



FROM THE PRESIDENT

Dan Eignor, Educational Testing Service

As I write this column, some five weeks have passed since the 2006 Annual Meeting in San Francisco. I continue to think about how well this meeting went, and again would like to acknowledge the work of Leslie Lukin, Chad Buckendahl and Lori Nebelsick-Gullet in putting together such an excellent program. I would be remiss if I didn't also acknowledge AERA's contributions to the success of our meeting, by scheduling our meeting at the Hotel Nikko, which provided excellent services, and for procuring a set of meeting rooms that allowed us to schedule five sessions per time slot. With the increased number of sessions, Leslie and Chad were able to accept 145 paper proposals, or 59% of the total number submitted, and 23 coordinated session proposals, or 52% of the total number submitted. With the rooms available at the Nikko, Lori was able to accept 14 proposals for pre-conference training sessions.

Finally, I'd like to acknowledge the work done by the Graduate Student Issues Committee in expanding the number of posters that could be given at the Graduate Student Poster Session, so that almost double the usual number of graduate students had the opportunity to present. This is a model we will clearly want to follow in the future.

Given our success in 2006, I am very much hoping that similar sorts of arrangements can be worked out with AERA for our annual meeting next April in Chicago.

2007 Annual Meeting News

2007 Program Co-chairs Mary Pitoniak and Mike Jodoin have been working on the program for next year and have met with me on a number of occasions to discuss issues, and, in particular, to begin making decisions on invited sessions for next year.

AERA Division D Vice President Mark Reckase and I have been in contact and are trying to work out a schedule for next year whereby the Division D Business Meeting and Reception and the NCME No-Host Reception do not occur on the same night at the same time. Next year we plan on reinstating the arrangement where the No-Host reception immediately follows the NCME Graduate Student session. Finally, I plan on starting a discussion soon with AERA and the NCME Run/Walk Coordinator, Brian French, on the possibility of moving the Run/Walk from Wednesday to Thursday morning, so that timing issues involving the Run/Walk and the Breakfast will no longer exist.

Journal Editors

As announced at the NCME Breakfast, Susan Brookhart will be serving as the new editor of *EM: IP*. Terry Ackerman, NCME Publications Committee Chair, has developed a call for the new editor of *JEM* that appears in this edition of the Newsletter and will be appearing in a number of other publications very soon. Terry hopes to have a slate of nominations in place by July 1. Both Susan and the new *JEM* editor will be actively involved in moving their journals to an electronic submission system.

NCME Committees

Each of the new committee chairs is working at present with the relevant Board liaison to fill membership slots on their committees, and also have started discussions about objectives to be accomplished in the coming year. Chairs of the Awards Committees have prepared calls for nominations for the various NCME awards, and these calls appear in this edition of the Newsletter and will be appearing in a number of other publications very soon.

Standards Committee Work

The chair of the NCME Standards and Test Use Committee, Doug Becker, has had the new committee involved in dealing with two issues: 1) preparation of a reaction statement to a set of proposed psychological testing standards submitted for international endorsement to the International Organization for Standardization (IOS) and sent to NCME by the American National Standards Institute (ANSI), the U.S. member of ISO; and 2) preparation of a document providing information on the

process NCME will use in reaching a recommendation about whether the *Standards for Educational and Psychological Testing* require revision. This document will be sent to the Management Committee for consideration.

Upcoming Board Meetings

The NCME Board of Directors is scheduled to meet August 18 and 19 in Madison, WI and on November 10-12 in Washington, D.C. Part of the November meeting will be devoted to long-term planning, and, in particular, to a discussion of ways in which NCME might increase its level of visibility in the public policy arena. If you have any matters you'd like to have discussed at these meetings, please send them to me via e-mail (deignor@ets.org).

The Burning Breakfast Question

After being duly embarrassed by not being able to respond to Ed Haertel's question about the NCME budget at the Breakfast, I have followed up and now have what I think is a suitable answer to Ed's question. Briefly, Ed wanted to know why NCME total revenue from publications had dropped from \$165,300 in 2005 to \$71,000 for 2006. The answer is that NCME institutional subscriptions are no longer paid directly to NCME; they are now paid to Blackwell and become part of the revenue from which our royalty is computed. A line item for royalty revenue did not appear on the spreadsheet that was handed out at the Breakfast, which made the difference in revenues very apparent.

2007 AWARD WINNERS

Barbara Plake Receives the NCME Award for Career Contributions to Educational Measurement

Robert J. Mislevy and Suzanne Lane, Chairs

The 2006 NCME Award for Career Contributions to Educational Measurement was presented to Barbara S. Plake in recognition of her substantial impact on the visibility and validity of testing practice through her research contributions, tireless service to the profession, and extensive activities in applying and evaluating tests herself and providing tools for practitioners.

Dr. Plake has expanded the scope, frequency, and comprehensiveness of the activities of the Buros Institute of Mental Measurements, including increasing the frequency of the widely-used *Mental Measurement Yearbook* series and broadening test evaluation services to educational and professional assessment. A past President of NCME, she is also a co-founder of *Applied Measurement in Education*. Dr. Plake has authored more than a hundred research articles and chapters in computerized testing, standard-setting methods, and differential item functioning and order effects. Her collaboration, mentoring, training, and advice have deepened the understanding and improved the practice of countless students, measurement colleagues, and public, professional, and scientific organizations that use testing in society at large.

Comments from Barbara Plake

It is indeed an honor and a delight to be the 2006 recipient of the NCME Career Award. As I look back over my career I see three major themes that I believe characterize my research and career interests. These themes are: a) factors that might affect valid assessment results, b) factors that could influence valid interpretations of assessment results, and c) improving the technical quality of tests and testing practices.

My first area of interest was in factors that might influence the validity of assessment results. Today we would identify these as factors that contribute to construct-irrelevant variance. In the days when I was focusing on these factors, I was concerned about how the structure of the assessment itself could interfere with an examinee's ability to demonstrate his or her knowledge or achievement. Among the many factors that have been studied that could distort the correct interpretation of examinee test performance, my main interest was in the physical ordering of the items. It occurred to me that some examinees might be disadvantaged by the sequence the test questions, especially when the test questions were ordered from easy to hard, as was the case for most standardized tests at that time. My principal interest was in mathematics assessments and I hypothesized that female examinees, who may have less confidence in their mathematics ability, might be discouraged by this sequence of items. So I prepared test booklets that contained the same mathematics test questions but arranged them in three orders: easy-to-hard, spiral cyclical, and random. The spiral-cyclical order had the items progressively becoming more difficult, but within each 4-item sequential cycle items of lower, moderate, and higher difficulty appeared. Females performed significantly better on these mathematics questions when they were ordered in a spiral cyclical manner and significantly worse when they were ordered from easy-to-hard. This was the start of my interest in how item presentation and delivery could influence the validity of assessment results.

Next I turned my attention to the delivery of test questions via a computer, especially with adaptive testing where the item-selection algorithm was designed to systematically present to examinees items that were tailored to their ability level. I had already learned that the sequence of item ordering could impact test results for some examinees in mathematics. I wondered if having the computer select the items, and the order of their delivery, would likewise have an effect on test results. My colleagues and I looked at a number of variables in this program of research including whether examinees would do better if they had the option of making their own item difficulty selection decisions, as opposed to allowing the item-selection algorithm to make this choice for them. Examinees were informed that they would receive more credit for answering difficult items correctly than they would for answering less difficult items correctly, but were asked as they progressed through the test whether they wanted a difficult or easy item to be delivered. By allowing some students the option to self-tailor the item difficulty sequence, we learned that these students tended to perform better overall on the test.

Continuing the theme of addressing validity of score results, I next turned my attention to ways to improve the validity of score interpretations, especially for tests where pass/fail decisions were being made based on examinees' scores on a test. At the time, most of these high-stakes tests were composed of multiple-choice questions and the methods of choice by nearly everyone doing standard setting studies were variations of the Angoff Standard Setting Method. When using this method, subject matter experts are asked to predict the probability that a randomly drawn, hypothetical minimally-competent candidate would be able to answer the questions correctly. These estimates were made for each item in the test, often twice with performance data presented in between rounds of ratings. This judgmental process was implemented without much question by most of the major testing programs for licensure and certification. The task for the judges seemed overly complex and convoluted, so my colleagues and I decided to try some variations on this approach. We studied whether the performance data influenced the panelists' judgments and if so were these influenced judgments more valid for making pass/fail decisions (they were) and whether simplified methodologies could make the panelists' judgmental process more reasonable (it did). Around this time, some of these assessment programs, primarily in education, added to the complexity of making score interpretations by adding more performance categories (Basic, Proficient, Advanced, for example, not just "pass/fail") and by adding more complex item types to their assessments. This started another line of research to develop new methods that could be used for complex assessments with multiple cutpoints or performance categories.

My third area of interest relates to the improvement of tests and testing practices. When I became the Director of the Buros Institute of Mental Measurement, the motto of the Institute was to "Continue to Proud Tradition" that had been started by its founder, Oscar K. Buros. Ten years before I became the director, James Mitchell had taken the leadership role for the Buros Institute upon the death of Oscar Buros and the successful competition by the University of Nebraska-Lincoln to become the new home for the Buros Institute of Mental Measurements. I became the director in 1988; the time was right to change our perspective from continuing the work of the founder to looking toward ways to improve the science and practice of testing. Over the next nearly 20 years, we have implemented a policy to only review tests in the *Mental Measurements Yearbook* that have adequate descriptive information upon which to base a review, standardized the reviews so they follow a systematic framework to improve their communication, increased the frequency of the print-based publications and embarked on a program to make test reviews available electronically through electronic databases and web-based applications, and increased the scope of our work by adding a division that focuses on non-commercially available tests and testing products. Now called the Buros Center for Testing, two Institutes comprise the Center: the long-standing Buros Institute of Mental Measurements (focusing on the technical quality of commercially available tests) and the Buros Institute for Assessment Consultation and Outreach (focusing on improving testing practices in all areas of assessment).

Throughout my career I have been blessed to be able to work with many students and colleagues. Most notably are my collegial relationships with Ron Hambleton, Gerald Melican, Steven Wise, and James Impara. During my career I have had moments when I cried tears of joy and sorrow. The sorrow was most profound over the deaths of my first husband, Donald H. Plake, my best friend, Brenda H. Loyd, and my exceptional colleague, Richard M. Jaeger. Although my heart still aches from their loss, my life has been enriched by the time and experiences we shared together. Also, throughout my life my father, Dr. John K. Sterrett, has served as an inspiration to me. He mentored me, supported me, and believed in me. It is to his memory that I dedicate this NCME Career Award.

Tim P. Moses receives the NCME Alicia Cascallar Award for an Outstanding Paper by an Early Career Scholar

Neil Dorans, Chair

Tim P. Moses of the Educational Testing Service won the Alicia Cascallar Award for his 2005 NCME paper, "Using kernel equating to check the statistical equivalence of nearly identical test editions."

Comments from Tim Moses

I am honored to receive the Alicia Cascallar Award for my 2005 NCME paper, "Using kernel equating to check the statistical equivalence of nearly identical test editions." My paper considered two of several previously-proposed procedures to check for item-order effects across test forms. The two procedures focus on different aspects of test scores but were both developed from

the perspective that item-order effects are primarily an equating problem (i.e., ensuring fairness among test forms containing the same items in different orders). The kernel equating framework was useful for integrating the procedures and also for extending them beyond their original proposals so that they not only would evaluate item-order effects in terms of a yes/no question, but would also suggest the appropriate equating adjustment for reducing item-order effects on scores. This framework was applied in an investigation of item-order effects for two exams and shown to be useful for assessing and correcting the impact of item-order effects on the score classifications of most interest to the testing program. I thank the following people for their help, collaboration and encouragement in writing this paper: Christine Wilson, Wen-Ling Yang, Neil Dorans, Alina von Davier, and Paul Holland. I also thank the College Board for generously allowing me to use their test data.

Yanmei Li receives the NCME Brenda H. Loyd Outstanding Dissertation Award

Cherly Cardell, Chair

The 2005 Brenda H. Loyd Award for outstanding dissertation work in the field of educational measurement is given to Dr. Yanmei Li for her dissertation entitled, "Applications and Extensions of the IRT Testlet Models." Dr. Li completed her research at the University of Wisconsin-Madison's Department of Educational Psychology, and her dissertation advisor was Dr. Daniel Bolt. The criteria used by the award committee included the 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.

Dr. Li is currently employed as a research scientist at the American Institutes of Research in Washington, D.C., where she continues her work on educational measurement and item response theory.

Comments from Yanmei Li

When tests are made up of testlets, standard IRT models are often not appropriate due to the local dependence present among items within a common testlet. Several testlet-based IRT models have recently been developed to model examinees' responses under such conditions (e.g., Bradlow, Wainer, & Wang, 1999). However, practical applications of these models in real testing contexts require development of new methods that permit the linking of multiple forms and/or subpopulations. My dissertation extends applications of testlet models by developing methodology that permits linking of testlet IRT calibrations, as well as the use of testlet IRT models to study differential item functioning (DIF). It also introduces methods for comparing different testlet models as a way of investigating how passage related factors influence item performance.

My dissertation consists of three studies. In the first study, a new test characteristic curve (TCC) method is presented to link calibrations based on the testlet model. Procedures for calculating the TCC linking coefficients are developed. The effectiveness of the method is investigated using simulated data. Results suggest that the linking coefficients are accurately recovered and superior to those obtained when using a traditional IRT model to link calibrations.

In the second study, a multiple-group testlet model is introduced and a marginal-maximum-likelihood estimation procedure is derived. The multiple-group testlet model can be applied to problems such as DIF. DIF is likely to occur in testlet-based tests due to the passage factors introduced by testlets. In this respect, the multiple-group testlet model offers greater potential in studying the causes of DIF in test items. A DIF detection procedure is proposed by which DIF can be attributed either to testlet factors or other characteristics unique to the item. It is argued that such studies may have important implications for understanding the occurrence of DIF in testlet items and thus assist in developing items for new testlet-based tests so as to minimize the effects of testlet-based factors on DIF.

Finally, in the third study methodology for comparing different testlet models is considered. The Bradlow et al. (1999) model introduces separate testlet factors to account for local dependence and applies a common item discrimination parameter to both the general ability and testlet factor. In this study, several alternative ways of accounting for local dependence that make different assumptions regarding the influence of testlet factors on item performance are proposed. To compare these models, several real test datasets are analyzed using Bayesian model selection criteria. Results suggest that an alternative model in which separate discrimination parameters are applied to the general ability and testlet factors provides a better fit to these data despite its greater complexity. As in the DIF study, it is argued that the results of this study may be of practical value in test construction.

Won-Chan Lee receives the NCME Bradley Hanson Award for Contributions to Educational Measurement

Deb Harris, Chair

The winner of the NCME 2006 Bradley Hanson Award for Contributions to Educational Measurement was Dr. Won-Chan Lee, of the Center for Advanced Studies in Measurement and Assessment (CASMA) at the University of Iowa. The award is given in support of a project that promises to make a significant contribution to the field. Dr. Lee's project, "Integrating Procedures and Software for Estimating Classification Consistency," has the goal of summarizing procedures for estimating classification consistency (both IRT and classical methods), creating integrated software to run all the procedures, and

disseminating this information to the measurement community. Dr. Lee has worked in the area of classification consistency for sometime, including co-authoring a paper with Brad Hanson (and Bob Brennan) on the topic. Dr. Tony Thompson and Dr. Robert Brennan provided letters of support.

Comments from Won-Chan Lee

I had the opportunity to work with Bradley Hanson on a paper entitled “Estimating consistency and accuracy indices for multiple classifications” (Lee, Hanson, & Brennan, 2002) published in *Applied Psychological Measurement*. Young researchers often sought Brad out as a collaborator, and I remember that his name appeared on nine conference papers at one NCME/AERA convention. He represented my ideal of a scholar and it’s a great delight and honor to have received the 2006 Bradley Hanson Award in his memory.

The purpose of my project entitled “Integrating procedures and software for estimating classification consistency” is: (a) to summarize procedures developed so far for estimating classification consistency; (b) to create integrated computer software for implementing those procedures; and (c) to communicate these issues to various measurement communities. Research on classification consistency has been on-going for about 30 years. No current document exists that systematically discusses the many approaches that have been developed. Furthermore, publicly available computer programs do not exist for many of the procedures, and there are no integrated and well-documented computer program packages that, in total, permit users to apply all of the available procedures. Hopefully, this project would remedy these problems. As an outcome of this project, a monograph will be written, which summarizes various IRT and non-IRT procedures for estimating classification consistency, and provides practical guidelines for users. Also, several computer programs will be created with written manuals. A brief summary of the estimation procedures and computer software that have been developed is provided below.

For the sake of convenience, the various estimation procedures are categorized here into two broad categories, IRT and non-IRT approaches. Non-IRT procedures for tests consisting of dichotomously-score items include Huynh (1976), Subkoviak (1976), and Hanson and Brennan (1990). When the test is composed of items more complex than dichotomous ones, the following non-IRT procedures can be used: Livingston and Lewis (1995); Breyer and Lewis (1994); Woodruff and Sawyer (1989); Brennan and Wan (2004); and Lee (2005a). IRT procedures include Huynh (1990); Schulz, Kolen, and Nicewander (1999); and Lee, Hanson, and Brennan (2002) for dichotomously-scored items, and Wang, Kolen, and Harris (2000) for polytomously-scored items. For tests that consist of mixtures of dichotomous and polytomous items, the procedure discussed by Wang et al. (2000) can be extended directly using mixed IRT models.

Some of the procedures mentioned above have been implemented in the following computer programs. Class Consistency (Hanson, 1995) computes classification consistency as discussed in Hanson and Brennan (1990). BB-CLASS (Brennan, 2004) is more general than Class Consistency and can also compute results for the Livingston and Lewis (1995) procedure. MULT-CLASS (Lee, 2005b) computes results based on the multinomial and compound multinomial models as discussed in Lee (2005a). Finally, IRT-CLASS (Lee & Kolen, 2006) is intended to be used for tests that are scaled using dichotomous, polytomous, or mixtures of different IRT models. These computer programs are available from CASMA website: www.education.uiowa.edu/casma.

Wainer, Bradlow and Wang Receive NCME Annual Award for Technical or Scientific Contributions to the Field of Educational Measurement

Barbara Dodd, Chair

Drs. Howard Wainer, Eric Bradlow and Xiaohui Wang received the NCME 2006 Annual Award for Technical or Scientific Contributions to the Field of Educational Measurement. Selection committee chair Barbara Dodd presented the award to Wainer, Bradlow, and Wang at the annual meeting for “their development of Testlet Response Theory.”

Comments from Howard Wainer:

IRT is a family of models that tries to describe, in a stochastic way, what happens when an item meets an examinee. To do this it makes some assumptions about the character of the response process. Until recently, all IRT models had conditional local independence as one of their primary requirements. As evidence accumulated and tests were being asked to enable more delicate inferences, it became clear that the assumption of local independence was often untenable. Test developers were finding that with increasing frequency it was important to build a test out of units that were larger than a single item. Wainer & Kiely (1987) argued for such a model, proposing the testlet as the unit of test construction. The testlet was meant as an extremely flexible unit. While it was originally proposed as an aggregation of items that were grouped together to act as a unit, the concept could encompass testlets with but a single item, or, at the other extreme, a test could be made up of a single testlet. Once the concept of the testlet was suggested, a number of alternative psychometric models were proposed and implemented.

Over the almost 20 years since the testlet was first proposed, Howard Wainer and his colleagues, in a series of more than a dozen papers have built testlet response theory. Over the past six years, Wainer, in collaboration with Eric Bradlow and Xiaohui Wang, developed a general testlet model embedded within a fully Bayesian framework. In addition to the technical developments that they have published they also have prepared the software package SCORIGHT that implements all of these developments. At this time they are currently completing a book that contains a coherent description of all of this work, and are just now finishing off work that will allow the general testlet model to contain covariates.

Sandip Sinharay receives Jason Millman Promising Measurement Scholar Award

Karen Mitchell, Chair

The 2006 Jason Millman Promising Measurement Scholar Award was presented to Dr. Sandip Sinharay of the Educational Testing Service in recognition for exemplary early scholarship. As a statistician, Dr. Sinharay brings a unique perspective to his work in educational measurement. His work with Bayesian hierarchical models, differential item functioning statistics, and other psychometric tools hold great promise for the measurement community. Dr. Sinharay is an impressive and highly productive measurement researcher, applied measurement practitioner, and educator. The NCME leadership is proud to recognize Dr. Sinharay's early contributions to the field of measurement and look forward to his future contributions.

Comments from Sandip Sinharay

It is a great honor to receive an award named after Professor Millman, a scholar extraordinaire and a perfect gentleman.

My primary research topic has been the application of a Bayesian model checking tool to item response theory (IRT) models. I am confident that researchers in educational measurement, even those not currently using Bayesian methods, will find interest in this research, especially the graphical displays assessing model misfit. In my work I have attempted to convince applied-measurement practitioners to: 1) be careful before employing any model fit technique and judge if the technique is good enough, 2) check several aspects of model fit in applied applications of IRT, and 3) apply model fit techniques to real data sets more often and assess the practical significance of model misfit (a recommendation made by Professor Ronald Hambleton).

Another area of my research involves test fairness. I, with colleagues like Neil Dorans, have employed a Bayesian approach to small-sample DIF estimation that uses information from past items that are similar to those appearing in a current test. The Bayesian approach offers improvement over existing methods and provides some support to the current practice of not performing DIF when sample sizes are small for tests with a proven record of having few items with substantial DIF. I would like to continue pursuing research utilizing existing data on past tests to improve the quality of statistical inference in applied measurement problems.

I am also performing research comparing post-stratified equating and chained-equipercenile equating in the context of the non-equivalent anchor test (NEAT) design, and in particular, the characteristics of anchor tests in NEAT. Ongoing research by myself and Paul Holland shows that the performance of an anchor test in equating that is proportionally representative of the two tests to be equated with respect to content, but consists only of medium difficulty items, may be just as good as that of the "mini-test," widely believed to be the optimum anchor test. If successful, this research will have practical applications because medium-difficulty items may have lower costs than other items.

Recently, the legendary statistician Prof. C. R. Rao (a member of the U. S. National Academy of Science and the recipient of the National Medal of Science in 2002) requested that I help him with the publication of a volume of the *Handbook of Statistics* (http://www.elsevier.com/wps/find/bookdescription.cws_home/709310/description#description) on Psychometrics. The book, to be published in the later half of 2006, will have a broad coverage, and is aimed to be a guide for research scholars, reading materials for graduate students and source materials for practitioners. Several exceptional educational measurement professionals have contributed chapters to this volume. There will be a chapter on "future challenges in psychometrics" with short articles by several prominent researchers and vice presidents from Research and Development divisions of leading testing companies. This will greatly help researchers looking for challenging measurement problems.

Finally, I have performed research on subscore, improvement of the statistical estimation techniques used in the National Assessment of Educational Progress (NAEP), application of a Bayesian hierarchical model to calibrate automatically generated items, assessment of convergence of the Markov Chain Monte Carlo (MCMC) algorithm, and application of time-series analysis to predict number of examinees who will take a future test. I hope to continue to serve educational measurement, a field that appears even more fascinating than it did when I entered it about five years ago.

APA, AERA, AND NCME BEGIN DISCUSSIONS ABOUT REVISING THE 1999 STANDARDS FOR EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT

Barbara Plake, Professor Emeritus, University of Nebraska-Lincoln

The Presidents of the three sponsoring organizations have each appointed one member to serve on the Standards Management Committee (SMC). The three members are Barbara Plake (NCME), Suzanne Lane (AERA) and Wayne Camara (APA). All three members are also members of NCME. Wayne Camara will serve as chair of the Standards Management Committee, serving a 5 year term; Suzanne Lane will serve a four-year term, and Barbara Plake's term will be for three years. All subsequent appointments to the committee will be for 5-years.

At the initial meeting of the Standards Management Committee a decision was made to seek input from the three sponsoring organizations about whether they believed it was time to initiate a revision of the *Standards*. Towards that end, all three organizations were sent a letter asking for a response by October 15 about their recommendation about whether to initiate the revision process.

The roles and functions of the Standards Management Committee (SMC) have expanded from what they have been historically. Most notably, it is now the responsibility of the SMC to appoint the members of the Joint Committee (including chairs). Another new role is to create the charge, financing and terms of the Joint Committee in consultation with the CEOs from AERA, APA, and NCME. In addition to these new roles, the SMC will continue to:

- Determine pricing of the *Standards*
- Determine number of advertisements for the *Standards* in organizational publications
- Approve unbudgeted expenses
- Oversee Finances/Development Fund
- Approve staff of Publisher dedicated to the *Standards*

The Standards Management Committee decided to raise the price of the 1999 *Standards* effective July 1, 2006 as follows:

Students	\$25.95 (no change)
Members	\$35.95
<u>Non Members/Institutions</u>	<u>\$49.95</u>
10% discount for order quantities larger than 10	

This pricing structure was decided in response to increase in publication costs for reprinting the 1999 *Standards*. Proceeds from the sale of the *Standards* go into the Standards Development Fund which will be used to support the revision.

If the decision is made to revise the *Standards* the SMC will immediately begin procedures for identifying the members of the Joint Committee and implementing mechanisms for getting members input on matters related to the revision. Typically, after a decision to revise the *Standards* is made, it takes 6 – 7 years to complete the revision. So that means that the revised *Standards* would be published about 2013.

NCME WEB SITE CHANGES

M. David Miller, NCME Web Site Editor

The NCME Web Site will be making substantial changes starting this summer. The Rees Group, Inc. will be working with us to completely redo the Web Site. The most significant change will be adding a Members Only section. Under the Members Only section, members would have password protected access that would:

- Allow online Journal access to Blackwell's Synergy system;
- Allow members to update contact information;
- Allow new members to join and current members to renew online with a credit card;
- Allow members, during the renewal cycle, to complete a survey indicating interest and willingness to participate in NCME activities;
- Allow members to vote on NCME issues and conduct Board of Director elections through the website; and

- Allow members to access an online membership directory.

These changes will make registration, journal access, and voting more efficient. A committee from NCME chaired by M. David Miller will work with Rees Group, Inc. during the process of rebuilding the Web Site.

CALL FOR 2007 AWARDS

Call for the NCME Award for Career Contributions to Educational Measurement

Description of the Award

The award honors living persons whose publications, presentations, and professional activities over a career have had a widespread positive impact on the field of educational measurement. These contributions may include theoretical or technical developments, service to professional organizations, conceptualizations of educational measurement that have enhanced public understanding of measurement problems, applications of theory that have influenced the nature of educational tests and measurement, or innovative ideas that have significantly affected measurement practices. Award recipients receive a check for \$1,000 and a commemorative plaque from NCME. In addition, recipients are invited to provide an address at the next year's NCME Annual Meeting.

Description of a Nomination

A nomination must include two items: (a) a 1-or-2 page summary of the nature, significance, and impact of the nominee's contribution to the field of educational measurement, and (b) an up-to-date copy of the nominee's vita.

Application Procedure

Submit eight (8) copies of all materials by November 30, 2006 to:

Bruno D. Zumbo, Ph.D., Professor
Department of ECPS
University of British Columbia
Scarfe Building, 2125 Main Mall
Vancouver, B.C.
CANADA V6T 1Z4

For Additional Information: Go to the NCME Website or e-mail Bruno Zumbo at bruno.zumbo@ubc.ca

NCME Award for Outstanding Dissemination of Educational Measurement Concepts to the Public

Description of the Award

In the year 2007, NCME will honor outstanding dissemination of educational measurement concepts to the public in 2004, 2005, or 2006. Examples of past awards that have been made in this category include: James Mitchell, Buros Institute, the Australian Council for Educational Research, *The Seattle Times*, the Admission and Guidance Services Division of the College Board, and Larry Rudner. NCME members and others are invited to identify candidates for this significant award. Selection criteria for the award will include quality, innovation, and importance of the contribution.

Description of a Nomination

Self nominations are encouraged as are nominations for others. Individuals or groups are eligible for this award. Nominees need not be NCME members. A nomination consists of 7 copies of a 3-5 page statement summarizing the technical or scientific contribution as well as an electronic version of the statement. Applicants should clearly describe and demonstrate the importance of the contribution to the field of educational measurement. Additional supporting documentation is welcome. Applications should include the names and addresses of two persons familiar with the specific application and its results. The committee may request further materials and may contact others who are likely to be able to evaluate the contribution.

Application Procedure

Nominations should be sent by December 8, 2006, to:

Xiang Bo Wang
Research and Analysis
The College Board
45 Columbus Avenue
New York, NY 10023-6992, USA.
Phone: 1-212-713-8287
Fax is 1-212-649-8427
Email: Bwang@CollegeBoard.org

For Additional Information: Go to the NCME Website or e-mail Xiang Bo Wang at Bwang@CollegeBoard.org

Call for the Alicia Cascallar Award for an Outstanding Paper by an Early Career Scholar

Description of the Award

Alicia Cascallar, who published most of her research as Alicia P. Schmitt, is best remembered for her work in the area of differential item functioning. The Alicia Cascallar Award for an Outstanding Paper by an Early Career Scholar has been established to honor Alicia's professional commitment and accomplishments and to continue her practice of mentoring promising new scholars in the area of educational measurement. The award consists of \$500, a commemorative plaque from NCME, and a waiver of NCME conference fees. The award will be given to an early career scholar who presented an outstanding paper at the 2006 NCME Annual Meeting.

Description of a Nomination

To be eligible for this award, the individual must have presented a paper at the 2006 NCME Annual Meeting. The author must have received his/her doctoral degree within 5 years of the Annual Meeting. In addition, professional colleagues who believe that this work represents a significant contribution to the field of applied measurement must endorse the paper by writing a letter to the Committee Chair.

Application Procedure

Nominator will submit his/her nomination letter and the curriculum vitae of the candidate to the Committee Chair. The letters of other endorsees of the paper should also be submitted to the Committee Chair. The candidate should submit the version of his or her paper submitted to the discussants for the 2006 meeting paper electronically to the Committee Chair. Materials should be submitted by November 1, 2006 to:

Mei Liu
Center for Statistical Analysis — MS 08P
Educational Testing Service
Rosedale Road
Princeton, NJ 08541
Attention Alicia Cascallar Award

For Additional Information: Go to the NCME Website or e-mail Mei Liu at mliu@ets.org

Call for the Bradley Hanson Award for Contributions to Educational Measurement

Description of the Award

The Bradley Hanson Award has been established to honor Brad Hanson's contributions to the field of educational measurement and to further advance the goals embodied in his work. Nominees must propose a project that promises to make a significant contribution to the field of educational measurement and/or to the development/instruction of new professionals in the field. A typical time frame for the completion of the project is one to two years. The recipient will be awarded \$1,000 and a commemorative plaque from NCME.

Description of a Nomination

Nominees may be non-NCME members. However, the recipient must be an NCME member by the time the award is presented at the Annual Meeting in 2007. The award is open to individuals at any stage in their careers, including graduate students, and may be awarded to either a single individual or a group of individuals.

Nominations for the award must include: (1) a letter of nomination. The nomination letter must describe the candidate, the specific project to be supported and timeline, including how the \$1,000 will be spent, and should be 1500 words or less. Self-nominations are welcome; (2) the candidate's curriculum vitae; and (3) at least one additional letter of recommendation addressing the qualifications of the candidate and the importance of the project.

Application Procedure

All materials should be received by November 1, 2006. The nomination letter and curriculum vitae may be submitted electronically to jeeseonkim@wisc.edu. The recommendation letter(s) should be sent as hard copies by regular mail to the following address:

Jee-Seon Kim
Department of Educational Psychology
University of Wisconsin at Madison
1025 West Johnson Street
Madison, WI 53705
Attention: NCME Bradley Hanson Award

For Additional Information: Go to the NCME Website or refer to Brad Hanson's homepage at <http://www.b-a-h.com/> or e-mail Jee-Seon Kim at jeeseonkim@wisc.edu.

Call for the Brenda H. Loyd Outstanding Dissertation Award

Description of the Award

The Brenda H. Loyd Award honors an outstanding dissertation in the field of educational measurement. The winner of the award will receive \$1,000 and a commemorative plaque from NCME. In addition, the advisor or committee chair for the award-winning dissertation will receive a letter of congratulations.

Description of a Nomination

Nominations will be accepted for dissertations completed between July 1, 2004, and June 30, 2006. The author of the dissertation need not be a member of NCME. However, the author's advisor must be a member of NCME. Nominations must include: (a) a letter of nomination from the author's advisor; (b) a summary of the dissertation research (up to 10 double-spaced pages), including the rationale for the study, research questions, methodology, results, and conclusions; (c) a table of contents (including a list of table and figures); and (d) a statement from the graduate school confirming the date of completion and acceptance of the dissertation.

Application Procedure

Submit seven copies of all materials by Friday, November 17, 2006 to:

Mark J. Gierl
Centre for Research in Applied Measurement and Evaluation
6-110 Education North, Faculty of Education
University of Alberta
Edmonton, Alberta, CANADA T6G 2G5

For Additional Information: Go to the NCME Website or e-mail Mark Gierl at mark.gierl@ualberta.ca

Call for the Jason Millman Promising Measurement Scholar Award

Description of the Award

The Jason Millman Promising Measurement Scholar Award is intended to honor the lifetime work of Dr. Millman, to recognize his contributions to the field of applied measurement, and to continue Dr. Millman's support of scholars in their formative years who are just beginning their research careers. The award recognizes a scholar at the early stages of his/her career whose research has the potential to make a major contribution to the applied measurement field. The winner of the award will receive \$1,000 and a commemorative plaque from NCME.

Description of a Nomination

To be eligible for the award, the candidate must have received his/her doctorate within the last five years. The nomination must include: (a) a letter of nomination from a professional colleague who is a member of NCME; (b) at least two letters of recommendation (from persons other than the nominator) that speak to (1) the candidate's contributions to the field of measurement as a teacher, and/or as an applied measurement practitioner, and/or as a measurement researcher, and (2) the reasons for which the candidate's work represents a significant contribution to the field of applied measurement; (c) two or more unique papers presented at any of the last 5 NCME annual meetings, or published in the last 5 years in an NCME publication; (d) the candidate's curriculum vita; and (e) a letter from the candidate outlining his/her career goals and how his/her work contributes significantly to the field of educational measurement.

Application Procedure

Submit 7 copies of all materials by November 4, 2006 to:

Kadriye Ercikan
ECPS, Faculty of Education
University of British Columbia
2125 Main Mall, Vancouver, BC V6T 1Z4
Canada

For Additional Information: Go to the NCME Website or e-mail Kadriye Ercikan kadriye.ercikan@ubc.ca.

CALL FOR NOMINATIONS: JEM EDITOR

The Publications Committee is soliciting nominations for the editor of the NCME publication *Journal of Educational Measurement* (JEM). The next JEM editor will be responsible for issues appearing between January 2008 and December 2010. The NCME Publication Committee will screen nominations. The Committee offers a slate to the president, who, in turn makes a recommendation for appointment to the NCME Board of Directors. This appointment is designed to provide a smooth transition between the incoming and outgoing editors.

We are seeking an editor that would help move *JEM* to an online format: that is submissions and reviews would be entirely online.

JEM is a quarterly journal that publishes original measurement research, reports on new measurement instruments, provides reviews of measurement publications, and reports on innovative measurement applications. The topics addressed will interest those concerned with the practice of measurement in field settings as well as be of interest to measurement theorists. In addition to presenting new contributions to measurement theory and practice, *JEM* also serves as a vehicle for improving educational measurement in a variety of settings.

If you are interested in this position, if you would like to nominate a colleague, or if you would like additional information, please contact Terry Ackerman, NCME Publications Committee Chair by email: taackerm@uncg.edu, or by phone: 336-334-3474. Deadline for nominations is August 15, 2006.

EVENTS OF REGIONAL INTEREST

The 22nd Annual Washington State Assessment Conference

The Washington Educational Research Association and the Office of Superintendent of Public Instruction present the 22nd Annual Washington State Assessment Conference: Preparing ALL Students for the 21st Century. The event will be held from December 7-8, 2006 (Pre-conference December 6) at the Seattle Airport Hilton Hotel Conference Center (17620 Pacific Highway South).

The keynote presenters will be Kati Haycock on Thursday and Pedro Noguera on Friday. Kati Haycock is the director of the Washington DC-based Education Trust. The Trust advocates for youth, and provides hands-on assistance to educators working to improve student achievement at all levels. Pedro Noguera is an urban sociologist, and professor at New York University. His work has focused on ways that schools are influenced by social and economic conditions in the urban environment.

On Thursday and Friday the conference will offer over 50 breakout sessions presented by local educators. Members of the OSPI staff will have sessions that review and present the latest information and plans from Olympia. There will also be pre-conference workshops on Wednesday that will be packed with useful ideas and information on current topics.

In addition to the above, the WERA annual Spring Conference will be held at the same location on March 28-30, 2007. Theme and keynoters have not yet been identified.

For further information, contact Leonard Winchell, WERA executive secretary, at lenwwa@aol.com, or the WERA Web Site: www.wera-web.org.

Southeastern Association of Community College Research 2006

The 35th annual Southeastern Association of Community College Research (SACCR) conference preparations are well underway. This year's conference theme is, "Tune-In Institutional Quality and Tune-Up Institutional Effectiveness at Two-Year Colleges." The event will be held at the DoubleTree Hotel in Nashville, Tennessee from July 23- 26, 2006. The DoubleTree is about 15 minutes from the Nashville International airport and easy to find.

Among those scheduled to present are Dr. Belle Wheelan (new President of the Commission on Colleges, Southern Association of Colleges and Schools), Dr. Trudy Banta (preeminent scholar, researcher and practitioner in the fields of institutional research and effectiveness), Dr. Joe Marks from the Southern Region Education Board, Dr. Kent Phillippe from the American Association of Community Colleges, and Dr. Linda Hagedorn from the College of Education at the University of Florida.

The pre-conference session this year will feature Dr. Ann Chard, Vice President of the Commission on Colleges of the Southern Association of Colleges and Schools. Dr. Chard's presentation is entitled, "An Overview of the Revised SACS Standards with a Special Emphasis on Institutional Effectiveness and Program/Learning Outcomes." Given all the recent, and newly proposed, changes to the SACS compliance standards, this promises to be an informative and pragmatic workshop.

This year SACCR will also be offering post-conference sessions dedicated to IPEDS Computer Training on the Peer Analysis System. These workshops will run both Wednesday late afternoon and Thursday morning. Nashville State Community College will host the computer-based workshops, which are free to conference attendees.

The 2006 Virginia Assessment Group Fall Conference

The Virginia Assessment Group (VAG) fall conference will be held from November 13-15, 2006 at the Stonewall Jackson Hotel and Conference Center in Staunton, Virginia. The call for proposals for this event is available at: <http://www.virginiaassessment.org>. The deadline for proposal submission is June 30, 2006.

A STATUS REPORT ON THE NCLB STANDARDS AND ASSESSMENTS PEER REVIEWS

Ellen Forte edCount LLC

Anyone who works on large-scale state assessment issues is aware that the *No Child Left Behind Act of 2001* (NCLB) requires states to implement systems of standards and assessments for reading/language arts, mathematics, and science. However, the federal criteria and the process for determining how well states' systems meet them aren't quite so well understood by those outside of the U.S. Department of Education and state education agencies. It is quite likely that a number of states will need substantial assistance in meeting the federal mandates over the next several months and will call upon their contractors and Technical Advisory Committee members to help.

All states are ultimately hoping to get "full approval" of their systems, meaning that, barring any major changes to their standards or tests, their systems meet the current requirements and won't be reviewed again until well into the next reauthorization of the *Elementary and Secondary Education Act* (ESEA). At press time, only three states have achieved this level of approval: South Carolina, Delaware, and Tennessee. All other states have been reviewed and will either be granted full approval or assigned one of the following other categories:

- **Approval Expected:** all components of the state's system appear to meet the federal requirements, but some activities or processes won't be completed until the summer or fall of 2006. For example, standard setting activities may be taking place in June 2006 and once the state submits evidence that cut scores have been approved by its State Board of Education the state is likely to be granted full approval.
- **Approval Pending:** some components of a state's system as administered in the 2005-06 school year did not meet federal requirements, but the state is making changes that will result in compliance with federal requirements prior to the 2006-07 administration. A state that administered an alternate assessment that was poorly aligned with its standards in 2005-06, for example, would need to provide evidence that it has addressed alignment gaps and weaknesses prior to the 2006-07 administration to qualify for full approval.
- **Non-Approved:** several key areas of the state's system are significantly out of compliance with the federal mandates and it would be impossible to address these areas prior to the 2006-07 test administration. Using the alignment example above, a state with no plan or an untenable timeline for addressing major alignment deficits might be "non-approved." States in this category might risk losing part of their Title I, Part A administrative funds.

(Additional details about these approval categories can be found at: <http://www.ed.gov/admins/lead/account/saapr3.pdf>.)

The NCLB peer review process began in April 2003 when the U.S. Department of Education (ED) released peer review guidance that outlines the criteria used to evaluate states' systems in relation to the NCLB requirements. This guidance is organized into seven sections:

- 1) Academic content standards
- 2) Academic achievement standards
- 3) Assessments in the grades and content areas required by NCLB
- 4) Technical quality of the assessments
- 5) Alignment among standards and assessments
- 6) Inclusion of all students in the assessment system
- 7) Reporting of assessment results

Between January 2005 and April 2006, all states submitted evidence to ED to document how well their standards and assessments meet the federal criteria in these seven areas; these evidence packages have all been reviewed and all but the most recently reviewed states have received specific feedback from these reviews.

In preparation for these reviews, ED vetted scores of measurement and testing professionals from universities, state and local education agencies, and independent consulting practices. Employees of testing contractors are not eligible to serve as peer reviewers and reviewers must certify that they do not have any relationship with the state or its testing contractor that would be, or could have the appearance of, a conflict of interest.

The evidence for each state was reviewed by a team of three peers from the peer-reviewer pool. Each team generally included a psychometrician, a special educator or an educator who works with English language learners, and another testing practitioner with extensive practical experience in large-scale testing. Teams were "reshuffled" after each review so that a particular combination of reviewers was never repeated for another state.

Reviewers considered a state's evidence in relation to each criterion in the peer review guidance and submitted a set of detailed notes on this evidence to ED; the peer review team made no recommendations regarding a state's approval status. ED uses the notes to help develop a formal letter that indicates whether the state has met NCLB requirements and outlines any areas where the state must submit additional evidence (see <http://www.ed.gov/admins/lead/account/letters/index.html>). Once this evidence is submitted to ED, the peer review team is reconvened to review it and provide a second set of notes to ED.

While specific issues vary across states, the areas that have posed the greatest challenges to most states have been inclusion of students with disabilities and English language learners, alignment of standards and assessments, and comparability of test scores across forms and testing modes. In many cases, states evidence packages are lacking some basic technical documentation (e.g., technical reports and reports on standard setting and alignment studies; test blueprints; long-term equating plans). Further, many states submitted little evidence to support the meaning of scores from accommodated testing conditions or from the use of alternate assessments for students with disabilities and English language learners.

Testing contractors and TAC members often provide essential support to states in designing and implementing their testing systems. Additional assistance in clearly documenting this work would be of great benefit for states that must submit additional evidence in the next few months. This documentation would also help states to prepare in anticipation of reviews under subsequent ESEA reauthorizations and promote implementation of the professional standards for educational testing among states' high stakes, large-scale assessment systems.

SOFTWARE EXCHANGE WEBSITE

Gary Skaggs and Risper Awuor, Virginia Tech

Two months ago, we posted a survey through the NCME listserve on the possibility of NCME sponsoring a software exchange website and how this website might function. One hundred and fourteen individuals responded to this survey: 61 from higher education, 32 from testing organizations, eight from a government agency, six from K-12 education, and seven "others." Seventy-eight percent of this sample reported that they had searched for software sometime during the past year. They looked for software to accomplish a wide range of analyses, including item and test analysis, IRT and Rasch measurement, DIF, test construction, latent structure analysis, simulation, equating, test scoring, scaling, generalizability, latent class modeling, non-parametric IRT, survey administration, and attitude scaling.

Thirty percent of the respondents said that they had created some type of programming to carry out a measurement-related task during the past year. Of that group, nearly 80 percent indicated a willingness to share their software with others on an NCME website, but several respondents were concerned about the user-friendliness of their software and how confidentiality or licensing issues would be addressed.

The major part of the survey concerned the content and operation of the software website. The table below lists the frequencies of responses to several survey items. We should first say that individuals from different types of organizations (i.e. higher education, testing organization, etc.) responded to these questions in essentially the same way.

Item	Disagree Strongly	Disagree	Agree	Agree Strongly
The author of freeware or open-source software should be an NCME member.	37	51	20	5
To be included on the website, software should meet quality standards determined by NCME.	3	10	39	61
NCME should require a minimal level of documentation but not an evaluation of software quality.	7	24	54	26
Users of software should be able to post comments on the website.	0	3	47	63
If the author is willing, freeware and open-source software should be downloadable from the website.	0	5	32	76
Only information about and links to freeware or open-source software should be provided. No downloadable files should be included.	31	63	11	5

In terms of the types of software that should be listed on the website, over 95 percent of the respondents favored posting freeware and open-source software. Sixty-five percent favored posting commercial software, and 79 percent favored posting the source code for algorithms. However, only 32 percent favored posting proprietary software. Over 85 percent thought that hardware requirements, operating system requirements, and a user manual or guide should be provided on the website. Smaller numbers of respondents favored posting sources for technical support (68%), version histories (54%), and source code (50%).

Finally, on an item asking for additional thoughts, many respondents indicated that they thought having a software exchange website was a “great” idea. Some other comments include:

“Perhaps references to scholarly works that have used the software? It’s not necessary but may prove useful.”

“What if an author provides the software or even algorithms and later gets involved in a commercial endeavor involving that intellectual property? This could get a little sticky.”

“If ‘source for technical support’ is included it should be made clear that this is not an open invitation for in-depth consulting on methodologies employed in the computer program.”

“I very much like the idea of having a ‘bank’ of software especially if it came with user-reviews.”

“Even an encyclopedic clearinghouse would be useful --just to have one portal that links to everything would be a huge step forward--so maybe this is the first step and the second step is to have the quality assured downloads.”

“I’d keep commercial/proprietary software off the list--especially if authors/companies are charging for the software. Shareware s/b acceptable, provided fee requests are voluntary.”

“I have a strong preference for open source programs and software.”

“This website could be of great value to researchers. Often analyses are not done because of the exorbitant cost of certain software or it is not available.”

“Start with few rules. If they are needed, let them grow. Sooner is better than later. An abstract and a contact might be the minimum. Perhaps a bulletin board eventually. RealBASIC is another useful language.”

“This could be a valuable service to members, who may otherwise be limited to a catch-as-catch-can process to locate special-purpose software.”

To sum up, the individuals who responded to this survey supported the idea of having a website devoted to exchanging software, and most who do some programming were willing to share their software. The respondents thought that NCME should determine some minimal levels of quality and documentation for software to be included, but that NCME should itself not evaluate the quality of the software, preferring instead to have users post comments on the website. Most respondents also thought that freeware and open-source software should be housed on the website. One area of disagreement concerned whether the website should be devoted solely to freeware and open-source software or should include information about commercial and proprietary software. Another set of issues concerns confidentiality, licensing, and user-friendliness. The comments above also provide some suggestions on the capabilities for the website. In the next few months the results of this survey will be used to develop a prototype website for you to examine.

NEWSLETTER ADVISORY BOARD

MARY LYN BOURQUE, Mid-Atlantic Psychometric Services
SUSAN BROOKHART, Brookhart Enterprises LLC
SUSAN L. DAVIS, University of Nebraska-Lincoln
ELLEN FORTE, edCount LLC
SARA HENNINGS, Riverside Publishing
JOAN HERMAN, CRESST/UCLA
THEL KOCHER, Edina Public Schools, Minnesota

WENDY MCCOLSKEY, SERVE
GERALD MELICAN, College Board
CAROL S. PARKE, Duquesne University
S.E. PHILLIPS, Consultant
DOUGLAS RINDONE, Consultant
GARY SCHAEFFER, CTB/McGraw-Hill
DONNA SUNDRE, James Madison University

SCOTT BISHOP, Editor, Data Recognition Corporation

Send articles or information for this newsletter to:

Scott Bishop Phone: 763.268.2029
Data Recognition Corporation Fax: 763.268.2529
13490 Bass Lake Road e-mail: SBishop@ DataRecognitionCorp.com
Maple Grove, MN 55311

Sara Hennings Phone: 630.467.7047
Riverside Publishing Fax: 630.467.7126
425 Spring Lake Drive e-mail: Sara_Hennings@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.