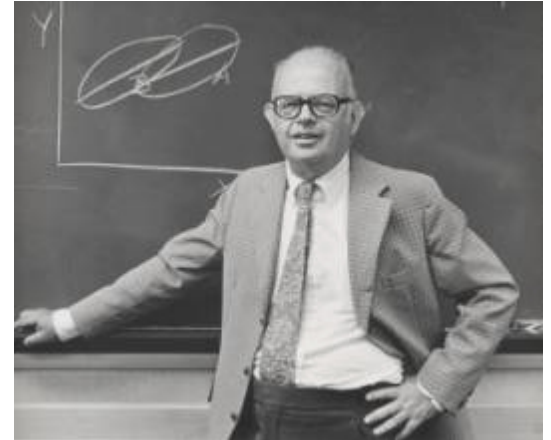




**A REMEMBRANCE OF
LEE J. CRONBACH, 1916-2001**

Edward H. Haertel
Stanford University

Lee Cronbach had just retired when I came to Stanford in 1980. We had many conversations over the years, commented on one another's papers, worked together with Lee Shulman developing prototype teacher assessments, and eventually co-authored a paper. I have a thick file of notes he sent me over the years about one thing or another, and draft copies of manuscripts covered with his comments, usually in green or purple ink. He will always be a role model for me; I can imagine none better.



For me, Lee was first and foremost a scholar. He loved ideas, especially in measurement and psychology. He'd happily look at anything I wrote, often giving me comments the next day that ran to a length greater than the few pages I'd given him. He also seemed to have a huge file of unpublished or unfinished work. When I was working on latent class models for cognitive test data, I mentioned to Lee one time that categorical classifications seemed to provide much less information than continuous scores. The next day, he gave me an unpublished paper he'd done years before, applying ideas from information theory to estimate the number of bits of information in an IQ score, and suggesting that the disparity in information was much less than I thought. I wondered sometimes if there was any problem in all of measurement that he hadn't considered.

Lee was also a role model for my teaching and advising. When I first met him in 1980, he had a stack a couple feet high of file folders for me, all neatly labeled and organized, of transparencies, worksheets, handouts, tests, and notes from his course on "Introduction to Test Theory," which I was to take over. I even inherited the ball of clay and tube of colored pick-up sticks that he used to represent vectors when he covered factor analysis. He explained to me that he tried to present each important concept three different ways, using a figure or diagram, a verbal description, and mathematical symbols. The first of many notes I received from Lee described all the students who had been working with him and had not finished, including a few who had dropped out of sight years before. He'd found new advisors for all of his current students before I arrived at Stanford, but he urged me to follow up with those students who had dropped out, whom he felt might be able to get a fresh start with a new advisor.

Lee accomplished what he did by working very hard. His multiple drafts of manuscripts were legendary. When Bob Linn, Bob Brennan, and I worked with him on our 1997 paper in *Educational and Psychological Measurement*, I could hardly keep up with his successive revisions. Our manuscript first got far too long, then gradually came back down to a length suitable for journal publication, with many ideas set aside for some future time. His file of unpublished or unfinished work was larger when we finished than when we began. Shortly thereafter, Lee's attention shifted to finishing the book that Dick Snow had been working on up until his death from cancer. Before he died, Dick accepted Lee's offer to see the book through to completion. Lee formed a group of Snow's colleagues and fulfilled his promise. The project took longer than anticipated, but Lee did get to see the prepublication copy a week or so before he died.

There's no way to describe all the ways Lee will be missed. His passing marks the end of an era, for Stanford's School of Education and for our profession.

TRIBUTE TO DR. SYLVIA TAYLOR JOHNSON

Gerunda B. Hughes, Howard University



Dr. Sylvia Taylor Johnson was truly a brilliant, dynamic, multi-dimensional woman. And yet, in spite of the reality of her death, the spirit of her work will live on in part because of the service that she rendered to students through her teaching, research and participation in professional organizations.

Teaching was one of the many things Sylvia Johnson did very well. Students who were enrolled in her classes knew that when they finished the course, they would have the knowledge and skills necessary to move on to the next level. She was demanding, but always gracious; she was firm about maintaining high standards, yet sensitive to the needs of students in meeting those high standards. No question was too "simple-minded." No thought or theory too "far out" to be given serious consideration by her.

Sylvia T. Johnson was a member of the faculty of the School of Education at Howard University for twenty-seven years. I believe her personal mission was to educate the students of Howard University, not just teach them. It should come as no surprise that mentoring of students and early-career professionals was a vital part of her modus operandi. Mentoring—at least the way Sylvia Johnson practiced it—provided access to knowledge and information and to policy and decision makers. These are the types of opportunities for students that Sylvia T. Johnson help to provide. And these opportunities, I dare say, are immeasurable.

She touched the lives of students in very measurable ways. During the last six years, specifically, over 25 graduates students were partially or fully supported financially with research dollars from projects on which she served as Principal Investigator. Without that support, many of those students would not have had an opportunity to pursue and obtain an advanced degree. For sure, she touched their lives profoundly in very real and measurable ways.

Finally, in the area of service to the broader education community, Dr. Johnson served minority student interests in both measurable and immeasurable ways. Her service on numerous policy making committees and boards will have far greater and far-reaching effects on their lives than they can even begin to imagine. The work that Sylvia Johnson did on their behalf related to equity and fairness in assessment and measurement will

indeed have immeasurable effects on their lives and their progeny.

Gerunda B. Hughes, Ph.D., is a former student and, more recently, a close colleague of Dr. Sylvia Taylor Johnson.

INSTRUCTIONALLY SUPPORTIVE ACCOUNTABILITY TESTS

*W. James Popham, Chair - Commission on
Instructionally Supportive Assessment*

State-level accountability tests, if they are to help rather than harm instruction, must assess only a state's most highly prioritized content standards so that students' standard-by-standard mastery can be determined. Nevertheless, accountability tests focused on only a modest number of high-import content standards can still be used to supply school-by-school evaluative data. These were key recommendations of a 10-member national commission whose conclusions were reported in late October.

The Commission on Instructionally Supportive Assessment was convened in the summer by five national associations of educators (AASA, NAESP, NASSP, NEA, and NMSA). The Commission's members were Eva L. Baker, David C. Berliner, Carol Camp Yeakey, James W. Pellegrino, W. James Popham (chair), Rachel F. Quenemoen, Flora V. Rodriguez-Brown, Paul D. Sandifer, Stephen G. Sireci, and Martha L. Thurlow.

(continued)

NEWSLETTER ADVISORY BOARD

BETTY BERGSTROM, Computer Adaptive Technologies, Inc.
GREGORY CIZEK, University of North Carolina, Chapel Hill
JOAN HERMANN, CRESST/UCLA
SHARON LEWIS, Council of the Great City Schools
DUNCAN MACQUARRIE, Tacoma Public Schools
KAREN MITCHELL, SRI International
LORA MONFILS, Graduate Student – Rutgers University
S.E. PHILLIPS, Consultant
NAMBURY RAJU, Illinois Institute of Technology
LAWRENCE RUDNER, ERIC/University of Maryland
STEVE SIRECI, University of Massachusetts, Amherst
LISA F. SMITH, Kean University
JON TWING, NCS Pearson
DENNY WAY, Educational Testing Service

DOUGLAS F. BECKER, EDITOR, Riverside Publishing

Send articles or information for this newsletter to:

Douglas F. Becker Phone: (800) 767-8420, ext. 7006
Riverside Publishing Fax: (630) 467-7126
425 Spring Lake Drive e-mail: douglas_becker@hmco.com
Itasca, IL 60143

The *NCME Newsletter* is published quarterly. The *Newsletter* is not copyrighted; readers are invited to copy any articles that have not been previously copyrighted. Credit should be given in accordance with accepted publishing standards.

The Commission identified nine requirements that must be satisfied if a state's accountability tests are to support instruction.

1. A state's content standards must be prioritized to support effective instruction and assessment.
2. A state's high-priority content standards must be clearly and thoroughly described so that the knowledge and skills students need to demonstrate competence are evident.
3. The results of a state's assessment of high-priority content standards should be reported standard-by-standard for each student, school, and district.
4. A state must provide educators with optional classroom assessment procedures that can measure students' progress in attaining content standards not assessed by state tests.
5. A state must monitor the breadth of the curriculum to ensure that instructional attention is given to all content standards and subject areas, including those that are not assessed by state tests.
6. A state must ensure that all students have the opportunity to demonstrate their achievement of state standards; consequently, it must provide well-designed assessments appropriate for a broad range of students, with accommodations and alternate methods of assessment available for students who need them.
7. A state must generally allow test developers a minimum of three years to produce statewide tests that satisfy Standards for Educational and Psychological Testing and similar test-quality guidelines.
8. A state must ensure that educators receive professional development focused on how to optimize children's learning based on the results of instructionally supportive assessments.
9. States should continually track progress to ensure that tests are a) appropriate for the accountability purposes for which they are used, b) appropriate for determining whether students have attained state standards, c) appropriate for enhancing teaching, and d) not the cause of negative consequences.

A report for policymakers describes and defends these requirements. A separate report contains illustrative language (addressing each requirement) that could be incorporated in a request for proposals if state officials choose to solicit external contractors to construct assessments satisfying the Commission's nine requirements. Both Commission documents are available online at:

- American Association of School Administrators
<http://www.aasa.org/>
- National Association of Elementary School Principals
<http://www.naesp.org/>
- National Association of Secondary School Principals
<http://www.principals.org/>

- National Education Association
<http://www.nea.org/>
- National Middle School Association
<http://www.nmsa.org/>

SOME RECENT AND PROVOCATIVE BOOKS

Lawrence M. Rudner, ERIC Clearinghouse on Assessment, University of Maryland, College Park

Classroom Instruction that Works: Research-Based Strategies for Increasing Student Achievement.

Robert J. Marzano, Debra J. Pickering, and Jane E. Pollock. ASCD, 178 pages.

An excellent distillation of the research literature on teaching strategies and student achievement, this book identifies and substantiates nine categories of instructional strategies that are most apt to maximize student learning. See:

<http://shop.ascd.org/ProductDisplay.cfm?ProductID=101010>.

Knowing What Students Know: The Science and Design of Educational Assessment.

James Pellegrino, Naomi Chudowsky, Robert Glaser (Eds.). National Academy Press, 382 pages.

This National Research Council Report identifies recent advances in cognitive and measurement science, explains how this expanding knowledge can serve as the foundation for improved approaches to assessment, and poses recommendations for future research, policy and practice. See:

<http://www.nap.edu/books/0309072727/html/>

Assessments in Educational Reform:

Both Means and Ends

Robert W. Lissitz & William D. Schafer (Eds.). Allyn & Bacon, 240 pages.

Well-known and respected professionals were challenged to provide an idealization of a portion of future assessments. By collectively looking forward, the editors hope this volume will both shape and speed the transformation to make assessments integral to all phases of education. See:

http://vig.pearsoned.com/store/product/1,3498,store-2924_isbn-0205332692,00.html.

Assessment: Issues and Challenges for the Millennium

Garry R. Walz, Ph.D. & Jeanne C. Bleuer, Ph.D. (Eds.). CAPS Publications, 416 pages.

This book offers a collection of papers that provide an update on new developments in assessment as well as discussions of important topics for counselor/educators, practitioners, scholars, and researchers. See:

http://ericass.uncg.edu/online_catalog.html.

REPORTING SCHOOL QUALITY IN STANDARDS-BASED ACCOUNTABILITY SYSTEMS

Robert Linn, *Center for Research on Evaluation, Standards, and Student Testing (CRESST)*

Editor's Note: This article is a shorter version of a recent CRESST Policy Brief. The full version complete with examples and tables may be found on the CRESST web site at: <http://www.cse.ucla.edu/CRESST/pages/newsletters.htm>.

Measuring School Quality—Current Status

Increasingly, states are using their accountability systems as much for measuring school status as for student achievement. According to a survey by Education Week (Orlofsky & Olson, 2001), 27 states assign ratings to schools or identify low-performing schools. President Bush's education plan (2001) would require "school by school report cards...for all public schools," and publication on the Internet. But the methods by which states rank schools or measure improvement vary greatly.

The most common method of reporting school results is in terms of current status, often done by reporting the school mean or median score for students in the grade assessed. States have been moving in recent years away from the use of scale scores and percentile ranks to the percentage of students who meet or exceed a performance standard or the percentage of students in each of several performance categories. The Florida school accountability system for example, grades schools from A to F based on current performance of students on the Florida Comprehensive Assessment Test (FCAT). Florida also reports performance in terms of improvement. But the rules are far from simple.

To receive an A, a Florida school must: "Meet grade B criteria AND the percent of students absent more than 20 days, percent suspended and dropout rate (high schools) are below state averages AND there is substantial improvement in reading AND there is no substantial decline in writing and math, AND at least 95% of standard curriculum students were tested." (Florida Department of Education, 1999). Further definitions are provided explaining both substantial improvement and substantial decline.

What appears to be a fairly straightforward A-F reporting system for current school status is considerably complex in its details. Most state accountability systems are at least this complex. Like Florida, each has different features making none of them comparable.

Measuring School Quality—Improvement Over Time

A preferable approach to measuring and reporting school achievement is to place greater emphasis on

improvement than on current status. A common method is to compare test scores between two years but for the same grade, for example, third-grade reading in 1998 to third-grade reading in 1999. Such "improvement-over-time" comparisons based on successive groups of students in selected grades are reasonable for schools with consistent student populations.

Williams School, a fictitious name but with actual school results, shows reasonable improvement in Grade 3 reading proficiency with slightly higher "proficient or above" scores in 1999 than 1998 (Table 1). The validity of inferences from such comparisons is questionable, however, for schools with rapidly changing demographics or with too few students tested in a specific grade.

Table 1: Williams School
Colorado Student Assessment Program Score
3rd Grade Reading Proficiency Levels*

Year	1998	1999
Unsatisfactory	13.13	12.61
Partially Proficient	25.25	20.17
Proficient	58.59	62.19
Advanced	3.03	3.36
Proficient or above	61.62	65.55
No scores	0.00	1.68
Total Students	99	119

*Scores are percentages of students at that level.

Longitudinal and Quasi-Longitudinal Reporting Methods

Another way to measure improvement is to track the performance of students from one grade to the next. The approach using only students with scores in both years of the comparison is commonly referred to as a *longitudinal model*. It has the appeal that the school is only held accountable for students who were in the school for the period between the first and second test administrations. Although this feature of the longitudinal approach may seem an advantage to schools, it has the clear disadvantage of excluding mobile students who change schools from one year to the next and students who for some other reason are tested in only one of the years being compared. Therefore, the educational system is not held accountable for these students.

An alternative that avoids this disadvantage is to base the accountability on a comparison of the performance of all students in the school in, say fifth grade in 2000, with that of all students in that school who are tested in the sixth grade in 2001. This approach has been called a *quasi-longitudinal* approach. It has the advantage that all students in the school at the selected grades influence the results in a given year.

Both the longitudinal approach and the quasi-longitudinal approach require comparable tests across each grade level compared. Both require annual testing in every grade used in the accountability system and are generally associated with the use of off-the-shelf tests or measures with characteristics similar to such tests. A downside to off-the-shelf tests is that they are usually not matched to state content standards. Because schools usually align instruction to the content that they are tested on, the state standards often become a lower priority.

Adjusting School Rankings for SES

Research has well documented the overriding effect of socioeconomic status (SES) on student achievement. To account for this SES factor, some states such as California and Pennsylvania report “similar schools scores” to supplement their regular school rankings. The use of background measures is controversial, however, because they imply a lower set of expectations for less affluent students. They may also mislead educators and policymakers to presume that schools are doing better than they really are or not as well as they really are.

Some educators have seized on the similar schools rankings, saying that their schools perform very well when compared to other similar schools, but failing to mention the schools’ lower results on most performance measures. Accepting or indeed promoting similar schools rankings may lead to lower expectations for students from different backgrounds. Because of the link between socioeconomic status and ethnicity, reducing the expectations for students from low SES backgrounds typically means lower standards for African American and Hispanic students.

Recommendations

No school reporting method is without some disadvantages; however, the recommendations below may enhance the likelihood that assessment systems will contribute to the overarching goal of improving student learning, while minimizing some of the potential negative effects. Although several suggestions repeat an earlier CRESST brief, they bear repetition in our high-stakes accountability environment.

1. Place more emphasis on school improvement than on current performance. This allows for differences in starting points while maintaining high standards and expectations of improvement for all.

2. Report the margin of error for any school result. All measurement systems, including polls, surveys, and even scientific tests, contain some degree of error. To avoid improper use of test scores, states should report the probability that a student or school has been misclassified.
3. As required by the Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999), evaluate the validity of the uses and interpretations of assessment results. Validate the *full* accountability system including standards, tests, alignment, professional development, rewards, sanctions, teaching quality, curriculum, and resources in addition to the positive and negative effects.
4. Validate trends with results from other indicators such as the National Assessment of Educational Progress, other state tests, and results from college admissions and placement tests such as the ACT, the SAT, and Advanced Placement tests.

Robert Linn is the co-director of the National Center for Research on Evaluation, Standards, and Student Testing, Distinguished Professor of Education at the University of Colorado at Boulder, former chairperson of the National Research Council’s Board on Testing and Assessment, and President-elect of the American Educational Research Association.

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Bush, G. W. (2001). *No child left behind. A blueprint for education reform*. Washington, DC: U.S. Department of Education. Available March 21, 2001: <http://www.ed.gov/inits/nclb/index.html>
- Florida Department of Education. (1999). *School accountability report guide: June 1999*. Available March 16, 2001: <http://www.firn.edu/doe/bin00018/guide99.htm>
- Orlofsky, G. F., & Olson, L. (2001, January 11). The state of the states. *Quality counts 2000* (Vol. 20, No. 17, pp. 86-88). Washington, DC: Education Week.



CALL FOR NOMINATIONS: BRENDA H. LOYD OUTSTANDING DISSERTATION AWARD

The National Council on Measurement in Education (NCME) requests nominations for the sixth annual Brenda H. Loyd Award for outstanding dissertation in the field of educational measurement. Nominations will be accepted for dissertations completed between July 1, 1999, and June 30, 2001. The author of the dissertation need not be a member of NCME; however, the author's advisor must be a member of NCME.

The winner of the award will receive a \$1,000 monetary award and a commemorative plaque. In addition, the advisor or committee chair for the top-rated dissertation will receive a letter of congratulations. The award will be presented at the NCME Annual Meeting in New Orleans in 2002. Also, an announcement of the award recipient will be published in the NCME newsletter. An honorable mention award may also be given; its recipient will be recognized with a certificate.

To nominate a dissertation, five copies of the following items should be submitted to the Chair of the Brenda H. Loyd Outstanding Dissertation Award Committee by January 31, 2002: (a) a letter of nomination from the author's advisor or a member of his or her dissertation committee, (b) a summary of the dissertation research (up to 10 pages), including the rationale for the study, research questions, methodology, results, and conclusions; (c) a table of contents (including a list of tables and figures); and (d) a statement from the graduate school confirming the date of completion and acceptance of the dissertation.

The criteria used by the Dissertation Award Committee include significance of the contribution to the field of educational measurement, quality of the literature review, technical quality of the research, and clarity of the writing.

Submit materials to: Jeffrey K. Smith, Graduate School of Education, Rutgers University, 10 Seminary Place, New Brunswick, NJ08901. If you have questions, please email me at jefsmith@rci.rutgers.edu.

