
**Pennsylvania Association of Elementary
and Secondary School Principals**

White Paper
on
High Stakes Testing



September 2008

High Stakes Testing

WHITE PAPER

Written and prepared for PAESSP

by

Joseph S. Yarworth, Ph.D.

and

Margaret M. Place

Professors

Albright College

Department of Education

Reading, Pennsylvania

**Pennsylvania Association of Elementary
and Secondary School Principals**

122 Valley Road

P.O. Box 39

Summerdale, PA 17093

(717) 732-4999 (Telephone)

(717) 732-4890 (Fax)

www.paessp.org

Copyright © 2008 by PAESSP

PAESSP's Mission

To ensure a quality education
for every child by
comprehensively supporting
the educational leaders of our schools.

PAESSP's Vision

To be the best service organization
of its kind in the United States.



Selection of the High Stakes Testing Task Force

Participants were selected to guarantee a representative sample from large, small, urban and rural school districts. In addition, an equal number from the elementary, middle and secondary level were selected. Consideration was also given to include representation from various geographical areas of the state and those with a background in testing.

The PAESSP membership was contacted by a broadcast e-mail requesting volunteers to serve on this task force. As is usually the case, many more members volunteered to be part of the task force on high stakes testing than could be accommodated.

The members of the task force were charged with doing research, reading literature provided on the topic, attending a Webinar and a meeting in Harrisburg and assisting with the editing of this position paper.

It was gratifying to the PAESSP Board of Directors and staff that these members provided their assistance and expertise in this important issue.

PAESSP will consider additional task forces in the future for other important issues as they develop.

High Stakes Task Force Members

Dr. Joseph Yarworth
Margaret Place

Facilitator, Albright College
Facilitator, Albright College

Dr. John Colantonio
Dr. Michael Young
Demetrius Thermenos
Tonya DeVecchis-Kerr
Dr. Russell Eppinger
Dr. Philip Domencic
Kim Olszewski
Dr. Alan Fauth
Andrea Coleman-Hill
Erica Kolat
Dr. Linda S. Mohler
Dr. Susan Evans
JoAnne Fox
Dr. Richard Pysch
Antoinette Miranda
Tim Meckey
Mark McElroy
Eric Heasley
Stephanie Turnbull
Vito Quaglia
Gwendolyn Damiano
Dr. Melissa Patschke
Mary Jo Walsh
Michael Pawlik
John Wilson

Clarion University of Pennsylvania
Conestoga Valley School District
Brandywine Heights School District
Mt. Union Area School District
Middletown Area School District
Cornwall-Lebanon School District
Erie School District
Dallastown Area School District
Governor Mifflin School District
Chartiers Houston School District
Mifflin County School District
Rose Tree Media School District
Upper Merion Area School District
Pine-Richland School District
Armstrong School District
Clearfield Area School District
Delaware Valley School District
A. W. Beattie Career Center
Pittsburgh School District
Wyoming Area School District
Scranton School District
Spring Ford Area School District
FELL Charter School
Greater Nanticoke Area School District
Altoona Area School District

PAESSP Staff Support

Joseph Acri
Sheri Thompson
Laurie MacAskill
Mary Snyder
Melissa Plantz

Assistant Executive Director
Director of Communications
Admin. Assistant/Receptionist
Office Manager
Membership Coordinator

Pennsylvania Association of Elementary and Secondary School Principals

White Paper on High Stakes Testing

On March 18, 2008, the Pennsylvania Association of Elementary and Secondary School Principals (PAESSP) assembled a task force of members to discuss and to recommend a policy position for the Association on the topic of "High Stakes Testing."

The timeliness of the task force's meeting was critical given the fact that two events were driving discussions of the topic at both the national and state level.

At the national level, the renewal of *No Child Left Behind* was becoming embedded in the primary election contests for both the Democratic and Republican parties. At the state level, a proposal to require *Graduation Competency Assessments (GCAs)* was on the front burner of the Pennsylvania State Board of Education at the request of Gov. Edward Rendell.

The PAESSP had previously published its recommendations for changes to No Child Left Behind in July 2007 and at that time had also started the process to establish a task force on "High Stakes Testing," almost six months before the current controversy on GCAs emerged at the state level.

When the task force met in March 2008, the session began with a presentation on GCAs and the current status of "High Stakes Testing" in Pennsylvania.

Following the introductory presentations, the task force held a general session to discuss the concept of "High Stakes Testing" in the public schools in Pennsylvania. After the conclusion of the general session, working groups for elementary, junior high school/ middle school and high school were formed to provide the principals' representatives with an opportunity to discuss the impact of "High Stakes Testing" at the building and divisional level.

This white paper presents a summary of the major points of that day's discussion as well as the concerns and recommendations of the task force members regarding "High Stakes Testing."

High Stakes Testing: The Concept

The concept of "High Stakes Testing" rests in a notion that consequences follow the taking of an examination; these consequences may be positive or negative.

For generations of students prior to the passage in 2002 of the federal No Child Left Behind (NCLB) Act, "High Stakes Testing" had an impact on few students. The most common meeting ground for public school students and the testing universe came with the Scholastic Aptitude Test (SAT), which has been used by many colleges and universities to determine the composition of their incoming freshman classes.

With the passage at the federal level of Public Law 94-142 (1975), testing spread to a new population of students within the public schools. The consequence of this type of testing was the placement of children in appropriate classes to address their educational needs identified through the referral and testing process.

In reality, other than the two aforementioned groups, the only “High Stakes Test” taken by most public school students was the test given by the Commonwealth’s State Police in conjunction with the state’s Department of Transportation to determine which students received a license to drive on the public highways.

Low Stakes Testing: Pennsylvania’s “Educational Quality Assessment”

To the credit of the Commonwealth of Pennsylvania, it began testing on a statewide basis in the early 1970s. Following the adoption of 10 Quality Goals of Education by the Pennsylvania State Board of Education, the Commonwealth initiated a statewide testing program, *Educational Quality Assessment (EQA)*, to determine to what degree schools within the state were meeting the state’s goals.

This testing process could be described as “Low Stakes Testing.” Data were reported on a building basis: no individual student scores were published, and there were no comparisons to other districts throughout the state.

The EQA testing system was rather sophisticated. It identified 35 variables within the research knowledge base that had been shown to correlate with achievement measures covered by the test. The statistical technique of stepwise multiple regression was used by the state to provide each school with a projected score range. Building scores fell within, above or below those ranges.

EQA tested all of the state’s goals, which included areas of “Self-Esteem,” “Understanding Others,” “Creative Activities” and “Societal Responsibility” in addition to the tests of “Reading Comprehension,” “Writing Skills” and “Mathematics.”

Pennsylvania continued to use the EQA testing system from the early 1970s through the end of the 1980s.

Medium Stakes Testing: Pennsylvania’s “TELLS Assessment”

Following the publication of *“A Nation At Risk”* by the federal government (1983), the Commonwealth of Pennsylvania embarked on its second generation of testing, *Testing for Essential Learning and Literacy Skills (TELLS)*, with a focus on the areas of reading and mathematics.

In this testing system, which was conducted through the 1980s, a movement to “Medium Stakes Testing” began. Individual student reports were given to school districts. On each test a basic “cut score” was established. If the student scored at or above that score, he/she “passed” the test. If the student scored below the “cut score,” the student was required to enter remedial study as a consequence of having “failed” the test. The student remained in remediation until exited by the district based on achievement of a level of proficiency established by the district.

The state published the passing percentage of students in each building in which students were tested; the state provided money to districts for remediation based on the number of student failures in each testing period.

Following the administration of the 1989 TELLs tests, the Commonwealth published a “matched pairs” study of children’s achievement from grade 3 to grade 5 and from grade 5 to grade 8.

For reading, a sample of the results were as follows: 72.1% of the children passed the test in both 3rd and 5th grades; 13.1% of the children failed in both years; only 4.8% of the children who failed in grade 3 passed in grade 5, while 10% of the children who passed in grade 3 went on to fail in grade 5. The sample size of the study numbered 84,055.

It is interesting to note that in 1987, after years of testing on a limited number of objectives and with specific funding for remediation, “proficiency” (passing) levels of grade 8 only reached 76% in reading and 75% in mathematics.

Unlike EQA, which was still available to districts, TELLS did not test the state goals in areas beyond the basic skills.

High Stakes Testing: NCLB and “PSSA”

Although the Commonwealth of Pennsylvania introduced the *Pennsylvania System of School Assessment (PSSA)* almost 10 years before the federal government adopted No Child Left Behind, the current version of the PSSA is used by Pennsylvania for its compliance with the NCLB regulations.

The data produced by the PSSA tests classify children into the categories of “Advanced,” “Proficient,” “Basic” and “Below Basic” with the goal that all children in the Pennsylvania schools will reach proficiency by 2014.

In order to meet full compliance with federal regulations, test data are disaggregated into groups for poverty, special education, ethnicity and English Language Learners.

As reported in its *“White Paper on the No Child Left Behind Reauthorization”* (2007), PAESSP supported the state’s efforts to move to a growth-based assessment model, which would allow the recognition of student and school improvement based on the entry level of each student. The Association opposed, and continues to oppose, pre-set required achievement levels for meeting *Adequate Yearly Progress (AYP)* without regard to the entry-level skills and knowledge children bring into the school setting, especially English Language Learners.

NCLB and PSSA moved children into the realm of “High Stakes Testing” by requiring that all students who did not reach proficiency on the PSSA receive remedial instruction (without the state aid provided under TELLS). It also required that students at the high school level who did not reach proficiency on the 11th grade measure be evaluated by a locally constructed and administered assessment system in order to graduate from high school.

Despite the changes in PSSA since the early 1990s, all of which were designed to bring the tests into a closer alignment with the state standards, two basic questions remain regarding the PSSA:

1. Are there sufficient items on the test for each standard tested to warrant the description of the PSSA as a “criterion-referenced” measure?
2. As the state changes the content measured by each test each year, how are the schools expected to hit a continually moving target?

High Stakes Testing: Graduation Competency Assessments

The battle over “High Stakes Testing” reached new levels of conflict in 2008 when the State Board of Education moved to adopt changes to its Chapter 4 regulations on “Academic Standards and Assessment,” which would require students who did not reach proficiency on the 11th grade or 12th grade retest of the PSSA to take a series of six Graduation Competency Assessments (GCAs) in English, math, science and social studies in order to receive a high school diploma.

PAESSP opposed the changes to Chapter 4 for the following reasons:

1. GCAs required additional state testing with the loss of instructional time in the classroom.
2. GCAs took the control of who graduated from high school from the local school district and its teachers and placed the decision-making process into the hands of test makers.
3. GCAs circumvented the process of local assessment required under the current version of Chapter 4.
4. GCAs forced the passing of a test or series of tests to be the final qualification for the receipt of the high school diploma.
5. The cost to develop, not to mention administer and score, the GCAs was estimated at \$50 million.
6. School districts, under the proposed changes to Chapter 4, would have been required to provide remediation without state funds to students who failed the GCAs.
7. The only local option available to school districts not wishing to use the GCAs was an expensive and difficult process required to validate the local tests psychometrically in order to insure that they appropriately assessed the state standards and were equally challenging as the PSSAs.

In July 2008, the Governor and the legislature reached a compromise on the issue that would authorize the state to develop graduation tests that school districts could use voluntarily in 2009-2010.

High Stakes Testing: Value-Added Assessment

With a view to implementing a growth-based model to evaluate AYP, Pennsylvania has developed the *Pennsylvania Value-Added Assessment System (PVAAS)* for all 501 school districts.

PVAAS provides feedback to three key questions:

1. Did each cohort of students (grades 4-8) make a year’s worth of growth in reading and math?
2. Did each NCLB subgroup (grades 4-8) make a year’s worth of growth in reading and math?
3. Is each individual student on a trajectory to reach proficient or advanced levels on a future PSSA?

Although PVAAS provides a sophisticated tool to evaluate measures of student achievement, it also requires that the data used in the system be accurate beyond any measure of doubt in setting the baseline from which “growth” and “change” will be measured; otherwise the test data may be misinterpreted by the public.

PAESSP’s Task Force Concerns on High Stakes Testing

The initial session of the March 2008 task force meeting focused on the principals’ concerns regarding the impact of “High Stakes Testing” within the Commonwealth. The major concerns were:

- The scheduling of the testing and the timing of the return of the scores do not help current classroom teachers utilize the data given in the PSSA reports.
- Curricular areas in the minor subjects, e.g. art and music, have been curtailed to spend time on test preparation
- A disconnect exists between assessment and accountability.
- The tests have had a negative emotional impact on the children, children crying as the test is given.
- There is no state curriculum, yet there is a state assessment.
- “Proficiency” is determined by a single measure at a given point in time; multiple methods of assessment should be used to measure proficiency.
- “Proficiency” has become a statistical concept, not a curricular concept.
- The content of the PSSA changes annually; it is a moving target. How does a school know when it hits the bull’s eye?
- The content of the PSSA may not be consistent with content taught in the classrooms. With the constant change in the PSSA, how can an appropriate curricular alignment take place?
- The focus on doing well on a state test does not equate to educating the “whole child.”
- With recent concerns expressed about the “Reading First” initiative, how much faith can be placed in programs described as “best practices” or “research-based” which are implemented to reach proficiency under NCLB?
- The entire testing process demonstrates the degree of mandated testing measures ordered by Harrisburg and Washington. Where is the concept of “local control”, both political and educational, which is often quoted as the centerpiece of school governance in our society?

PSSAs

Elementary Concerns: PSSAs

- PSSA provides only a single measurement point on the barometer of measuring proficiency; multiple criteria for proficiency are needed.
- PSSA becomes a hammer to penalize districts; there is no positive reinforcement for success.
- PSSA scores may be the result of economic factors within communities, class size issues within districts and supportive services available to districts — none of which are accounted for in the test score itself.
- Students with Individual Educational Programs (IEPs) receive instruction at their instructional level but are assessed based on their grade level of assignment.
- Classroom grades are given based on true criterion-referenced grading systems; PSSA uses a norm-referenced-based system of reporting. Parents are easily confused by the incompatibility of the grading systems.

Middle School/ Junior High School Concerns: PSSAs

- As proficiency levels of AYP rise, will more areas within the curriculum, e.g. art or music, be reduced or eliminated in order to provide more “practice” for the test?
- PSSA should be used as a “formative” approach to assessment, not as a “summative” judgment of child’s and class’s performance. Formative assessment may be used to diagnose areas of weakness in order to achieve improvement while summative assessment may become an autopsy which simply shows where the child has failed.
- Is PSSA data on children in special education compatible with the data requirements of the IEP? Which supersedes which?
- Sub-groups of IEP students and English Language Learners should be judged as proficient and meeting AYP based on a growth model of assessment.

High School Concerns: PSSAs

The secondary group identified three positive aspects of PSSA testing:

1. The program provides a requirement for remedial support for low-achieving students.
2. The testing process has given an impetus to curriculum revision.
3. The necessity to improve test scores has improved district efforts to plan and provide professional development for faculty.

The secondary concerns were as follows:

- The “turn around” time to receive the test results is too long.
- State funding is needed for remediation efforts in the districts.

-
- Too much instructional time is lost to test preparation and test taking.
 - The curriculum is becoming a mile wide and an inch deep.
 - The entire process returns to a “one-size-fits-all” model; the state is far too diverse for this type of approach.
 - Schools should be allowed to use multiple data sources to determine “proficiency;” some students do not perform well on standardized tests or in an atmosphere that the PSSA test creates.
 - What is the motivation for students to perform well on the test other than to face remediation or more testing?
 - The tests discriminate against special education and English Language Learners.

GCAs

Middle School/ Junior High School Concerns: GCAs

- If remediation is required, state funding should be available at lower levels as well as at the high school level.
- How will the IEP students and English Language Learners be tested in the areas of science and social studies? In the areas of mathematics and English grammar and literature?
- Given the proposal to test in the four areas at the high school, will there be similar tests eventually for middle school and junior high school students?
- Will career-track students at the high school have the appropriate access to courses as the college-bound students in order to provide a level playing field for all students who may eventually be required to take the tests?

High School Concerns: GCAs

The secondary group identified three positive aspects of the proposal to institute Graduation Competency Assessments (GCAs):

1. There is focus on content-specific information related to specific high school courses.
2. The test is given at the completion of the course; hence, students should be prepared to take the tests.
3. The curriculum’s scope and sequence will be standardized statewide which should level the playing field for those who frequently transfer among schools and school districts.

The secondary group’s concerns about GCAs were as follows:

- Although a common curriculum should lead into the GCAs, how will these exams address student differences and student learning styles?
- Curriculum in the tested areas may become too narrow since teachers will teach to the test.

- Will the test results be available in time to provide remediation in time for students to re-take the test in time for graduation?
- What actual data exists to demonstrate that the currently developed local assessment systems required under Chapter 4 are NOT valid measures to assess proficiency?
- More instructional time will be lost to additional testing.
- Will the GCAs test higher order thinking skills or merely assess basic content mastery within the subjects?
- If students fail the GCAs, are their only choices to drop out or to leave the senior year without a diploma?
- Who decides the content of the GCAs? Will the same process be used to develop the GCAs as was used to develop the PSSAs?

Value-Added

Elementary Concerns: Value-Added Assessment

- Are there sufficient measures to determine an accurate assessment of what constitutes a child's and a class's previous performance?
- Will the system be used as an evaluation of individual teacher performance?
- Will initial high achievers face a "ceiling effect" on the tests and not be able to show their full growth?

Middle School/Junior High School Concerns: Value-Added Assessment

- Value-Added Assessment will be supportable if cut-scores are the same for all grade levels, if students can make AYP through a growth model and if the process can be applied to all students.

High School Concerns: Value-Added Assessment

- The concept's validity rests on the quality of the tests and the statistical model which predicts the level of expectation of achievement by the students.
- The process should not become the base for the evaluation of faculty.
- Value-Added Assessment is based on a growth model of achievement. The academic entry level of students is critical to the process.
- In urban districts with large in and out migration during the school year, how will the process deal with these transfers?
- Professional development is needed for staff to understand and to learn to work with the system.
- Since Value-Added Assessment deals with the basic skill areas of NCLB, how will the non-tested areas (e.g. art, music, and elective subjects) be affected as districts strive to increase the value added scores?
- The growth/value-added model will require an even closer alignment of the PSSA test, the standards and the local curriculum. How will this be accomplished?

PAESSP's Task Force Recommendations

- PSSA tests need to be true criterion-referenced measures.
- PSSA tests must not change constantly; the moving target must stop moving.
- PSSA tests should be subjected to a rigorous study of their reliability and validity semi-annually to guarantee that the data produced for students and school are accurate.
- The school experience must be more than teaching to an annual state test.
- The richness of the local curricular program, e.g. art, music, and social studies, must not be sacrificed to "improve" test scores.
- Multiple measures are needed to determine the "proficiency" levels of the students.
- Data from Value-Added Assessment measures should NOT be used to evaluate professional staff members.
- A clear distinction needs to be made between the concept of a "growth" model for AYP and the concept of "Value-Added" Assessment.
- GCAs should be an optional assessment process available to local school districts.
- Neither the PSSA nor the GCAs should be the only and final determiner as to whether or not a student receives a high school diploma in Pennsylvania.

Summary

PAESSP stands ready to enter into a productive dialogue with the Pennsylvania Department of Education and with the Commonwealth on the issues surrounding the impact of "High Stakes Testing" on Pennsylvania's children. PAESSP supports the concept of accountability and will continue to serve the citizens of Pennsylvania in order to assure and provide a quality education for all students of the Commonwealth.



122 Valley Road
P.O. Box 39
Summerdale, PA 17093
www.paessp.org