

# **National Association of Test Directors 2000 Symposia**

Organized by:

**Ray Fenton**  
**Anchorage Public Schools**

Edited by:

**Joseph O'Reilly**  
**Mesa (AZ) Schools**

This is the sixteenth volume of the published symposia, papers and surveys of the National Association of Test Directors (NATD). This publication serves an essential mission of NATD - to promote discussion and debate on testing matters from both a theoretical and practical perspective. In the spirit of that mission, the views expressed in this volume are those of the authors and not NATD. The paper and discussant comments presented in this volume were presented at the April, 2000 meeting of the National Council on Measurement in Education (NCME) in New Orleans.

The authors, organizer and editor of this volume are:

Ray Fenton  
Anchorage Public Schools  
PO Box 196614  
Anchorage, AK 99519

Anne Chartrand  
NCS-Pierson  
3310 Dresden Dr.  
Montgomery, AL 36111

Frederick Stofflet  
Anchorage Public Schools  
PO Box 196614  
Anchorage, AK 99519

Michael W. Strozski  
Garland Independent School District  
870 West Buckingham Rd.  
Garland, Texas 75040  
[mystro@tenet.edu](mailto:mystro@tenet.edu)

Robert Tobias  
New York Public Schools  
110 Livingston St.  
New York, N.Y.11758

Joe Hansen  
Western Oregon University  
345 North Monmouth Ave.  
Monmouth, OR 97361

Joseph O'Reilly  
Mesa Public Schools  
549 North Stapley Drive  
Mesa AZ 85203

**Table of Contents**

**Whose Standards Are They Anyway?**

Anne Chartland.....1

**Alignment of Curriculum and Instruction To State Standards & Assessments: A visit to the real world**

Frederick Stofflet.....10

**High Stakes, High Standards, Massive Testing & Improved Student Success**

FredMike Strozeski.....10

**Discussant Comments**

Joe Hansen... ..56

# **WHOSE STANDARDS ARE THEY, ANYWAY?**

**Anne Chartrand**

**NCS-Pierson**

Over the years, terminology and definitions for what we currently are calling standards have been tweaked, redefined, adjusted to new circumstances, and combined, often to the point of confusion. Many years ago, when discussing “setting standards,” we clearly were discussing the pros and cons of cut score methodologies suggested by researchers such as Ebel, Nedelsky, and the many ways to modify Angoff. Today, standard setting has come to be associated more with the articulation of curriculum, as well as indicating performance levels established on assessment results. The definitions of standards have become more varied and appears to change based on situation, purpose, objective, and interpreter. For example, just examine some of the past years. We could begin with behavioral objectives, and move on to goals, objectives, scope and sequence, competencies (minimum and otherwise), essential skills (as opposed to nonessential skills), curriculum frameworks, courses of study . . . you know, standards. Clarification also would be helpful for the variety of proficiency

levels used across the country. Can we adequately describe a student who is basic, proficient, above average, below average, advanced, level 1, level 4, below basic . . . you know, standards.

This is not meant to be glib, but if we continue to confuse the issue, how can we expect the public to understand what we are trying to convey? Educators and policymakers constantly are making the statement that they want to “raise the standards.” Does this mean the skills and knowledge expected of students are to be more difficult, or are we demanding an increase in test scores, a higher “cut” score on the same standards, or an increase in the number of students in higher proficiency levels? How do we best explain “standards of performance”?

For whom are the content standards developed? The answer to this question helps determine who takes them seriously. If the standards truly are a guide for teachers to use in everyday instruction, then the standards become the curriculum and those in individual schools must pay attention. If the standards are simply a list of skills contained in a document to prove to policymakers that standards exist, then they often are meaningless to teachers. Although Gandal reports that there has been improvement in the revising of state standards that previously were identified as vague and weak, his report still concludes that the efforts were “far from acceptable” (Gandal, 1996). In his recommended guidelines for the large-scale assessment community, Popham opines that most of the state-level content standards that he has seen “represent little more than pious wish-lists at generality levels little better than the gunky state-level curricular syllabi of yesteryear” (Popham, 1999). If standards are vague, nondescriptive, and not easily converted to instructional activities, then they are useless to the teachers.

Parent and students often believe the textbooks to be the standards because tests and grades are based so heavily upon textbooks. Parents and students also may have the impression that teachers are establishing their own content standards, as well as the standards of acceptance, i.e., grades. E.D. Hirsch published what he believes to be the “core knowledge” that all students should learn by grade level. His contention is that, while most curriculum documents are vague, there is some essential common content that is specific and that should be available to all students regardless of where they live. He believes that this core knowledge should be about 50 percent of the curriculum in schools (O’Neil, 1999). Not only are there many schools that have adopted this concept, but also some military parents who use it adamantly to determine whether they believe their children are receiving an adequate education as the family moves from location to location. Their determination of essential standards is derived, not from the state or school documents, but from what is happening in the current classroom and how it compares with their experiences in other towns.

Establishing clear content standards certainly plays a critical role in an instructional program. The value of content standards can vary, however, based upon who

developed them, for whom they are intended, and the realism of their application in the classroom. Some of the common problems that must be avoided are related to identifying standards for students that are too broad or too narrow, too rigorous or too trivial, too many or too few. If standards are merely the skills measured on a statewide test, then it is evident that the test controls the educational system. Standards must be communicated to and understood by both the teachers and the students. They must be able to be successfully taught, contribute to thinking and reasoning, broad enough to be sampled but identified sufficiently well enough to be taught (Schomaker and Marzano, 1999), and given the attention to allow teachers to be proficient in their teaching. This is not an easy task. We can give so much attention to the development and refinement of standards that we neglect the pedagogical support system that must impart them.

Any discussion about establishing content standards also must include a discussion of how to assess standards. Some of the same issues encountered in developing content standards also apply in choosing or designing a measure for the standards. The specifications for an assessment should come directly from the content standards if the state or district truly believes the content standards are the essential skills and knowledge students should learn. The assessments, like the content standards, should not be too broad or too narrow (item-focused).

Even if a state or district has acclaimed content standards, if all emphasis is on a norm-referenced test (NRT), it is at best difficult, if not impossible, to report exactly what the student knows and can do, but rather how that student compares in performance with other students in a norm group. In Popham's continuing discussion of the shortcomings of NRTs, he points out that in the "quest for score variance in a standardized achievement test, items on which students perform well are often excluded. However, items on which students perform well often cover that, because of its importance, teachers stress" (Popham, 1999, p. 12). There are other problems when an NRT is taking the major media role in large-scale assessment. Regardless of what standards have been established and, in some cases, what instruments measure them, the magic of the 50<sup>th</sup> percentile gets the most attention. Politicians (and many others) still equate increased test scores with increased student knowledge. Worthiness is defined by the national average, and the demise or salvation of a school can be determined by a single score interpretation. In one particular instance, a method was established to identify those schools "in danger." In order to be "cleared," a certain number of students had to move out of a specific, narrowly defined, band of percentile ranks. It was quite possible for the school to improve significantly its overall test scores, yet remain on the danger list because of this one small group of students. The achievement of the other students was ignored. When only a few students' scores can "make or break" a school's rating, the philosophical practice of teaching *all* students becomes less of a reality.

The overemphasis on a single test score still prevails. Many states have developed customized assessments that are in congruence with their standards and have followed intense development procedures to assure their assessment's reliability and validity. When the scores are released, however, attention befalls the low-scoring schools, where teaching to the test is then predictable. Although many educators would like to deny that this is the case, states where either their state-developed tests or NRTs are the sole determiner of the quality of a school, teachers readily will admit that all they do is concentrate primarily on what is measured on the test. Other activities and subject areas are ignored or dropped to the bottom or a priority list. Accountability, rather than instructional prowess, has control in many places. Have we created slaves to increase test scores at the expense of the best educational practices for all students? Yes, and this is certainly not a new problem. More effort needs to be expended in studying the few locations that appear not to have fallen victim to this undesired outcome so that we can determine the feasibility of assisting others who are overwhelmed with this predicament.

In considering the demand for increased test scores, when will they be high enough? This is not to suggest at all that we are not in need of vast improvement in many areas, but it is possible that the continuing cry for increased test scores may result in demanding cognitive skills that are unrealistic for many students? David Hoff addresses the issue of the predictability of ups and downs in test scores in a recent article. He cites several sources and reasons for test scores to "start low, rise quickly for a couple of years, level off for a few more, and they gradually drop over time" (Hoff, 2000). He goes on to quote researchers who agree that schools take care of the easy things first and neglect to make systemic changes that can make a difference in the long run. The cycle continues when policymakers then get anxious and search for a new test. How much growth is reasonable? How much higher can the standards be without becoming illogical to those who teach and to those who must demonstrate knowledge? Should all schools be expected to have the same amount of growth or the same score levels?

On a positive side, large-scale assessments, for the most part, are developed much more carefully than many years ago, and with much more input from content and instructional leaders who take their tasks quite seriously. Closer attention is paid to bias issues, and procedures and consensus attempts are put into place that involve many more people than just a test developer. Piloting, field testing, and scrutiny of every item for content are intense, many times due to the fear of litigation, but add to the quality of the instrument. The addition of performance assessments has contributed a component that seems to add a more realistic and instructional (if constructed well) element to the previously monotonous selected-response instruments of the past. The problems that stem from assessments are not often due to the assessments themselves, but rather to the handling of the results. This is why it is still so critical to continue to use other measures in the evaluation of students, schools, districts, and states.

Having discussed a few issues and concerns regarding standards and assessments, we must look at the impact these efforts have had on instructional improvement. There has been some research and much discussion on whether this emphasis on standards for student performance and assessment systems has proven to be a positive venture. Comments span both ends of a continuum. Some believe that there is so much emphasis placed on teaching to a test that overall student achievement has suffered because of the narrowed curriculum. There have been problems not only with being too item-focused or skill-specific, but also with being focused only on the students that can pull up the test results and relieve any fear of retribution for low scores.

Others, however, contend that, in certain circumstances, there is improved student achievement as a result of improved instruction. “New accountability systems that are well-designed (with fair, comprehensible, meaningful, and stable features) are associated with improved student achievement when adequate capacity to improve instruction is present in schools or can be provided by an outside partner” (Fuhrman, 1999, p. 10). The qualifiers in the parenthetical statement are not to be taken lightly. These are difficult elements to assure in a program. Furham goes on to say, “in the absence of explicit attention to capacity, the new systems are insufficient approaches to improving student achievement “ (1999, p.10). The critical nature of professional development to provide teachers with the capacity to teach the standards proficiently needs to be at the forefront of most discussions and yet, often is ignored. This ability of the teachers to impart knowledge and provide assistance to those needing additional help, for the most part, will determine the success or failure of a “standards and assessment” program designed to increase student achievement. Again, it appears that, in many instances, much more time, effort, and funding are spent on developing and documenting the standards, than on providing the necessary assistance for the teaching of the identified essential skills.

In discussing “lessons from last decade’s reforms,” it has been noted that states with the highest test scores “have long supported high-quality teaching and teacher learning.” These states do not necessarily have strict statewide curriculum or high-stakes testing programs, but they “do have a long history of professional policy. Reform strategies that did not make substantial efforts to improve teaching have been much less successful” (Darling-Hammond and Ball, 1998, p.3). Teachers in the classroom must have the opportunity to secure the skills necessary for success. Therefore, teacher-training programs must become involved more adequately so that standards for students and standards for teachers are precisely aligned.

The underlying question is whether the emphasis on standards and assessment is having a positive effect on instructional improvement and overall achievement. Many say the verdict is still out, especially since there are so many places having to take to heart the effects of initial failure of some of the newer, tougher standards and the assessments that measure them. A few places are retreating from their initial

requirements in light of all the students that will be impacted negatively, while others are crying they should not kill the messenger but stand by the rigorous advances.

So, are all of these efforts working? Yes and no. (Hasn't this always been true?) Some who have researched specific locations declare that these systems have "helped channel teachers' work to the most important goals of the system . . ." and that some of the consequences have helped to motivate teachers to work in "more focused ways to produce improved student achievement" (Fuhrman, 1999, p. 6). Organizations such as Achieve are assisting in the dialogue of helping states improve their systems and have been encouraged by the direction and commitments of some states and districts, as well as their policymakers.

One significant problem in answering this question of effectiveness of the systems, lies in how often programs are changed before they can be declared successful or unsuccessful. Newly elected or appointed officials often want to make their presence known by restructuring or hastily changing the programs in place. We seem to start over constantly with something new before there has been sufficient effort and research to make an educated statement about the impact of a specific program. It may take three years to put a program in place, but positive results are expected within six months or the program is doomed. In addition, it is not uncommon for a new system to be mandated, yet, not adequately funded. Policymakers and the public, however, still hold the schools and districts accountable for the program's success.

Perhaps, the very best result of the efforts toward higher standards has been the dialogue created. When educators and policymakers gather and critically discuss the reality of the nature of our schools and how to improve what happens in classrooms, some good comes of it. Thousands of groups of teachers have been brought together to review and analyze their curriculum and discuss it across grade levels. This is not new, but it must continue if teachers are to keep abreast of what remains essential in curriculum and process. Each time there is a new effort that forces people to evaluate critically the current system, many are effected in a positive way – if simply by a deeper knowledge and expression of what they are doing and why.

This critical evaluation of classroom goals occurs at many levels, but when it occurs with teachers it can affect their students directly. One example is from an experience with the development of a graduation exam and one of the necessary tasks of the process to assure instructional and curricular validity. A statewide survey was conducted with teachers of various grade levels who were to answer questions regarding specific skills that they taught in their classrooms. The subject area was math and this particular incident was with seventh and eighth grade teachers. They were to answer whether they taught certain specific skills to their students. If the answer was no, they were to state why (the skill is too easy, the skill is too difficult for that grade level, etc.). The results for several skills astounded the teachers. Teachers had confirmed already that these skills were, indeed, essential skills for the

curriculum. The seven grade teachers stated that there were several skills that they did not teach because they were too difficult for the grade level. The eighth grade teachers declared that they did not teach these very same skills because they were too easy for the grade level! If nothing else, they certainly learned about the lack of communication and set out to establish exactly where these “essential skills” had fallen through the crack. Dialogue and critical evaluation by discussion create knowledge and understanding. An effective standards and assessment system generates this dialogue.

There are other benefits of these current efforts. There seems to be an attempt in many places to do a better job in bridging the disciplines and there is more emphasis, at least, on addressing the critical need for meaningful professional development. Another benefit has resulted from the fact that we must address more explicitly the issues of diverse populations in our schools. The cry for higher standards has brought more national attention to the need to provide for all students.

Still, there are so many places where programs have not made an impact and, in some cases, may have created problems. There is still too much attention on a single score on a single test at the expense of overall achievement of all students. When these assessments dictate, the interpretation still seems to be extremely grade-oriented. If it is a sixth grade test then the burden is on the sixth grade teacher. This is astounding since we have been battling this since the days of minimum competency testing. An entire school must accept responsibility for the growth of a child and not lay the burden on only the grade levels tested. Something about “it takes a village . . .” Funding efforts are often expended for the development of standards and especially for the assessments that measure them, ignoring the professional development that is needed to assure that the teachers have the capacity to teach well. Changes in subject areas, particularly with the onslaught of technology, demand that teachers have every opportunity to hone their skills and enhance their practices.

Identifying standards does not make “kids smarter.” Administering assessments does not make “kids smarter.” However, if a system is implemented with the goal of a more effective instructional program to maximize the potential of each student, then a significant impact can occur. What happens in the classrooms is still the heart of the matter. One thing seems clear. Even those of us who feel there is a lot of “deja vu all over again” realize, if we never embraced new efforts, discussions would cease, evaluation of current practice would diminish, and stagnation surely would result. We must keep everyone talking and analyzing if our programs are going to be successful. Are these programs working? In some places, yes in some places, no – just as in the past. Are these programs worthwhile? Absolutely. They keep us trying to find the answers.

High Stakes, High Standards, Massive Testing, and Improved Student Success. The Stakes and Standards Increase Each Year, While Instruction and Student Performance Continue to Improve. It's A Moving-Target Kind of World

**Michael W. Strozeski**

**Garland TX Public Schools**

What do you think of when you hear the expression: “*The Real Thing?*”

What do you think of when you hear: “*Just Do It!*”

What do you think of when you hear the words: “*Tony The Tiger?*”

What do you think of when you hear the words: “*Public Education?*”

What is the public’s general reaction to “*Public Education?*” Why does the public react that way?

We have come to believe the mythology as Truth! As an example of a significant Myth – we frequently hear that the United States is at the bottom of international lists when it comes to education. That is simply “Hogwash!” In case you are not familiar with the term “Hogwash,” suffice it to say that it doesn’t smell too good.

Frequently teachers and other educators have a view of public education much like the rest of the American public. Teachers have “bought” the mythology. Why is that? Why are there frequently negative perceptions about public education? For teacher, it’s usually because most teachers have inaccurate and incomplete information about public education to the United States. Usually we receive incomplete and often inaccurate information about our schools – be it locally or nationally. Media focus is always on the negative. Much of what the media produces is complete in error, or at best negatively slanted. Because of this in part, there is growing support for alternatives to public schools. Fewer and fewer people “know” anything about the public school unless they hear about it on the radio, TV, or read it in the newspaper. Ask today’s parents about media headlines versus “My School.” Make note of the fact that we will all be drowned in political rhetoric for the next several months about the “plight” of our public schools. There is growing disparity between the “haves” and “have-nots” of public education. Examine the Phi Delta Kappa – Gallup Poll. Look at the perceptions of parents versus non-parents in those polls. Try to “weed-out” the perceptions from the facts – Provided you can obtain access to the facts. And, unfortunately, today there are far too few spokespersons for our public schools.

How do we know that “Myths” are just – Myths? We know the reality only because we have measured the academic achievement of our students, -- i.e., *We* know because we give *TESTS* to our students. And, in Texas, we give *LOTS* of tests!

In 1997 the Commissioner of Education for the state of Texas, Mike Moses, said in an address to a convention of school boards and school administrators: “We should be telling people about the rising student achievement in Texas because the story is a good one. Students’ scores are up in every category on T.A.A.S. and also on N.A.E.P.”

The state of Texas has one of the most diverse populations of public school children in the United States. This diversity is heavily weighted in terms of student populations

that have historically demonstrated lower-than-average student academic success. Between 1987 and 1999, the public school enrollment in Texas increased by more than 20 percent, while the number of economically disadvantaged students increased by almost 70 percent. In 1998 the percent of students identified as being economically disadvantaged in Texas was in excess of 48 percent. During the same ten-year span, the number of students in Special Education increased by 75 percent statewide, and the number of students enrolled in English as a Second Language (ESL) or Bilingual programs doubled. Minority student enrollment in the state increased by almost 45 percent.

Texas has been a leader in the assessment of public school students for the past twenty years. Texas students have been given, T.A.B.S. (Texas Assessment of Basic Skills) tests, T.E.A.M.S. (Texas Educational Assessment of Minimal Skills) tests, and T.A.A.S. (Texas Assessment of Academic Skills) tests (all CRT tests) during that period. Subjects measured for inclusion in the accountability system are reading, mathematics, and writing. The State of Texas has operated a rigorous accountability system (Academic Educational Indicator System, A.E.I.S.) in “full-swing” for the past seven years. The A.E.I.S. accountability system is both rigorous and comprehensive. It rates several indicators of progress for students and schools across grades three through eight and ten. Virtually all students in those grades are included in the accountability system and, in addition, the system requires increasingly (every year) high standards for all student groups (i.e., African American students, Hispanic students, White students and Economically Disadvantaged students). This accountability system is committed to achieving high performance for all of its students, regardless of ethnicity, socio-economic status, academic ability, or primary language. Students, educators, parents, and communities are responding positively to the performance expectations dictated by this accountability system.

Several years ago, then Commissioner of Education for the State of Texas, Lionel “Skip” Meno set a standard for performance on the state CRT assessments of 90 percent. (The 90 percent standard was set for districts and schools to acquire a state rating of “Exemplary.”) He reasoned that 90 percent of all students and 90 percent of each student groups should be expected to pass the statewide assessments. “How can he set such a ridiculous standard?” was an early response throughout the state. My response, “I agree how can he advocate throwing away ten percent of our kids?” It probably should be noted that, in the beginning, my response frequently brought an end to further discussions of the subject.

Statewide we test more kids than ever before. In 1998 91 percent of all students enrolled in grades three through eight and ten on the day of testing were tested with T.A.A.S. We have increased standards every year for campus and district ratings and academic performance has improved every year for every students group.

The grade ten T.A.A.S. tests are required for high school graduation and each year more and more students meet the testing requirement for graduation on their first try. At the same time dropout rates (another indicator required by the A.E.I.S. system) have steadily declined for all students and all student groups. All the information from the tests (as well as dropout, grade level performance, student group performance, exemption rates, and attendance rate information) is made public every year for every campus and every school district. You can find about six years worth of data for every campus and district in the state on the Internet at: <http://www.tea.state.tx.us>. It's all made public! The actual T.A.A.S. tests are released to the public each year after all students have been tested and school accountability ratings are delivered to schools and districts.

The Texas Education Agency (TEA) has evidence that greater writing proficiency is demanded of today's fourth graders than was required of ninth graders in 1980. That means today's fourth graders are writing better compositions than ninth graders were twenty years ago. Since 1995, average T.L.I. (Texas Learning Index) scores (a transformed raw score – similar to Z-scores) have increased for all rating levels in the state. Schools are not just getting kids to pass the test with minimum scores, but are concerned (and successful) with improving high achievers as well. The Texas Learning Index scores also allow longitudinal analyses of the same students across the years.

An indication of the progress made for student groups in academic performance in writing is indicated in Table 1 below.

**Table I**  
**T.A.A.S. Scores (percent passing) in Writing**

	<b>1994</b>	<b>1999</b>	<b>Up</b>
African American	65%	83%	+18
Hispanic	69%	84%	+15
White	87%	94%	+7
Economically Disadvantaged	67%	83%	+16
All Students (except Sp Ed)	78%	89%	+11

An indication of the progress made for student groups in academic performance in reading is indicated in Table II below:

**Table II**  
**T.A.A.S. Scores (percent passing) in Reading**

	<b>1994</b>	<b>1999</b>	<b>Up</b>
African American	60%	79%	+19
Hispanic	64%	81%	+17
White	87%	95%	+8
Economically Disadvantaged	62%	80%	+18
All Students (except Sp Ed)	76%	88%	+12

An indication of the progress made for student groups in academic performance in mathematics is indicated in Table III below:

**Table III**  
**T.A.A.S. Scores (percent passing) in Mathematics**

	<b>1994</b>	<b>1999</b>	<b>Up</b>
African American	37%	74%	+37
Hispanic	46%	82%	+36
White	72%	94%	+22
Economically Disadvantaged	44%	80%	+36
All Students (except Sp Ed)	59%	87%	+28

Improved student performance is not just measured on the Texas statewide tests. Our students are scoring better on T.A.A.S. to be sure, but also improvement is noted on N.A.E.P. scores for Texas students (particularly minority students). Advanced Placement tests are being taken by more students than ever before and the scores on AP tests are higher than ever before.

**Table IV**  
**1996 N.A.E.P. Scores**

- ☞ Texas fourth graders' math scores up 11 points over 1992 scores
- ☞ Only five states had higher average scores (Maine, Minnesota, Connecticut, Wisconsin, and North Dakota)
- ☞ Eighth graders' scores up 12 points over 1992
- ☞ Hispanic fourth graders' scored sixth in math among 39 participating states
- ☞ African American fourth graders' scored higher in math than their counterparts in all other states
- ☞ White fourth graders' also scored higher in math than their counterparts in all other states

**Table IV**  
**1998 N.A.E.P. Scores**

- ☞ Texas fourth graders' improved their reading scores from 213 in 1992 to 217 in 1998 (National average was 215)
- ☞ Highest reading score among the six demographically similar states
- ☞ Eighth graders' scored 262 on reading, one point higher than the national average
- ☞ Hispanic eighth graders' scored highest among the six demographically similar states

**Table V**  
**1999 Advanced Placement**

- ☞ Students taking A.P. tests up 142 percent (1994-1999)
- ☞ Number of A.P. exams taken up 160.7 percent
- ☞ Schools offering A.P. courses up from 544 in 1994 to 971 in 1999
- ☞ Number of students taking A.P. up 16 percent over last year (51,228 students)
- ☞ Number of A.P. exams taken up 19 percent (88,484 exams)
- ☞ Number of 3 or higher scores up 15.9 percent (49,721 score of 3 or more)

The scores and statistics reported above are statewide statistics. In my district, the Garland Independent School District, most of the above T.A.A.S. performance scores would earn a principal a one or two year probation period before being non-renewed. The Garland I.S.D. has over 50,000 students in grades K-12. The student population is approximately: 49 percent Anglo, 18 percent African American, 26 percent Hispanic, and over 35 percent economically disadvantaged. In the Garland I.S.D. we give lots of tests: T.A.A.S., I.T.B.S., Cog. A.T., and S.A.B.E., as well as diagnostic reading tests for students in grades K-2. Today all of our six full-time high schools have approximately 90 percent or better of our tenth grade students passing each of the T.A.A.S. tests on the first administration of the tests.

Statewide, in 2003, students failing to pass T.A.A.S. reading will be retained in grade three. In 2005, students failing to pass T.A.A.S. reading and mathematics will be retained in grade five. In 2005, students failing to pass the high school exit-level T.A.A.S. tests at grade eleven (Mathematics including Algebra I, Science including

Biology, Physics and Chemistry, English language arts including writing, and Social Studies including U.S. History) will be denied a diploma. Students will have multiple opportunities to take the tests for graduation starting in grade eleven.

Public Education is indeed a *moving target* sort of world. Standards are being made more rigorous, more students are being tested, and more subjects are being tests. In Texas, all of our students are doing better now than they ever have before. Attitudes and expectations were our first challenge to success. Now, we just have to produce.

***Expect It, or Forget It!***

# **Discussant's Comments.**

**Joe B. Hansen**

The three papers presented in this symposium deal with some of the most important core issues facing public education today. Each takes a different approach, but as a group they come together to provide the basis for a stimulating, albeit frustrating discussion of the foibles of standards based education as a tool for reform. In this brief paper I review each paper separately, bridging where I can the thoughts and themes that are common and critiquing where appropriate.

The first presentation is by Ann Chartrand, who asks the question – Whose Standards Are They Anyway? I can summarize my reaction to this paper in four words “thoughtful and well grounded.” Dr. Chartrand begins by identifying four key issues involved with standards and then deals with them in a thoughtful and informed manner.

First is the definitional issue. Here she makes the point that there is a need for a more common understanding of what is meant by “standard” regardless of whether the focus is on assessment or content. Confusion as to the meaning of the term standard is not uncommon.

Second is the issue of need for support systems of professional development, and instructional systems to make standards work.

Third, she raises the central question of “for whom are content standards really developed? Is it students, teachers or perhaps politicians? A similar question could be asked regarding performance standards.

Fourth, she points out problems associated with over-emphasis on norm referenced assessment in particular and especially those problems stemming from focusing too much attention on a single test score on a single test as the criterion for reaching a standard.

Together, these issues reveal that a fair amount of confusion exists regarding what standards are or ought to be and how to make the most effective use of them.

Ann then discusses the instructional impact of standards, pointing out the importance of developing the professional capacity to teach the standards. This leads to a somewhat muddled, but realistic view of the effectiveness of standards as a vehicle for instructional reform. She reminds us of the impatience of the political process in expecting quick results from reform efforts that are not well conceived.

Three benefits of standards are identified, the first of which is the dialogue about education, curriculum, instruction and assessment that has helped to create knowledge and increase understanding. The bridging of disciplines that has occurred under the standards movement is another benefit. Standards have resulted in greater integration of instruction across disciplines and presumably has led to more integrated learning by students. Yet another benefit is the increased emphasis on “meaningful” professional development. I found this intriguing but was left wondering about the meaning of “meaningful.” I would expect that meaningful professional development would have a positive impact on student learning.

On the question of whether standards are working or not, that is, are they effective in improving student achievement, the answer is at best mixed. There is still too much emphasis on single scores from what, in my opinion are often inappropriate tests. And standards are still too grade level focused.

Finally, Ann makes the excellent point that what happens in the classroom is really the “heart of the matter.” I couldn’t agree more with that statement if there were two of me.

Critically speaking, I felt that this paper raised but failed to answer the question of “whose standards are they anyway?” Are they for the politicians who want to score points with the public for looking like educational reformers? Are they for the teachers who must understand how to teach standards and how standards based teaching differs from a traditional approach? What about the students? Shouldn’t standards really be for them? Or are standards for the broader community, to provide them with assurance that their educational tax dollars are well spent and their students are learning?

My recent personal experience with a longitudinal study of early career teachers in a standards based setting has revealed that, at least in the state of Oregon, there is no common understanding or approach among the teacher preparation institutions to training teachers to teach in a standards based system. The seventy or so teachers in our study come from seventeen different institutions in Oregon and it is thus far evident that there is no shared vision of what standards based teaching is or ought to be. There is great variability in how standards based education is defined and in how teachers are prepared for it. It is no small wonder that we have a muddled view when we look out across all the classrooms in all the schools in all the districts, not just in Oregon, but throughout the nation. I have but one thought to offer and that is – perhaps it is time to establish a set of meta-standards for standards. These meta-standards would have to be developed by representatives from higher education, pre-k through 12 and would have to be endorsed by appropriate professional organizations. They may take years to develop and gain acceptance. But if we believe that the standards movement is a non-transitory phenomenon then we owe to ourselves as educators, researchers, evaluators and measurement professionals to develop a common understanding of what “standards” really are.

Mike Strozeski’s presentation – High Stakes, High Standards, Massive Testing and Improved Student Success was a good paper, but I’m somewhat concerned about the things that it didn’t say. Mike begins with a challenging assertion that the often heard statement that the United States is at the bottom of international lists, educationally is “Hogwash”; implying that he will provide evidence to the contrary. What he does instead however, is to regale us with data showing how much improvement Texas students have made over the last five or six years, since standards based educational reform went into effect. Mike makes a legitimate point that much of the negativism about American public schools is mythical and that most people lack the information they need to fully understand the true situation. Of course, Jerry Bracey has crusaded under this banner for over a decade with his annual Bracey Report published by Phi Delta Kappa. According to Mike, we have the media to blame for this epic misunderstanding.

He shows us that while the enrollment in the Texas Public School System has increased by twenty percent over a twelve year span, the increase in economically

disadvantaged and ESL/bilingual students was far higher, implying that scores might be expected to go down rather than up.

In describing the Texas accountability system Strozeski states that the system is “committed to achieving high performance for all of its students.” I question or not this is the proper role of an accountability system. I believe the accountability system should limit itself to monitoring and reporting on the performance of students and schools and on the extent to which the educational system is being successful in helping students to reach high standards. The accountability system itself should be independent of and external to the standards. When the accountability system becomes responsible for *causing students to reach high standards* confusion occurs over what the means and the ends are. This can result in an over emphasis on the test scores, as Ann Chartrand points out.

The Texas commissioner’s goal that 90 percent of all students and 90 percent of each student group passing the statewide assessments has still not been met after more than five years of “reform.” I confess that I didn’t understand what Mike meant when he said that “the standards have increased every year.”

Tables I through II of Strozeski’s paper show data on the percent of students passing the TAAS in 1994 and 1999 for three ethnic groups plus economically disadvantaged and for all students except special education across the subject areas of reading, writing and mathematics. In 1994 87 percent of students passed in writing and reading, increasing to 94 percent in 1999 for a seven percentage point gain. While African-American and Hispanic students started with an 18 to 22 point lower passing rate and improved 18 and 15 percent respectively. A similar pattern occurs in reading. This raises in my mind a few questions that are not asked in the paper and for which no data are shown.

If 78 percent of students passed the TAAS in 1994, was the test cut score set high enough to begin with? What was the cut score and how was it determined? By whom? What do the quartile distributions look like and how did they change over time? Did lower quartile students improve as much as higher quartile students? Were changes equitably distributed across quartiles?

Another I would raise is whether we should have different standards of improvement for different levels of performance. In the end I was left with questions about the rigor and difficulty of the Texas accountability system that were not answered by the data presented.

Bob Tobias begins by telling us that New York State has been witness to both the promise and the challenges of standards based education. He provides us a list of Essential Elements of a Standards Based Education System, which properly used, could be a great tool for evaluating the quality of a standards based system. These

seven standards for a standards-based system, begin to approach the meta-standards I called for upon reading the Chartrand paper.

1. Clearly stated standards supported by consensus...
2. High quality assessments ...
3. Necessary resources...
4. Safety net for students who don't meet the standards...
5. Feedback (a reporting system) to teachers...
6. Stakes ~ an accountability system. Students and staff need consequences...
7. Continuous evaluation and revision...

I would award the New York State system especially high marks for standards 1,2, and 5. But has the system made a difference for students? Has it lived up to its promise?

A much greater percentage of students passed at all levels in 2000 than in 1999 and there was a larger percentage of students taking the test. The coursework based on the test became more difficult resulting in fewer students completing the work needed to stay on track with their cohort. 21.5 percent of those taking the ESL test dropped out.

My questions are: What's going on in the classroom? Do the teachers know how to teach to the standards? Do the teacher preparation institutions know how to prepare teachers to teach in a standards based educational world? Do the school districts have the capacity to develop either the curriculum or staff competencies?

As Tobias points out *necessary resources* are a key issue. Reform efforts that don't address the issue are not likely to succeed, yet we see the pattern of under funded reform repeated throughout the land. The *safety net* called for in the standards checklist, in the form of alternative programs for students who don't meet the standard, has not been put in place. These are common problems for many school districts struggling to make a standards based system work. We need to find solutions to these problems if we really expect standards based reform to make a difference for our nation's children. These problems are not created by educators, but by politicians and only they can solve them. While they are so busily trying to hold public education accountable we should hold them accountable for providing the resources and support systems needed to make their reform plans work.

