

MEET YOUR NEW PRESIDENT – JOHN FREMER

Our new NCME President is as likely to describe himself as a fisherman or a granddad as a measurement person, but he has actually earned his living in educational testing for forty years. Just from that sentence alone, you might guess that he is pretty old, probably likes the outdoors, and is committed to his family. All of these inferences are correct. It does not take much to get him to tell you about his last fishing or crabbing trip or one of his travel adventures with Judith, his wife and friend of forty years. (John and Judith are often accompanied by one of their eight cherished grandchildren, who range in age from six to eighteen.)



John has had an unusually broad career in testing, beginning with a stint as an item writer for the Gates-MacGinitie Reading series in the early 1960's. Since that time he has had development and management roles in tests at the local, state, and national levels, and for several different testing companies. He worked on projects related to many major testing trends during this time period, e.g., minimum competency testing, criterion-referenced testing, the growth of state assessments, the early development of the National Assessment, test disclosure, the move to CBT, and the upsurge in performance-related measurement, to name just a sample.

NCME Roles – Throughout his career, John has been an active member of NCME, including tenures as editor of *Educational Measurement: Issues and Practice* (1992-94) and as a member of the Board of Directors (1995-97). In addition he has held a number of NCME committee appointments, including chairing the publications, training, and fellowship development committees. Fremer also represented the NCME for several years on the Joint Committee on Testing Practices and chaired the work group that developed the *Code of Fair Testing Practices in Education*.

Education – John attended New York City schools and colleges, including Brooklyn College, City University of New York, where he graduated Phi Beta Kappa and magna cum laude. His undergraduate grades and graduate admissions test scores brought him a fellowship to Teacher's College, Columbia University where he earned master's and doctoral degrees in Educational Psychology and Measurement. His Ph.D. dissertation looked at the language and memory development of deaf children. Walter MacGinitie of the Gates-MacGinitie Reading Tests was John's Dissertation Chair and Robert L. Thorndike and Sam Ball were the other committee members.

ETS Employment – Fremer joined ETS directly from graduate school in 1965, working in the statistical analysis and test development areas. Over his 30+ years at ETS John has been involved in development and administrative work for tests in numerous subjects and covering a wide range of age levels, with a special emphasis on the measurement of verbal ability. Much of his work has focussed on the development and revision of national testing programs, communicating with various audiences about the strengths and limitations of tests, and the development and use of testing standards.

He currently holds a joint appointment at Educational Testing Service in the International Projects and the School and College Services areas. On the international side of his work life, he is the Director of Education of the Global Institute and works with ETS and other institute faculty to develop and offer assessment courses. Wearing his other hat, Fremer works on the Siemens Westinghouse Science and Technology Competition and the Siemens Awards for Advanced Placement.

Goals as NCME President – The next issue of the *Newsletter* will provide John an opportunity to speak directly to NCME members about his goals as NCME President. As a preview, though, it is clear that working with the Board, association committees, and our members to help people better understand testing and to promote high standards in assessment development and use will be key targets for him.

Prepared by John Fremer

NCME ANNOUNCES ELECTION RESULTS

The NCME Elections Committee is pleased to announce the following election results:

Vice President – H.D. Hoover

Board of Directors: At Large – Rebecca Zwick
 Local Education Agency – LeAnn M. Gamache

Congratulations!

LEGAL CORNER: *GI Forum v TEA*
S.E. Phillips, Michigan State University

In January, a federal district court upheld the Texas graduation test. The graduation test, which requires high school students receiving diplomas to pass subtests in reading, mathematics and writing, was challenged on adverse impact and due process grounds. The plaintiffs chose not to appeal. Newspaper accounts attributed the decision to advice from civil rights lawyers that: (1) there was virtually no chance of reversal by the appellate court; and (2) the decision would provide a weaker legal precedent for planned challenges in other states if it remained at the district court level.

The adverse impact challenge to the graduation test alleged that the test was unfair because the passing rates for Hispanic and African-American students were lower than for white students. The Court was persuaded that cumulative passing rates for seniors were the appropriate statistics for comparison but relied on inferential difference statistics applied to subpopulations of Texas minority and majority students to hold that the test had a legal adverse impact, though a positive practical effect, on minority students. The burden of proof then shifted to the state to demonstrate an educational necessity for the test. The state met this burden by demonstrating that the purpose of the test was to identify students who had not learned the state-mandated curriculum and to provide remediation and retesting. Data presented by the state indicated that passing rates for minority students had increased steadily since the graduation test was implemented and that large numbers of minority students had been successfully remediated. The burden of proof then shifted back to the challengers to provide evidence of an equally effective but less discriminatory alternative. The Court rejected proposals to allow students with high grades to pass with test scores below the passing standard or to use a sliding scale in which high performance on one subtest (e.g., reading) could compensate for low performance on another (e.g., mathematics).

Responding to the due process challenge, the Court adopted the *Debra P.* holding that a high school diploma is a property interest “*created by the requirement of compulsory education, attendance requirements, and the statute detailing graduation requirements.*” Evidence concerning test fairness, opportunity to learn (OTL), validity, reliability, passing standards, differential item performance, professional standards, and dropout/retention rates was heard by the court. The challengers argued that the graduation test did not meet professional standards because criterion and construct validity evidence was required in addition to content validity evidence, test/retest and alternate forms reliabilities should have been calculated instead of KR₂₀s, minority students had less opportunity to learn the tested

content than majority students, the test was used inappropriately as a sole criterion for graduation, test development was flawed because point biserials were used and factor analyses were not used to select items, and the test caused minority students to be retained and to dropout. The Court held that the graduation test was valid, reliable, and developed according to accepted professional standards, that the state was not required to use a variation of the discredited *Golden Rule* procedure to select items, and that a variety of nontest factors caused higher minority dropout and retention rates. The Court was also persuaded that mandatory remediation with 8 retest opportunities virtually eliminated the possibility of a false negative while giving all students with true achievement just below the passing standard a substantial probability of success.



The *GI Forum* case was similar to *Debra P.* in dealing with adverse impact and opportunity to learn for a high school graduation test. However, Texas had a state-mandated curriculum, its students subject to the graduation test had

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, Washington OSPI
KAREN MITCHELL, National Academy of Sciences
LORA MONFILS, Graduate Student – Rutgers University
S.E. PHILLIPS, Michigan State University
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
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 email: 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.

attended integrated schools, and it had over a decade of experience with the current graduation test and its predecessor. On the other hand, in the late seventies when the Florida case was heard, the state curriculum was not mandated by statute, some students had attended statutorily segregated schools for part of their education, and graduation testing was new in the state and the nation. But in both cases, the Court found that the test was a remedy for past discrimination. The Texas Court stated: “[The state’s] evidence that the [graduation test] . . . helps address the effects of any prior discrimination and remaining inequities in the system is both credible and persuasive.”

In summary, the Court held: “While the [graduation test] does adversely affect minority students in significant numbers, the [state] has demonstrated an educational necessity for the test, and the Plaintiffs have failed to identify equally effective alternatives. . . . The [state] has provided adequate notice of the consequences of the exam and has ensured that the exam is strongly correlated to material actually taught in the classroom. In addition, the test is valid and in keeping with current educational norms. Finally, the test does not perpetuate prior educational discrimination Instead, the test seeks to identify inequities and to address them.”

In its decision, the *GI Forum* court provided guidance for graduation testing programs in several important areas. The *GI Forum* court:

- upheld the *Debra P.* requirements of notice and curricular validity implicated by a property interest in a high school diploma;
- credited the professional judgment of psychometric experts who had extensive, direct experience with large-scale achievement testing;
- used reasonable, common-sense interpretations of professional standards for evaluating test quality;
- did not require perfection or mandate secondary or conditional professional standards;
- supported the use of cumulative passing rates and the 80% rule for adverse impact analyses but found legal adverse impact based on subpopulation inferential statistics;
- recognized that there was no evidence of a causal link between the graduation test and differential minority performance and that a variety of nontest factors may have contributed to the observed differences;

- found that the graduation test was not the sole criterion for receipt of a high school diploma;
- indicated that graduation test developers are not required to minimize differential performance among racial/ethnic groups, or to validate the test against criteria, such as subjective teacher grades, which measure different student attributes than those measured by the test;
- found evidence of successful remediation of minority students convincing and compelling;
- upheld passing standards for a graduation test based on all the facts and circumstances, including multiple retakes but absent evidence of research-based methodologies;
- found a graduation test developed using content and OTL procedures and criteria valid, and with reported total score KR₂₀ reliabilities by subgroup above .85 sufficiently reliable;
- found a graduation test administered only in English to be valid for all students, including those whose native language is not English;
- found notice of the graduation test 3 years prior to initial testing adequate;
- found adequate OTL on all the facts and circumstances, including successful remediation but absent formal surveys of teachers and students;
- noted that high dropout and retention rates among minority students were cause for concern but not shown to have been caused by the graduation test; and
- indicated that decisions of whether and what to test for high school graduation are the province of the legislature, not the courts.

In the *GI Forum* case, while the plaintiffs tried hard to discredit the psychometric quality of the test as a graduation requirement for individual students, they supported the continuation of the test as part of a system of school accountability. The Texas accountability system measures school test performance separately by ethnic group and schools must meet performance targets for each group separately to receive a recognized or exemplary rating. The state asserted that to ensure a common standard of achievement for high school diploma recipients, the responsibility for achieving the state curriculum must be shared jointly and concurrently by students and schools. The Court held that: “[The state] has shown that the high-stakes use of the [test] as a graduation requirement guarantees that students will be motivated to learn the curriculum tested. . . . In addition, the evidence

was unrefuted that the State had an interest in setting standards as a basis for the awarding of diplomas. The use of a standardized test to determine whether those standards are met and as a basis for the awarding of a diploma has a manifest relationship to that goal.”

Effective in 2003, the Texas graduation test will be moved to the junior year and will include academic tests in four subject areas: English language arts, mathematics, science and social studies. According to the new law, the English language arts test will cover English III and writing, the mathematics test will cover algebra I and geometry, the science test will cover biology and integrated chemistry and physics, and the social studies test must cover early American and U.S. history.

ATP CONFERENCE INTRODUCES COMPUTER BASED TESTING GUIDELINES

Betty Bergstrom, CAT, Inc.

The Association of Test Publishers (ATP) is a non-profit organization representing providers of tests and tools related to assessment, selection, screening, certification, licensing, educational or clinical uses. On February 17, 2000 in Carmel California, ATP held a conference to introduce the ATP Draft Guidelines on Computer Based Testing. These guidelines were developed to supplement the Joint Standards for Educational and Psychological Testing issued by AERA, APA and NCME (November, 1999). John Oswald, ATP President, described the conference as “A fitting way to celebrate the century: Gathering together as colleagues to explore new frontiers in the testing industry and looking ahead to challenges in the future”.

The conference featured Ron Hambleton, University of Massachusetts, Amherst as the opening keynote speaker. Barbara Plake, The University of Nebraska, Lincoln, James Olsen, Alpine Media, and Cyndy Fitzgerald, Microsoft presented a panel session designed to place the ATP CBT Guidelines into a historical context, relate them to other extant standards and provide practitioner examples of their use. Craig Mills, Association of International Certified Public Accountants, also a keynote speaker, closed the conference with a presentation entitled: “Unlocking the Promise of CBT.”

Other speakers presented technologies related to the ATP Guidelines, reviewed historical perspectives on computer based testing and discussed issues related to applications of CBT. These speakers included Tony Zara, National Council of State Boards of Nursing (NCSBN); Martha Stocking, ETS; James McBride, HUMRRO; Steven Gzesh, Microsoft Corporation; James Adair, Lotus Development Corporation; Gage Kingsbury, Northwest Evaluation Association

(NWEA); Michael Yoes, NCS; Jeff Sugerman, VUE; Betty Bergstrom and Patrick Meehan, Computer Adaptive Technologies (CAT); Eric Wenck, Sylvan Prometric; David Vale and Lauren Hebert, The Chauncey Group International; David Foster, Galton Technologies; Jamie Mulkey, Hewlett Packard Company; and J. Patrick Jones, PES.

The ATP Draft Guidelines cover Planning and Design, Test Administration, Scoring and Reporting, Psychometric Analysis and Stakeholder Communications for CBT. ATP anticipates that the guidelines will provide useful information for:

- Test development organizations
- Test publishers
- Test delivery organizations
- Test takers
- Research and evaluation specialists
- Teachers
- Technology companies
- General readers

The ATP Draft Guidelines on CBT are available for review at <http://www.testpublishers.org/>. If you have comments about the guidelines, please send them to William G. Harris, Executive Director of ATP at wgh.atp@worldnet.att.net. ATP welcomes input from members of NCME.

FULL-TEXT ERIC DOCUMENTS NOW AVAILABLE ON DEMAND, ONLINE

Lawrence M. Rudner, ERIC/University of Maryland

The ERIC Document Reproduction Service (EDRS) now offers online access to the full text of nearly all ERIC documents (shown as "ED" followed by six digits) produced since 1993! EDRS now offers *instant* electronic access to more than 85 percent of releasable ERIC documents produced since approximately 1995. No more waiting for credit card verification! Order fulfillment for older documents (produced from 1993 to 1994) that require conversion to electronic format may take two to three days.

If your host institution is an E*Subscribe customer, ERIC documents are available free of charge when you search ERIC from a registered IP address (for example, a campus address) or a valid E*Subscribe customer address (you will see a PDF icon). If your host institution is not an E*Subscribe customer, you can purchase on-demand electronic documents with a credit card or through an EDRS deposit account.

Institutional cost for E*Subscribe varies from \$960/year for a small organization to \$1,800/year for a medium sized school district to \$5,200 for a very large university. For more

information about E*Subscribe, EDRS's online subscription service for research libraries, schools, and other institutions, visit the EDRS Web site at (<http://www.edrs.com>).

You can search the ERIC database online and link to full-text ERIC documents on the following ERIC Web sites:

- * Assessment and Evaluation (<http://ericae.net>) (Full-Text is now their default database)
- * EDRS (<http://www.edrs.com>)
- * Information and Technology (<http://ericir.syr.edu/Eric>)

You can also search ERIC online and link to full-text documents through SilverPlatter's WebSPIRS and Ovid Online. If you work at a college or university, check your campus library's online offerings. For more information, call EDRS at 1-800-443-3742.

MEASURING STUDENTS' OPPORTUNITY TO LEARN AND THE QUALITY OF PRACTICE

Joan L. Herman, CRESST

The ever-growing use of high stakes, standards-based assessments demand attention to students' opportunity to learn. Do students have the opportunity to learn what is expected of them? Do they have the opportunity to engage in high quality curriculum and instruction that will prepare them to meet the rigorous standards that have been established? Currently, the answer is probably no, but how will we know whether progress in this critical domain – which surely must precede real improvement in performance – is being accomplished? Equally important, how can we provide feedback to schools that will enable them to improve?

Researchers at the national Center for Research on Evaluation, Standards and Student Testing (CRESST) believe that better measures of classroom practice are needed to answer these questions. While survey measures may be low cost and efficient, they have serious deficiencies in detecting the *quality* of practice and show little relationship to student learning. Classroom observation and intensive case study techniques, on the other extreme, are too labor intensive and costly. CRESST research, conducted by Pamela Aschbacher and Lindsay Clare, is exploring the potential of analyzing teachers' assignments and students' work to obtain more intermediate indicators of practice. Based on work conducted in language arts, their results are promising. Teachers were asked to provide a limited numbers of assignments, including a 'typical' writing, 'typical' reading comprehension, content area writing (elementary school only), and a 'challenging' major project, with accompanying samples of student work for each assignment. Assignments were rated on six different dimensions: challenge of the task, clarity of the teacher's

goals for student learning, clarity of the grading criteria, alignment between goals and task, alignment between the goals and grading criteria, and overall quality. Results show reasonable inter-rater reliability and generalizability. Relationships between assignment ratings and observation data also are encouraging. The researchers anticipate that the instrument also will serve important professional development purposes.

For further information, copies of the full CSE Technical Report 513, *Developing Indicators of Classroom Practice to Monitor and Support School Reform*, by Pamela Aschbacher may be downloaded from the CRESST web site, www.cse.ucla.edu. Or call 310-794-9140 for purchase. Validating Teachers' Assignments as an Indicator of Classroom Practice, by Lindsay Clare, is forthcoming.

IOMC CONFERENCE ANNOUNCEMENT

The Institute for Objective Measurement, MESA at the University of Chicago and the Rehabilitation Institute of Chicago will host the third International Outcomes Measurement Conference (IOMC3) on June 16 and 17, 2000. The conference will be held on the Chicago campus of Northwestern University. The Conference goal is to advance our understanding of health outcome issues, particularly in the realm of measurement. We are seeking speakers who can present data on the benefits of specific treatments, consequences of treatment, and measurement of treatments, using objective measurement (Rasch) approaches. The conference will have a health services research flavor with health outcomes being the focus.

IOMC1 and IOMC2, held at the University of Chicago in 1996 and 1998, drew nearly 100 participants. Speakers at the earlier Conferences included David Cella, Karen Cook, Anne Fisher, William Fisher, Norbert Goldfield, Carl Granger, Gunnar Grimby, Dennis Hart, Mark Johnston, Michael Linacre, Larry Ludlow, Richard Smith, Craig Velozo, Alan Tennant, Luigi Tesio, and John Ware.

A series of events are planned as part of this Conference. A pre-session on Rasch measurement software will be presented by J. Michael Linacre, Ph.D. of the University of Chicago on June 12 and 13, followed by a pre-session on objective measurement by Richard Smith, Ph.D. on June 14 and 15 at the Rehabilitation Institute of Chicago.

Continental breakfast, lunches and refreshments at breaks are included in the registration fee of \$200. To request registration information, please contact Patricia Hill at phill@rehabchicago.org (312.238.2802) or visit the IOM website at <http://www.rasch.org>.

CCSSO ANNUAL ASSESSMENT CONFERENCE

The theme of this year's conference - *A Look at Assessment: Past, Present, and Future* focuses on trends in large-scale assessment over the last 30 years. Educators at the local, state, and national levels must understand how and why assessment has evolved over time in order to develop and implement an effective large-scale assessment program. The sessions for this year's conference were very carefully selected to provide a theme-based, cohesive program.

The 30th National Conference on Large-Scale Assessment addresses continuing developments and innovations in the field of student assessment by providing an opportunity for participants to discuss and examine relevant issues, to examine the current status of research and practice in assessment, and to explore future initiatives and endeavors in the field. The conference begins with an opening luncheon session in which Pascal Forgione, Superintendent of the Austin Independent School District, will address "Trends in Large-Scale Assessment" and continues with more than 80 large and small group sessions, 2-hour symposia, and roundtable sessions addressing critical issues such as the assessment and evaluation of limited English proficient students and students with disabilities, requirements for assessment for Title I purposes, and test development and issues related to reporting of results. The sessions also address a wide range of technical concerns related to

developing new standards-based assessments, performance assessments, and portfolio assessments. Judith Johnson of the U.S. Department of Education will address "Political and Policy Issues Related to Large-Scale Assessment" at the Monday luncheon session. A breakfast session Wednesday morning begins to bring the conference to a close, providing a "futurist" perspective on large-scale assessment.

Visit the CCSSO website at <http://www.ccsso.org>.

AN INVITATION TO GRADUATE STUDENTS

The Graduate Student Issues Committee would like to invite you to participate more fully in NCME by becoming a member of GSIC. This is a great opportunity to network with graduate students and measurement professionals across the country as you work together on organizational issues such as graduate student recruitment and outreach, as well as planning for the annual NCME conference. We are looking for creative, hard-working students who represent the many areas of educational measurement. If you are interested, please contact:

Michelle Mengeling <mmengel@blue.weeg.uiowa.edu> or Lora Monfils <lmonfils@eden.rutgers.edu>, or speak with a committee member at the conference in New Orleans. And while in New Orleans, be sure to attend the GSIC sponsored sessions!
