



PRESIDENT'S MESSAGE

Judith Levinson, Evanston Township High School

One cannot read the newspaper these days without bumping up against articles on the state of the nation's schools. It is that time again...the time when schools' testing results are published and it is revealed which school districts have been naughty and which have been nice. Only this year the stakes are higher as the sanctions of the No Child Left Behind law are put into effect. Those that have been naughty are noted as low-performing schools that have failed to meet state standards for adequate yearly progress. They may already be losing students and resources or may be on their way to doing so. Those that have been nice have, at least for the time being, avoided the "failing label."

During the AERA convention last year, the NATD/NCME symposium focused on the No Child Left Behind legislation one year into implementation. Several perspectives were shared. Zollie Stevenson (Supervisor of Standards, US Department of Education) provided a summary of criteria that the federal government used to review state plans for meeting the requirements of NCLB. Phil Hansen (Chief Accountability Officer, Chicago Public Schools) reviewed activities at the state and school level to address the new mandates. Glynn Ligon (Evaluation Software Planning) spoke of the work that states are doing to set up systems for collecting and monitoring adequate yearly progress of schools and districts. Allan Alson (Superintendent, Evanston Township High School) discussed the ramifications of NCLB on smaller urban-suburban districts that have stellar reputations for instruction but will be considered failing under NCLB. The

symposium was received well with lively discussion and debate. Although many in education accept the intent of NCLB, that of the need to focus on and improve the performance of traditionally under-performing groups, they have concerns about the use of widely varying state tests to determine high stakes for students and schools.

Recently, I reviewed the purposes of the NATD organization. They are:

- To share information about testing in educational settings.
- To encourage the appropriate use of testing in educational settings.
- To improve the applications of measurement to students and educational programs.
- To encourage research in the area of elementary and secondary school testing and measurement.

These are lofty goals and the impetus for our involvement in the Joint Council on Testing Practices. This group represents AERA, APA, NCME, ACA, ASHA, NASP and NATD. The bylaws of this organization state that JCTP provides "a means by which professional organizations and test publishers can work together to improve the use of tests in assessment and appraisal." JCTP work groups have produced a number of important publications including "The Code of Fair Testing Practices in Education", "Test User Qualifications: A Data-Based Approach to Promoting Good Test Use", "Responsible Test Use: Case Studies for Assessing Human Behavior", and "The ABC's of School Testing". Steve Henry and I attended a

meeting of JCTP in September that focused on final approval of an update to the “Code of Fair Testing Practices in Education”. NATD’s involvement in the Code helps further our stated purpose of encouraging the appropriate use of testing in educational settings.

In attending the JCTP meeting and reading the fall newspaper barrage of articles on NCLB, I have been thinking about some of the testing/measurement dilemmas of the NCLB legislation. For example, the law requires that not only the whole school but also racial, economic, IEP and LEP subgroups within a school make adequate yearly progress. Some have referred to this method of calculation as a “trip wire” system, in which poor performance by one subgroup in one subject area can “trip up” the whole school. It is possible under this law that the performance of one subgroup such as special education students can result in identifying an entire school as a failing school even though the majority of the student population and subgroups are making AYP.

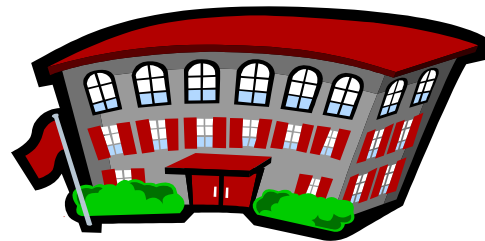
Another problem concerns the performance of students with limited English proficiency. Once LEP students become proficient, they exit the group. This “Catch 22” makes it difficult for this subgroup to meet performance targets because students who score at proficient levels meet exit criteria and are no longer part of the group.

Robert Linn of CRESST and the University of Colorado recently reported results of studying different methods for defining proficiency on tests (Linn, R.L., 2003, Education Policy Analysis Archives, 11, 31). Using three methods, he analyzed the results of middle school students’ performance on a reading test administered by the state of Kentucky. When he used the bookmark method, he found 61 percent of the students were considered proficient. With the contrasting-groups method, he found a pass rate of 22.7 percent and with the Jaeger-Mills method, he identified 10.5 percent as proficient. The implications of this study for the accountability systems developed for NCLB are clear. States were given the flexibility to choose their methods, resulting in a variety of definitions of proficiency across the nation.

Some schools and districts not meeting AYP in one state might meet proficiency criteria in other states.

There are also indications that some states are changing their standards as NCLB moves from the planning stage to implementation. Many states use test results to determine if students will graduate high school. Since graduate rate is tied into achieving AYP, states are beginning to review their cut-offs. New York, for example, adjusted its passing score on its mathematics test after a large number of students failed according to the standard and were not going to graduate high school. As W. James Popham noted in a recent Education Week article, “State officials realize that if they simply lower what’s expected in order for a student to be deemed proficient, fewer schools will fail. Unfortunately, lower expectations almost always yield lower accomplishments – from both teachers and students.” (Education Week, 9/24/03)

These are just some of the testing issues associated with the NCLB legislation. As the law moves into its third year of implementation, more problems will likely be identified. My questions to NATD members are: Are you concerned about the implications of No Child Left Behind and is this an issue that the membership wishes to address? If so, in what ways would you want to address it? Should this be a discussion item at our annual meeting? Or are there other venues in which to address it? Please share your views with the membership by using our NATD list serve: natd@coollist.com



Criteria to Evaluate Products and Services to Prepare K-12 Students for High Stakes Testing: White Paper Development

Report by: Peter Hendrickson,

One of the handmaidens of high stakes testing in a free market economy is opportunity for entrepreneurs, large and small, to sell test preparation products and services. The at least implicit promise is that purchase and use of the products/services will result in elevated scores. Many school districts, schools, and individual teachers commit fiscal and time resources to raise scores. These interventions may extend for weeks or months, gaining de facto program, if not curriculum, status.

Selection of curriculum materials has become increasingly sophisticated and districts are beginning to eliminate from consideration those materials without a credible research base validating their effectiveness to impact student achievement. Requirements in the No Child Left Behind legislation will sharpen this practice.

There are no such standards for selection of test preparation programs and there is scant evidence that any test preparation program improves what students know and are able to do beyond what other curriculum and instruction produce.

Conversations in several districts suggest significant dollars are currently spent on test preparation relying only on face validity and word of mouth to make purchase and use decisions. In some cases it appears that test preparation materials have replaced the regular curriculum as teaching units.

The Washington Education Research Association (WERA) and the Oregon Program Evaluators Network (OPEN) are jointly funding a study to produce a White Paper on preparation products and services for high stakes state tests. WERA has a tradition of gathering thoughtful practitioners and researchers together on high profile learning and assessment issues to influence policy and provide guidance to practitioners. OPEN has a history of supporting the community of

program evaluators while seeking to improve program evaluation practice in the Northwest. The joint White Paper on this topic will address, at least, the following issues:

- What data support the efficacy of high stakes test preparation programs?
- What criteria and process might educators use in selecting or eliminating test preparation programs?
- What guidelines, ethical and pedagogical, are there for use of test preparation programs?
- What standards are there for the use of test preparation programs?
- Develop and test guidelines and standards applicable to high stakes K-12 testing.

White Paper convener Peter Hendrickson of Evergreen School District in Vancouver, WA hopes that there will be a national audience for the publication which will be posted on the WERA website (www.wera-web.org) in 2004 and presented at the annual WERA Spring Conference. Members are considering surveying NATD members on current practices and services.



**NATD Recognizes the
Professional
Contributions of Two of its
Members**

At the annual meeting in Chicago last year, NATD presented its first *Certificates of Recognition for Outstanding Contributions to Educational Assessment* to two highly respected members of the association: Dr. Carole Perlman and Dr. Edward Drahozal. Each was recognized individually for her and his service to the organization. The certificates contained the following inscriptions.

Dr. Perlman. *Dr. Perlman's leadership, as a member of numerous national, state and local commissions on testing, as an author of several important scholarly articles on measurement and assessment, and as a board member of various professional organizations concerned with testing and assessment, has contributed significantly to the status and recognition of the profession. By serving in these capacities with class and integrity Carole has set a standard all members of the profession to follow.*

Dr. Drahozal. *Throughout his long career Dr. Drahozal has been a powerful force in the development of community among testing and assessment professionals in the schools. His enthusiastic support of research and evaluation organizations at the national, state, and regional levels has directly or indirectly affected the professional lives of nearly everyone in this association. Ed's friendship and humor are enduring assets that have and will continue to be appreciated at every level of the profession.*

Congratulations Carole and Ed!



**Nominate a Colleague for the NATD
Outstanding Contributions in Educational
Assessment Award.**

The NATD Board of Directors recently established an award to recognize individuals who have made outstanding professional contributions in the area of applied educational assessment in schools. One or two individuals may be recognized each year conditional upon approval by the Board of Directors. Individual(s) recognized by this award can be nominated from any of the myriad areas in the wide field of educational assessment including but not limited to measurement theory, promotion of best professional practices, teaching of measurement and assessment, and use of assessment information for curriculum, instruction, policy making and communication with stakeholders.

To nominate a candidate for this award, send a letter (preferably email) to Dr. George H. Olson at the address below. The letter should provide strong justification for why your candidate should receive the recognition. Nominations should be received before March 1, 2004.

Send nominations to:

Dr. George H. Olson
Leadership and Educational Studies
Appalachian State University
Boone, NC 29608

YOUR help is needed. You, too, can help build NATD membership. Please share the membership form from this newsletter with five colleagues whom you feel would benefit from being a member. It's just \$20.00 and it's a *wonderful* bargain.

NATD members regularly present papers at the Annual Meetings of the American Educational Research Association. These papers often address issues of interest to the NATD membership. The following is an extended abstract of a paper presented as part of the symposium “Do standards-based tests differ fundamentally from norm-referenced tests in practice?” at the Annual Meeting in Chicago, April 2003.

Validity Evidence for Washington Assessment of Student Learning (WASL) Performance Standard Cut-Scores for Reading and Mathematics

Duncan MacQuarrie, Currently with Harcourt

The No Child Left Behind (NCLB) federal legislation mandates standards-based assessments in grades 3 – 8 and at one grade level in 9-12. In addition, performance on these assessments must be classified into at least three levels of performance: “basic,” “proficient,” and “advanced.” The validity of these classifications, and the inferences about students and schools that are based on them, are of great importance. Lewis (2002) has suggested that one dimension of the validity of such performance categories is the degree to which they are based on consistent cut-scores across grade levels within content areas. MacQuarrie and Willhoft (2002) have suggested that such cut-scores (performance standards) must also be perceived as fair or reasonable.

The state of Washington has instituted standards-based assessments, named the Washington Assessment of Student Learning (WASL), at grades 4, 7, and 10. These assessments will form the core of the system required under NCLB. Differences in the performance of students across grade levels within subjects and between subjects within grade levels on the WASL have raised questions about the cut-scores in reading and mathematics for this assessment system. Reading performance at 7th grade deviates markedly from those at 4th and 10th grades. Mathematics performance is consistently low across grade levels, however performance at 7th grade is even lower than that at 4th and 10th grades. These patterns have raised questions about the consistency and reasonableness of the WASL performance standards.

In addition to the WASL, the Washington State Assessment System includes norm-referenced assessments in grades 3, 6, and 9.

This alignment of “basic skills” assessments in the years prior to the more application oriented standards-based assessments provided an opportunity to index the difficulty of the performance standards in norm-referenced terms. Data sets were available from the spring 2000 4th grade state assessment containing student records with both standards-based reading and mathematics assessment results and corresponding prior year norm-referenced reading and mathematics results. In both 2001 and 2002 comparable data sets also included student records for 7th and 10th grades.

In 2000, 2001, and 2002 over 70% of the students entering 4th grade with 3rd grade ITBS national percentile rank (NPR) reading scores of 50 to 54 subsequently met the spring WASL performance standard in reading. In comparison, in 2000 and 2001, only 30% of the students entering 4th grade with 3rd grade ITBS NPR math scores of 50 to 54 subsequently met the WASL performance standard in math. In 2002 that percent rose to about 43. In 2001 a little over 30% of the 7th grade students entering with ITBS NPR reading scores of 50-54 subsequently met the WASL performance standard. In 2002 that percent rose to about 38%. The comparable percents for 7th graders in math were 10% in 2001 and 14% in 2002. In 10th grade the percents for reading were just under 70 in 2001 and 75 in 2002 or very similar to the findings for 4th grade. The 10th grade math results were just over 10% in both 2001 and 2002.

The correlation coefficients between the standards-based and norm-referenced reading scores, as well as those for math, at all grade levels suggested a moderately strong

relationship between the performance on the norm-referenced and standards-based measures. Given these relationships, equipercentile equating of the standards-based and the corresponding norm-referenced distributions were developed. Once equated, the standards-based cut-scores could be expressed in terms of their equivalent norm-referenced NPR scores. When expressed as NPRs it is clear that the reading performance cut-scores in 4th and 10th grades are at the lower end of what would be considered the normal or average range of traditional norm-referenced test performance. The math cut-score at 4th grade appears to be at the upper end of the normal range as does that for 7th grade reading. The math cut-scores for the standards-based assessment at both 7th and 10th grade appear to lie slightly above the normal range.

The use of both kinds of normative portrayals of the performance standards has proved to be helpful in understanding the magnitude of the observed inconsistencies in WASL cut-scores. This clear demonstration that performance expectations at different grade levels within content areas were quite inconsistent raised questions relevant to the credibility of the WASL assessment system. States would be well advised to be intentional in their effort to avoid such inconsistencies when they engage in standard setting activities. Better articulation of cut-scores across grade levels is advisable for the purpose of creating an assessment system that the general public and policy makers will find to be believable.

References

Lewis, D. M. (April 2002). Standard Setting with Vertical Scales. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans, LA.

MacQuarrie, D. M. and Willhoft, J. (March 2002). Validity Evidence for WASL Reading and Mathematics Performance Standards. Paper presented at the Spring Meeting of the Washington Educational Research Association, Seattle, WA.

Note from the Treasurer

Thanks to all of you who have paid your 2004 dues in advance. The checks started to arrive in November showing that some of us are way too organized. If you have not paid your dues for 2004 please use the Membership Form that is part of this newsletter and highlight any title, address or phone number changes for me to put into our NATD.ORG membership list.

Remember to share a copy of the Membership Form with any in your office who should be National Association of Test Directors Members and might be able to attend our 2004 meeting in San Diego.

The Board will review our dues structure again at the San Diego meeting. Our \$20 dues still seem to be the best deal in town. Our expenses for the annual convention, Joint Committee on Testing membership and activities, keeping up the NATD.ORG site and the newsletter are running just a bit above our dues income but we still have a little over \$10,000 in our reserve account.

I hope to do a membership mailing late in December or early in January.

Ray Fenton, Treasurer
FentonResearch@aol.com

In Memorium Dr. David S. Ireland 1937 -- 2003

The NATD Organization is saddened by the death of David Ireland.

David collapsed and died while playing soccer on Saturday, September 13, 2003.

David lived in Ottawa, Ontario for the past 31 years and was manager of Research and Evaluation at the Carleton Board of Education.

David was an active NATD member who contributed much to the organization over the years. Ray Fenton noted that David "was often the strongest voice at the 'sing-a-long' when he did not bring his trumpet."

David will be missed at future annual meetings.

National Association of Test Directors Membership Application Form

US Department of Revenue Taxpayer ID# 222659646



Please type or print neatly.

Name: Dr. Mrs. Ms. Mr. _____

Title: _____

Organization: _____

Mailing Address: _____

City: _____

State: _____ Zip: _____

Phone: _____ FAX: _____

E-Mail: _____

Check here to request that your directory information not be published on our NATD web site: _____

Membership Category: (please check one)

_____ Active Member: Responsible for educational testing programs in settings not primarily for profit.

_____ Emeritus Member: Active NATD member for at least five years and no longer employed on full time basis.

_____ Associate Member: Not directly responsible for testing programs and/or involved in test development primarily for profit.

Annual dues are \$20.00. Please make checks payable to "NATD".

Mail your check and completed application to:

Dr. Ray Fenton, NATD Treasurer

Fenton Research

8520 North Coral Ridge Loop

Tucson, AZ 85704