

**Public Hearing
Regents Learning Standards
and High School Graduation Requirements**

**NY State Standing Committee on Education
Albany, NY, October 22, 2003**

**Testimony submitted by Dr. H. Evan Powderly
President, Westchester/Putnam, Middle School Principals Association**

Middle schools in New York State presently are in a Dickensian period. It is the best of times for middle-level education, and it is the worst of times for middle schools. The best of times is indicated by the recently adopted New York State Board of Regents Policy Statement on Middle-Level Education. In the Policy Statement the Board of Regents affirm their belief "that the time students spend in the middle grades, 5 through 8, is critical to both their personal growth and development and their success in high school." The Policy Statement identifies "seven essential elements that must be in place in standards-focused schools with middle grades if young adolescents are to succeed academically and develop as individuals." The Policy Statement has been embraced by middle-level educators across the State because the Statement reflects widely accepted middle-level theory and philosophy written over the years in documents like Turning Points, espoused by middle-level scholars like the late William Alexander, and endorsed by the National Middle School Association.

It is the worst of times for middle schools in New York state because we have been mentioned in various media as "the wasteland of our primary and secondary landscape, "a crack in the middle" and "the muddle in the middle." Middle schools are labeled with derogatory terms owing to our students results on various New York State standardized assessments. In a recent publication of the New York State School Board Association it was reported: "The Regents have identified middle school performance as the Achilles heel of the state's education system. The most recent scores on the eighth-grade English Language Arts (ELA) exam illustrate the problem. The percentage of students meeting all the standards (Levels 3 and 4) increased by only one percentage point from the previous year to 45 percent. The results are down from 1999, when 48 percent met all the standards." Statements like these, and the validity of the New York State Assessments bring me before this committee today on behalf of the Middle School Principals Association of Westchester and Putnam counties, and all the middle schoolers, parents, and teachers we serve.

In the best of times for middle schools we, middle school principals, want what the Regents Policy Statement envisions "an educational program that is comprehensive, challenging, purposeful, integrated, relevant, and standards-based" resulting in meaningful achievement; but we fear even worse times, when we will be limited to preparing students to look good on the New York State Assessments by reducing or eliminating time students spend in music, art, social studies, and physical education, so

that students can have more prep time for test taking. A recent study examined data from 18 states that have implemented high stakes testing programs to assess whether students gained any knowledge that they could apply elsewhere, other than learning the necessary facts for performing on a state's high stakes test (Amrein & Berliner, 2002). They concluded that, "if the intended goal of high stakes testing policy is to increase student learning, then the policy is not working." While a state's high stakes test may show increased scores, there is little support in these data that such increases are anything but the result of test preparation and/or the exclusion of students from the test process (Amrein & Berliner, 2002, p.2).

We, Middle School Principals, believe, as stated by the National Forum to Accelerate Middle-Grades Reform, "in standards and assessments that lead to high expectations, foster high quality instruction, and support higher levels of learning for every student. At the same time, we believe that no single test should ever be the sole determinant of a young adolescent's academic future, whether it be promotion to the next grade, special placement, or transition from middle grades to high school." We believe the Grade Eight Language Arts Assessment of 2003 is an invalid measure of the achievement of eighth graders, thereby, erroneously assigning students to academic intervention placements and limiting high school course selection. These limitations imposed on students destroy an adolescent's self-esteem and promote thoughts of dropping out.

I am certain during these sessions you have heard from many who have told you the New York State Assessments are "ill-advised, badly administered, and of poor quality." You may have heard that there is "little time to teach anything not on a test" and "that state tests make educators teach in ways that contradict their own ideas of good educational practice." You may have heard "the tests will not motivate previously unmotivated students;" "the tests make everyone anxious, students, parents, and teachers;" and "overall, the benefits of state-mandated testing programs are not worth the investment of time and money." You may have heard, as middle school principals have argued since 1999, that there are "too many tests in grade eight with too much emphasis on the results of these one-time exams." I am certain you have heard of the severe consequences for failing New York State Assessments and Regents exams. I am not here today to speak to you about consequences, but truth. The truth in testing, and the validity of the Grade Eight New York State Assessments. It is not necessary to reiterate for you the debacle of the recent New York State Assessments, including the 2003 Physics Regents Exam and Regents Math A. The tragedy resulted in a committee review of the challenged exams, rescoring of some examinations and a delay on raising high school graduation requirements. The troubles with the New York State tests is not limited to the recent Regents exams, but include problems with the scaling, equating, and scoring of fourth and eight grade English Language Arts (ELA) and Mathematics tests. Educators in the field are aware of the irregular cut-score (setting of the passing score) of the grade four ELA Assessment and problems of the grade eight Mathematics Assessment including the heavy reliance on reading skills, ambiguous language in directions, and inadequate time to complete the multiple choice portion of tests. The Commissioner of Education considers the problems with New York State Assessments as a mere "stumble

from time to time" on the accountability path. We consider these problems akin to a fall from a narrow winding road running along a steep slope. The result of that fall has caused severe injury to our students. To illustrate the lack of truth or validity in the New York State Assessments, we will use our experience with the State Education Department regarding our concerns about the cut scores of the Grade Eight ELA Assessments administered in January 2003.

In March 2003 middle schools throughout the state received preliminary results of the Grade Eight ELA Assessments administered in January 2003. Soon after the arrival of the results, I began receiving telephone calls from many Westchester middle school principals about the scores. Many districts reported a dramatic drop in the number of three's and four's on the exam which are the performance levels needed to meet proficiency standards. After the results became official in May, 2003, we learned 32 of the 50 districts in Westchester County reported lower or the same scores as the year before. Of 40 middle schools in Ulster, Sullivan, and Orange counties 22 lost ground compared to the previous year.

During the Commissioner of Education's press conference releasing the scores, he expressed deep disappointment in the results of the eighth-grade tests, and promised changes at the middle school level. The Commissioner also reported "the mean scores decreased more than expected in Rural, Average, and Low Need Districts." Did this mean he expected our scores to decrease? Would he use the ELA scores to accelerate proposed changes in middle-level education to include reduction in other subject requirements, such as technology, music, and art?

On behalf of the middle school principals of Westchester and Putnam counties, I carried on a continual dialogue with members of the State Education Department about our concerns. I expressed alarm that after aligning our curriculum with the New York State English Language Arts Standards; and after providing academic intervention services to many students over the past few years; and even after reluctantly adding test preparation to our instructional practices; exemplary middle schools in Westchester and Putnam counties had dramatic reduction in the number of students demonstrating proficiency on the January 2003 ELA Assessment. I told officials at the State Education Department that we did not believe the ELA Assessment results reflected the true achievement of our students, and we believed an error had been made in the calculation of the cut scores. We questioned the methodology used by CTB/McGraw Hill, the commercial testing firm that constructs and scores New York State Assessments, personnel to adjust the cut scores to compensate for moving the test administration date from March to January.

During the three months of dialogue officials at the State Education Department (SED) were firm that no errors had been made and appropriate statistical methodology had been used to establish cut scores. They further refuted our assertions that scaling and equating methodology was flawed; and that there was an unfair scoring emphasis on the multiple choice section of the ELA Assessment. SED officials ignored our charges that a low Grade Four ELA Assessment cut score, combined with a 55 passing score on the

high school English Regents exam and inaccurate Grade Eight ELA scores, resulted in unfair comparisons between elementary, middle, and high schools in New York State.

Frustrated with SED responses to our concerns about the validity of the Grade Eight ELA Assessment, in June 2003 the Westchester/Putnam Middle School Principals Association contracted with Dr. Richard M. Wolf, retired Chairman, Department of Measurement, Evaluation and Statistics, Teachers College, Columbia University, to examine the Grade Eight ELA Assessment scoring procedures. In a report to the middle school principals, Dr. Wolf concluded, "the Grade Eight ELA exam has real problems that can only be corrected with a great deal of additional work." In a presentation to the middle school principals he stated, "I was appalled by what has been done." Dr. Wolf believes the standard setting was wrong, and arbitrary; and that the Bookmarking standard setting procedure conducted on February 24 to 25, 1999, was not representative of a random sample of teachers and was froth with social pressure.

Dr. Wolf identified real problems with the transformation of raw scores to scaled scores because the methodology was never described in materials given to middle school principals by SED officials. Dr. Wolf determined there was a big problem when the cut scores were reset to the middle of the year from June. He finds the explanation by SED officials and CTB/McGraw Hill researchers, "a lot of bluster amounting to nothing more than the extrapolation with very, very high proficiency cut scores." He conjectures the cut scores "keeps the pressure on middle schools."

Dr. Wolf's findings are consistent with other current research that reports "the high error rate that has been experienced in virtually every major testing program and state in the nation in which high stakes tests are used calls to question the validity of such tests" (French, 2003). In a monograph, "Errors in Standardized Tests: A Systemic Problem", Rhoades and Madaus cite numerous errors in educational testing. "Beginning in 1999, Ms. Rhoades and Professor Madaus conducted a systematic search for reports of testing errors and found more than 100 in the United States, Britain and Canada from 1976 to 2002, a period that saw extraordinary growth in school testing. One major testing company, for example, had its revenues rise more than tenfold during those years" (NY Times.com, Article: Rising Demands for Testing Push Limits of its Accuracy). From 1999 to 2003 Rhoades and Madaus found dozens of errors discovered by school officials, teachers, parents and students. Testing companies had not discovered most of them. The number of test errors in the monograph are too numerous to detail today. I would cite one example of a documented testing error made by CBT/McGraw Hill, the commercial testing firm that produces and scores the ELA Assessments, that had disastrous effects on New York students.

In 1999 educators in Indiana and Tennessee reported a sharp drop in average percentile scores on the Terra Nova test. After weeks of wrangling with CTB McGraw Hill, rerunning of scores, and auditing the data, it was discovered that a programming error caused rankings on the Terra Nova to be too low at the lower end of the scale and too high at the upper end. As a result, approximately a quarter of a million students in six states were given the wrong national percentile scores. In addition to Tennessee and

Indiana, the error also corrupted scores in New York City, Wisconsin, Nevada, and South Carolina. The students and staff in New York City and three Nevada schools were among the most seriously affected because officials used the scores to make high-stakes decisions.

To understand the impact of the error in New York City requires a review of decisions made in the months before the error was found. In early 1999 the Terra Nova was used for the first time in New York City to measure achievement in reading and math. News of unexpectedly low scores hit the press that May. New York State Education Commissioner Richard Mills suggested that these new lower scores accurately reflected poor student performance and recommended that low scoring students attend summer school that year. Thirty-five thousand New York City students attended school that summer, and thousands of others were notified that they were to repeat a grade because they did not comply with the summer school attendance requirement.

On September 15, 1999 days after school started, McGraw Hill admitted that the same error that affected Tennessee scores had also incorrectly lowered New York City students' percentile ranks at the lower end of the scale. Thus, 8668 children whose correct scores were above the cutoff had mistakenly been compelled to go to summer school. The irony in this story is that the Terra Nova scores had actually risen substantially over those of the year before.

In Nevada, state officials used percentile rankings on the Terra Nova to identify "failing" schools, which received state funds for improvements. Recalculation of the scores showed that three schools were erroneously cited as "failing." The negative publicity that accompanied inclusion on such a list could not be undone. Following discovery of the error on the Terra Nova, CTB McGraw Hill posted a warning on their website. "No single test can ascertain whether all educational goals are being met" - a strong caution against using test scores in isolation for making high-stakes decisions. (Rhoades & Madaus, March 2003)

Rhoades and Madaus also cite examples of errors in cut-score setting and standard equating procedures similar to the mistakes middle school principals allege occurred with the Grade Eight ELA Assessments of 2003. The decision that underlie the formation of cut scores and passing scores are largely subjective. Glass (1977) pointed out that the idea of objectively setting cut scores that accurately differentiate between students who know and students who don't know is largely a fantasy - there is no mathematical or logical support for such an idea in the realm of education testing. He wrote, "If ever there was a psychological educational concept ill-prepared for mathematical treatment, it is the idea of setting cut scores. (Rhoades and Madaus, March 2003).

Given the multiple errors in standardized educational testing we, the middle school principals of Westchester and Putnam counties, have identified; and given the affects of testing on our students' lives; we ask the New York State Assembly and Senate appoint a special independent panel to investigate the validity and usefulness of high stakes standardized tests in New York State and further we recommend the Assembly and

Senate consider creating an independent oversight panel that audits the processes and products of testing agencies to assure the consumer that the tests are "technically sound and appropriate for its stated purpose" (National Commission on Testing and Public Policy, 1991). We are certain that the panel will recommend less state testing and more flexibility for local school districts to demonstrate student proficiency with multiple measures. We hope as the Board of Regents deliberates regulations to implement the newly adopted Policy Statement on Middle-Level Education, they will ponder the relationship between standardized tests and their vision of middle-level education, and consider alternative forms of assessment that honor this vision.

I would like to conclude with the words of a recent graduating valedictorian from another state:

"So I'm the Valedictorian. Number one. But, what separates me from number two, three, four, five, six, 50, or 120? Nothing but meaningless numbers. All these randomly assigned numbers reflect nothing about the true character of an individual. They say nothing ... about desire or will. Nothing about values or morals. Nothing about intelligence. Nothing about creativity. Nothing about heart. Numbers cannot and will not ever be able to tell you who a person really is. Yet, in today's society we are sadly becoming more and more number oriented. Schools today are being forced to teach to the numbers....the state tests serve as just another set of meaningless numbers that add one more reason to focus on scores and forget learning....Judging us by our competency on a biased test in perhaps the biggest injustice that the state could ever inflict upon us... Does anyone care about the human beings behind the numbers?" (Schantz, 2000)