 87 boxes in Region 10, a small district testing about 1,600 students in two towns. The shrink-wrapped skid of boxes, delivered “a little late,” almost collapsed coming through the door. That logistical problem resolved, the test coordinator (who is the director of curriculum and instruction) and his secretary then spent two full days preparing the booklets: placing labels with students’ names and other information on the covers, sorting into school and classroom sets and so on. After delivery to schools in two towns, there were numerous meetings with teachers to review test protocol. This year the testing coordinator made *triple* the number of daily calls during testing to the State Department of Education to resolve problems such as, “This student used a ruler and shouldn’t have; do we discount his test?” “This student threw up on the test, should we copy his answers onto another answer sheet or put it in a plastic bag and send it to the test company?” Because of the flu and the increased number of students taking the test, the make-up period for testing individual students who had

missed portions of the test disrupted the regular school routine weeks beyond the regular testing period. Post-test work included collecting, collating and ensuring absolute security until the testing company's delivery service arrived to pick up the test booklets. This year, pick-up for Region 10 was two weeks late.

Imagine what this must have been like in a large district – Bridgeport, for example, which serves more than 22,000 students, nearly 10 times as many as Region 10. It received 673 boxes of booklets in order to test approximately 13,000 students at 43 schools or program sites this past March.

Now the critical question: What will this massive, costly testing – costly not only in money, but also in time and effort – tell Connecticut educators that they don't already know from the amount of testing done prior to NCLB? Nothing.

The results of tests already given by Connecticut and other states are consistent across the nation. In general, wealthier students do better than poorer students, nonminority students do better than minority students, students without disabilities do better than students with disabilities, and there are significant gaps among urban, suburban and rural students. Performance between male and female students varies, too, depending on the subject.

So, what needs to be done to close these unacceptable gaps in achievement? We need to establish a system of education reform that incorporates accountability and provides to all children programs we know, based on research, will work to diminish these gaps. While I have specific suggestions about programs that should be offered to all students to diminish the gaps, I will focus only on accountability and assessment, because these are the topics of today's hearing.

An effective accountability system uses the strengths of a reasonable amount of the type of standardized, summative testing required by NCLB. The primary purpose of such tests is to show how *groups* of youngsters perform and how effective *programs* of instruction are, identifying where the flaws and issues in these programs might be. For example, if only 27 percent of fourth graders appear to be mastering the estimation objectives, the district must ask: (1) Is estimation of the type tested included in our curriculum prior to fourth grade? (2) If it *is* in the written curriculum, are teachers devoting enough time to teaching it? (3) If it *is* being taught, what instructional techniques are being used? (4) How can we modify our instructional techniques to increase our students' achievement of the estimation objectives? Such testing is important and useful if employed in this way.

The tests required by NCLB are not, however, useful as tools to shape instruction for *individual* students. They serve only as signals to do further diagnosis of a student's skills if there appears to be a large discrepancy between what the teacher already knows about the student's skills and the results of the test. In fact, it would be a misuse of such tests to use them as sole indicators to place students in particular programs. After all, these test results are provided months after students have taken the tests, when they should have mastered new material and, more often than not, have moved on to the next grade in school. Thus, a high-quality assessment system also incorporates the use of common formative tests – short tests that are given every four to six weeks and provide immediate feedback to teachers and students. Teachers use this feedback to “customize” instruction for each child throughout the year.

Considerable research shows that formative tests are not only effective in raising the achievement of all students, but are particularly effective in raising the achievement of lower-performing students. *This closes achievement gaps*. Paul Black and Dylan Wiliam, writing in the October 1998 *Phi Delta Kappan* (“Inside the Black Box: Raising Standards Through Classroom Assessment”), said formative testing is “at the heart of effective teaching” and “helps low achievers more than other students and so reduces the range of achievement while raising achievement overall.”

*There is no such body of evidence regarding the standardized, annual, summative tests states are being required to administer beginning this year under NCLB.*

An examination of our schools that have been most successful in raising students’ achievement significantly – schools such as Amistad Academy, a charter middle school in New Haven, and Dwight, a Hartford elementary school – shows that these schools have developed their own set of formative tests and use them regularly to modify each student’s instruction based on the results of these tests.

While using both annual, summative tests in alternate grades and formative tests throughout the grades, a robust accountability system also incorporates open-ended response items that require critical thinking and articulate response – rather than tests that are limited *solely* to multiple-choice items that are easier, faster and cheaper to score, but don’t tell us enough about how our students think. Open-ended items require students to demonstrate deep understanding and represent more of what they will encounter beyond the classroom.

Assessment is useful only if what we test is important and challenging. Connecticut has been testing students in Grades 4, 6 and 8 since 1984 and Grade 10 since

1994 – and we have publicly reported the results by subgroup: race/ethnicity, economic status and one characteristic NCLB does not require, gender. Our Connecticut Mastery Test (CMT) in the three earlier grades and Connecticut Academic Performance Test (CAPT) in the tenth grade are widely considered “the gold standard” across the nation. In great part, this is because we *do* test what is important and challenging, including a separate writing assessment that requires students to produce first-draft essays in response to a prompt. Even in our math and reading tests, about one-third of the items are open-ended questions that require students to explain in writing how they got their answers.

Let me share with you an example of an open-ended mathematics item and a student response that received the highest score of 3 (both released from the 2005 administration of the Grade 10 CAPT). The problem is called “Walk to the Stadium,” and it tests knowledge – and how to use it – in the areas of algebra and functions. Please note what I believe are the most important words on the page that presents the problem: “Remember to show your work and write your answer in your answer book.” (Please see Attachment 1.)

Another example from the 2005 CAPT – the Interdisciplinary Writing Test, represented here by the “Overview,” “Your Task,” “Preparing to Write Your Letter” and excerpts from an example of student work that received the highest score, 6 – also is instructive. (Please see Attachment 2.) NCLB requires a third academic indicator in addition to reading and math tests, and Connecticut has chosen to continue its testing of students’ writing skills. We believe writing is fundamental to success in all other learning and in life; good, sound writing requires good, sound thinking. That is why we have

tested writing, requiring students to produce an essay or letter, since the very beginning in 1984. The writing test is costly and time consuming to develop, administer and score, but writing skills are so important that we would never consider dropping this test from our assessment system. Multiple choice has its place, but using this format alone on tests cannot approach Connecticut's "gold standard" of assessment.

Finally, a robust system of accountability would use the technology currently available in the real world as part of the assessment system. Such a system would, for example, have youngsters write the answers to their prompts using a word processing tool, just as I did when I wrote this testimony. A robust accountability system enables students to use technology as a tool for answering questions, and it would also use currently available technology to score students' responses – even those that are in open-ended, written formats. This would require a major investment in development and in hardware for students, but would result in economies in the end. This would make test taking a more authentic experience and bring our students into the 21<sup>st</sup> century.

The annual testing of every student in seven grades – administered as it is under the current practice and requirements of NCLB – has caused an unprecedented crisis that has begun to unfold. Testing-company insiders will admit privately that this industry is begging for qualified personnel – with firms "stealing" staffers from each other and from state departments of education. Before NCLB was enacted, it appears that no one ever looked nationwide and asked if those few companies had the resources to generate accurately results for 45 million tests in a few months' time. Only recently, after much-publicized errors in major testing programs, has the secretary of education raised the issue with company presidents.

I predicted these problems in an article I wrote for Education Week in June 2004; a recent report, *Margins of Error: The Education Testing Industry in the No Child Left Behind Era*, written by Thomas Toch and published by Education Sector Reports, confirms that they have happened. “The mounting scoring errors and reporting delays that have resulted from the many challenges confronting the testing industry and state testing agencies as they struggle to respond to NCLB’s testing mandates,” Toch wrote, “have tarnished NCLB’s testing-based system of school accountability.” The recent stories about misreporting the scores of more than 4,400 students who took the SAT in October 2005 suggest widespread problems even in the most established testing programs.

Connecticut, too, has been affected by these problems. We had delays in reporting CMT scores two years ago with CTB/McGraw Hill and, recently, incorrect reporting of 355 scores (and failure to report 12) on the CAPT with Harcourt Assessment, Inc. Very recently, we have been informed by Harcourt that “issues” with its “scanning/network server have forced us to interrupt processing of documents for the [2006] CAPT program.” The “worst case” presented to Connecticut would have the state failing to meet NCLB-mandated reporting deadlines, which is clearly unacceptable. I have no doubt that Harcourt’s scanner/network server problems must be affecting all other states with which it has testing contracts.

How many more problems will Connecticut and other states experience due to a small number of testing companies overwhelmed by NCLB requirements? And how will our students be affected? My greatest fear is that this situation will result in a nationwide push toward tests made up of only multiple-choice questions – cheaply, easily and

quickly scored by machine, but unable to tell us enough about what our students know and are able to do with that knowledge – tests that don't test what is important and challenging. This is exactly what the United States Department of Education has suggested Connecticut do in Grades 3, 5 and 7 in order to save money – that, and eliminate the writing test that has been fundamental to our assessment system for more than two decades.

Less large-scale, summative testing as currently required by NCLB, but better and more frequent formative testing – testing that works to *improve* student achievement, not just record it – would allow us to better focus on closing the unacceptably large differences in achievement among subgroups of children.

NCLB, with its heavy emphasis on academic achievement measured by annual, summative tests alone, will have our nation's children meeting minimal targets – busily filling out those 6.4 billion test pages. I *do not* believe this will result in the kind of educated adult we all envision. I *do* believe that this nation wants its public school graduates to have not only the knowledge critical to the essential disciplines – from reading, writing and mathematics to science, history, world languages and the arts – but also the human qualities that make a good friend, a good neighbor, a good colleague, a good citizen.

Claiborne Pell, former U.S. Senator from Rhode Island and father of the federal Pell Grants that have helped so many students attend college, said, “The strength of the United States is not the gold at Fort Knox or the weapons of mass destruction that we have, but the sum total of the education and the character of our people.” The sum total of our children's education must be more than 38.4 billion multiple-choice test items – no

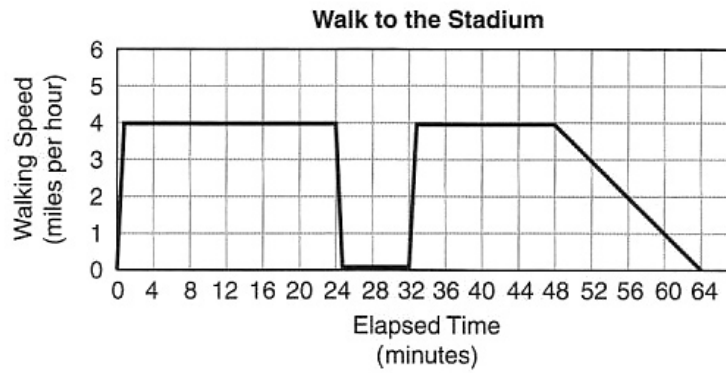
matter how many they get right. The sum total of our children's education must foster in our students the character of which Sen. Pell spoke. I see character as the qualities of responsibility and integrity; the willingness to work hard and to persist despite challenges; intellectual curiosity and independent thought; imagination and creativity; respect for oneself and others; strength and kindness; and a commitment to being active, constructive members of one's community, state, nation and world. The sum total of our children's education must help create an academically astute, responsible, caring, compassionate and wise citizenry. To succeed in that, we *must* fix the assessment requirements of NCLB.

# ATTACHMENT 1

## *Walk to the Stadium* Algebra & Functions

### 6. Walk to the Stadium

Ernest walked to the stadium to watch a football game. The graph below illustrates his walking speed on the trip.



- About how many miles did Ernest walk to the stadium?
- Based on the information in the graph, describe what may be happening during his walk to the stadium. Be sure to include a description of each change shown in the graph.

**Remember to show your work and write your answer in your answer booklet.**

---

# ATTACHMENT 1

## Scored Student Responses for *Walk to the Stadium*

Question 6: Walk to the Stadium

Score 3

a. 3.2 miles

Show Calculations

$$\left(\frac{24}{60} + \frac{16}{60}\right)4 + 2\left(\frac{16}{60}\right) = 3.2$$

time spent walking 4mph      time spent slowing down: avg speed

b. From 0-24 minutes he walks normally. 24-32 rests his feet. 32-48 walks normally. 48-64 reaches the crowd outside game which gets thicker and slower the closer he gets to the game.

This student correctly calculates the miles walked and has an adequate description of events, resulting in a complete and comprehensive response. The student has demonstrated a full and complete understanding of the task.

## ATTACHMENT 2

### CAPT Interdisciplinary Writing Test Session 1

#### Online Learning

#### Overview

The purpose of this Interdisciplinary Writing test is to determine how well you can write to persuade others to think as you do about an issue. In this test, you will read three short articles about an important issue, take a position on the issue and write a first draft of a persuasive letter. You must support your position with information from *each* of the source materials. Your response will be read and scored by trained readers.

#### About This Test

In this Interdisciplinary Writing test, you will think about and take a position on an important issue: **online courses for high school students**. While you are working on this test, you will use skills and knowledge you learned in your language arts, mathematics, science, social studies, the arts and other classes.

#### The Issue

Online courses have been offered to college students for a number of years. Now high schools are starting to become involved in the process. Proponents argue that online offerings allow students in small districts to have access to classes that the school wouldn't be able to afford or staff. Opponents counter that online courses cannot be equivalent to the experience of having the interaction of a teacher and classroom of students.

# ATTACHMENT 2

## Your Task

You will read three articles about online courses, take a position on the issue and write a persuasive letter to your local school board. In your letter you must support or oppose a proposed plan to offer online courses in your high school. Your letter must include information from *each* of the source materials.

Below are the steps you will follow:

**1. Read the source materials.**

Before taking a position on the issue, it is important that you read all the source materials related to online courses. (Take a minute to locate the source materials beginning on page 13 in this booklet.) As you read the source materials, you may underline information, highlight or write notes on the articles themselves.

**2. Prepare to write your letter.**

You have been given two organizers which you may find useful as you consider the various arguments for and against the proposed plan to offer online courses in your high school. You may use the organizer that best suits your position. In addition, scratch paper has been included for any additional notes or outlining you may wish to do in preparing to write your letter. The organizers and scratch paper can be found on pages 19–21. **You are not required to complete these pages, and they will not be scored.**

**3. Write your letter in the answer booklet.**

Write a letter to your local school board either supporting or opposing a proposed plan to offer online courses in your high school. You must include information from **each** source.

**Write your response in your answer booklet. Only what is written in the answer booklet will be scored.** Three pages have been provided for your response. Any additional pages will be removed prior to scoring.

# ATTACHMENT 2

## Your Score

Your score will be based on the following criteria:

1. **Position**—Did you take a clear position on the issue?
2. **Comprehensiveness**—Did you use information from *each* of the three sources that are provided?
3. **Support**—Did you support your position with accurate and relevant information from the source materials?
4. **Organization**—Did you organize your ideas in a logical and effective manner so that your audience will understand and follow your thinking?
5. **Clarity and Fluency**—Did you express your ideas clearly and fluently using your own words?

## Organizing Your Time

You will have 65 minutes to complete this test. The following schedule is suggested:

30 minutes for reading the source materials and planning your letter; and

35 minutes for writing your letter.

**You may re-read or refer to the source materials at any time during the test. You may also highlight or make notes on the sources if this is helpful to you.**

Know your purpose for writing and remember to:

READ, THINK, PLAN, WRITE.

# ATTACHMENT 2

## Preparing to Write Your Letter

Arguments *FOR* a plan to offer online courses in your high school.

Based on your reading of the source materials, list below the most important arguments, or points of view, presented to **support** a plan to offer online courses in your high school. Also list the evidence or claims which support each argument.

Arguments For a Plan to Offer Online Courses in Your High School	Supporting Evidence or Claims

**This page will not be scored**

# ATTACHMENT 2

## Preparing to Write Your Letter

Arguments *AGAINST* a plan to offer online courses in your high school.

Based on your reading of the source materials, list below the most important arguments, or points of view, presented to **oppose** a plan to offer online courses in your high school. Also list the evidence or claims which support each argument.

Arguments Against a Plan to Offer Online Courses in Your High School	Supporting Evidence or Claims

**This page will not be scored**

# Scored Student Responses for Interdisciplinary Writing Session 1

## This is an example of score 6:

(What follows are the first and last paragraphs of a handwritten, three-page persuasive letter that received the highest score of 6 on the 2005 Connecticut Academic Performance Test [CAPT] Interdisciplinary Writing Test. All writing samples are scored holistically. The original work has been retyped without changes.)

Dear Board of Education,

The groan heard on a weekday morning is very frequent at the homes of many high school students. A solution to this daily ritual is virtual learning. Although being face-to-face with an actual teacher and classmates provides that classroom feeling, virtual learning allows students to take extra courses, prepare for the future and it allows students to be comfortable in their space.

...

Although a regular classroom gives a nonsynthetic feeling, virtual learning is an excellent, up and coming idea to let students get courses they otherwise could not receive, prepare for the future and feel comfortable. As many know, a classroom is not the place for everyone and this may be the key to the future for many students. Virtual learning is not for everyone, but it may be able to change how some kids learn which will, in the end only be beneficial.

Sincerely,