NCME national council on measurement in education

Volume 13, Number 2 **NEWSLETTER** **June 2005**

Incoming President Jim Impara, left, congratulates outgoing President Dave Frisbie, at the 2005 NCME Breakfast in Montreal.

FROM THE PRESIDENT

By James C. Impara, Buros Institute for Consultation and Outreach and Caveon Test Security

Greetings. To the 1,100 members who attended the Annual Meeting in Montreal, *bonjour*. To those unable to attend, I hope to see you next year in San Francisco. You missed an excellent meeting. The program was varied and interesting. The Training Committee put together the largest set of professional development sessions that we have had in recent memory (there were 16 different workshops). Thanks to all those who participated and who helped to make the 2005 meeting memorable.

This will be the first of several President's notes in the NCME Newsletter this year. This note covers two topics: (1) an update on the status of the move from AERA to Blackwell for publishing our journals and to The Rees Group for Central Office Services, and the status of the Annual Meeting contract that will stay with AERA; and (2) NCME's mission and focus.

Status of contracts. By now you should have received at least one issue of each journal (*JEM* and *EM:IP*) that has been produced by Blackwell. You should have noticed essentially no changes in the look and feel of *JEM*, but you might have noticed that the inside paper for *EM:IP* was lighter and not slick. This change will make the journal a bit less expensive to print but should have no impact on the production quality. Both editors have expressed their pleasure at how easy the Blackwell staff is to work with. In addition to the hard copy publishing, both journals are now available to members electronically through Blackwell's Synergy service. This should continue to be a positive change for NCME. We recently received a report from Blackwell that our institutional subscriptions now exceed the number of such subscriptions for both journals that we had when AERA was our publisher. This means more income to NCME for the journals, and more importantly, more visibility in the measurement community.

Beginning in June, we will officially have our central office services at The Rees Group (TRG) located in Madison, WI. June is the transition month. Beginning July 1, 2005, TRG will assume full responsibility for NCME's Central Office services. Our new Executive Director, Bruce Wheeler, attended the annual meeting to get a feel for our organization. He had lots of positive things to say about how we operate and how he is looking forward to working with us. Part of TRG's activities will be in member services, which will include maintaining up-to-date membership rosters. A planned activity is the establishment of a member's only access on the NCME web site. This access will permit members to change their address or other pertinent information directly. In addition, we are moving our web site to The Rees Group server and will be setting up a mechanism for submitting annual meeting proposals to the 2006 Program Committee (co-chairs are Chad Buckendahl and Leslie Lukin) electronically. This same site can be used for electronic submissions for Training Session Proposals so they may go directly to that committee (chaired by Lori Nebelsick-Gullet). We will also be working toward permitting payment of dues electronically. Stay tuned for more developments.

At present we are targeting a one-year annual meeting contract with AERA, with a multi-year contract in development. Because this is still in the development stages, more detailed information is not available. We will be working with AERA to smooth out some of the rough spots that have plagued us in recent years (e.g., the lateness of the program, the long lines/waits to register on-site).

The NCME mission and constituency. An important topic at the upcoming Board meeting in July will be a review of NCME's constituency and organizational focus. Our membership includes a wide variety of jobs and interests. The majority of our members are employed in university settings. Other members are employed in K-12 education settings (including local and state education agencies) or with test publishers or similar employers (e.g., contractors doing testing or test-related business, credentialing agencies). This configuration raises several questions about both the direction we should take when recruiting new members, the publication policies in our journals, and the make up of the Board, which now requires a distribution across each of the constituencies by specifying elections from within various constituencies on a rotating basis.

Over 300 people attended the 2005 breakfast meeting in Montreal, and many completed the questionnaires that were on the tables. One of the questions related to recruitment of new members. Many of the responses to that question focused on increasing our outreach to organizations that are involved in credentialing (licensure and certification, e.g., members of the National Organization for Competency Assurance -- NOCA), employment testing (e.g., Society of Industrial and Organizational Psychologists - SIOP), and other arenas that are not presently an organizational focus (e.g., Division 5 of APA). Such a change in the focus and interests of our members would have implications for the content of our annual meetings and our publications. It could lead to a greater emphasis on topics that cut across many different fields including education (e.g., standard setting, DIF, performance testing). Such a change may lead to greater

(continued on page 2)

(continued from page 1)
emphasis on the practical and applied aspects of testing, rather than the theoretical, and the contexts of applications and practice may also change from educational to other settings.

Our current by-laws are specific in that NCME's principal focus is on measurement issues (theory and practice) related to education. The by-laws do not specify what level of education (i.e., K-12, higher education), but many have interpreted the charge as having a principal focus on K-12 educational issues. One could easily argue that many of the issues and problems related to educational measurement are also of interest to the entire measurement community, including higher education, employment testing, and credentialing.

This is an exciting and interesting time in measurement because of NCLB and the proliferation of testing outside the educational community. These changes and the changes in our organizational management make this a doubly exciting time for our community.

LEGAL CORNER: RECONCILING IDEA AND NCLB

By S. E. Phillips, Consultant

For special education students and the schools that serve them, the requirements of two federal education laws and their implementing regulations, the Individuals with Disabilities Education Act (IDEA) and the No Child Left Behind Act (NCLB), are in conflict. This conflict has caused major problems for schools seeking simultaneously to comply with special education law and meet adequate yearly progress (AYP) requirements for the special education subgroup.

Under IDEA, the curriculum for each special education student is determined by the student's Individualized Education Plan (IEP), agreed to by a committee of educators and the student's parent(s)/guardian (IEP team). The IEP team selects the specific math and reading content for which a special education student will receive instruction during the school year, plus other activities and services that will be provided. The specified reading and math objectives are memorialized in the IEP along with recommendations for assessing the special education student's progress in achieving those objectives, including test accommodations, modifications or alternate assessments where appropriate. For example, a 6th grade student with a learning disability (LD) may have an IEP that calls for math instruction on 3rd grade math content and assessment of achievement based on a 3rd grade level math test administered orally with frequent breaks. However, to count toward the NCLB 95% participation requirement for the special education subgroup, the student's school must administer the 6th grade state test designed to measure achievement of the state's 6th grade math standards (unless the student is severely cognitively impaired and qualifies for the 1% exemption).

Administering a test of 6th grade state math standards to a 6th grade LD student who has received instruction on 3rd grade math content is problematic psychometrically and legally. Such testing is inappropriate because the school is being held accountable for performance on content it is legally proscribed from having an opportunity to teach. The student is required to sit for an exam unrelated to the instruction received, raising the possibility of extreme frustration and parental objections. And, because academically challenged students typically learn math primarily in school, one can predict with a reasonably high degree of certainty prior to testing that the student will not be proficient on the 6th grade math test. Alternatively, the school could administer two tests: the NCLB 6th grade math test and the IEP prescribed 3rd grade math test. However, the loss of instructional time and administrative scheduling difficulties associated with double testing make this option unappealing. The school could skip administering the 6th grade test and accept the consequences if the special education subgroup participation rate falls below 95%. However, this option may cause negative repercussions for special education students and teachers who may be blamed for the school's failure to make AYP. Further, if the school is meeting AYP for all other subjects and subgroups, the resulting sanctions may be inappropriately punitive. Moreover, both options leave this 6th grade LD student behind in the AYP calculations because there is no NCLB accountability for the student's progress in achieving the specified 3rd grade math objectives and the school has little incentive to devote extra resources to that student because the school already knows that the student will be counted as not proficient.

There is a reasonable solution to this dilemma that satisfies the spirit of NCLB, allows schools to be in compliance with both IDEA and NCLB, and requires schools to be accountable for the results of the instruction provided to special education students who are not severely cognitively impaired but whose academic achievement is significantly below grade level. Rather than including these special education students in the grade level corresponding to their chronological age, for AYP purposes these students could instead be classified in the grade corresponding to their ***instructional level*** as determined by their IEP teams. That is, for NCLB purposes, the 6th grade LD student described above would be classified as having a 3rd grade *instructional level* because the IEP team determined that the student should be taught 3rd grade math content. Under this system, the school could count this student as proficient if the student scored at or above the proficiency level on the state 3rd grade standards test with appropriate accommodations (e.g., larger font size, more white space on a page, frequent breaks) that did not alter the construct intended to be measured (oral administration of a reading comprehension test or use of a calculator on a math computation test would alter the skills intended to be measured to listening comprehension and calculator literacy, respectively). Using an instructional level testing system, reporting of 3rd grade math results at the school level could include both aggregated data (chronological and instructional grade level students) and disaggregated data.

An important key to success for instructional level testing is clear specification by the IEP team of the state grade level content standards for which the student is to receive instruction. To facilitate appropriate instructional level designations by IEP teams, state content standards must be vertically aligned across grade levels. This differs significantly from the earlier practice of out-of level testing in which special education students were given tests a fixed number of levels below their grade placement without aligning the testing to the instruction the students received and without clear plans for measuring progress and determining when the student should move to the next level. Instructional level testing is different because it is based on testing aligned with the state grade level content standards that the IEP team has determined are most appropriate for the student, and it creates school accountability for student proficiency on those content standards.
(continued on page 3)

Send articles or information for this newsletter to:

Susan M. Brookhart
2502 Gold Rush Avenue
Helena, MT 59601

Phone: (406) 442-8257
Fax: (406) 442-8257
e-mail: susanbrookhart@bresnan.net

The *NCME Newsletter* is published quarterly.

(continued from page 2)
In addition to training IEP teams on instructional level selection, states would also need to extend their reading and mathematics content standards below 3rd grade so appropriate tests of enabling skills could be constructed for special education students with instructional levels below 3rd grade.

Although some educators believe that differences in interests and experiences render a 3rd grade math test inappropriate for a 6th grader, Texas found that a wide variety of topics worked equally well in testing younger and older students on the same content knowledge and skills. Thus, a carefully constructed test of a state's 3rd grade math standards that excludes those few topics appropriate only for younger or older students could serve as an appropriate accountability test for older special education students with an IEP math instructional level at the 3rd grade. The advantages of testing such special education students at their instruction levels include matching testing to instruction, compliance with IDEA and the spirit of NCLB, avoiding extreme student frustration and parental objections, avoiding inappropriate or double testing and holding schools accountable for student proficiency on the instructed content.

Allowing special education students working significantly below grade level to be classified for NCLB purposes at the instructional level determined by their IEP teams would require a change in the NCLB regulations. In addition, implementing rules would need to be devised to discourage low expectations and gaming the system. For example, *significantly below grade level* would need to be defined. One possibility would be documented individually administered achievement test results two or more grade levels below chronological grade placement. Such documentation should be available in the IEP as part of the required student evaluation or re-evaluation for special education services. Schools might also be expected to design an accelerated instructional program for mildly cognitively disabled students who are no more than one year behind and will be tested on-grade-level.

Additionally, rules that discourage inappropriate multiple repetitions of the same grade level content may be useful. For example, there could be a presumption that a special education student who scores proficient or higher on the current instructional grade level test will move on to the next grade level content. If the student scores below proficient, the student could remain at the same instructional level for at most one more year unless the IEP documents a deteriorating medical condition or serious disability for which additional time is indicated.

Another area for which implementing rules might be crafted includes a cap on the percent of students who may be classified by instructional level. Some policymakers are concerned that low expectations have caused some special education students to achieve much farther below grade level than they would have if appropriate instruction had been provided. However, the percent of such special education students is unknown and low expectations that may have persisted over an extended period of time cannot be corrected in a year or two in students with learning difficulties. Thus, there will be a period of transition in which it may be advisable to allow a gradually decreasing percentage of students to be classified by instructional level. For example, 3% might be allowed the first year, 2½% the second year, and so on, until the desired percentage is reached.

Note that in all the examples given above, the implementing rules would specify conditions under which students tested at their instructional levels would count in determining AYP participation and proficiency. However, they would not overrule the IEP team decisions. Special education students would still receive instruction and testing as dictated by their IEPs. Allowing limited instructional level testing for NCLB purposes would be fairer to schools, benefit students whose schools are accountable for the their instructional level learning, and provide incentives for schools to teach special education students at the highest levels possible while still honoring IDEA requirements.

The National Association of Test Directors has an online newsletter. You can access NATD Newsletters on the web by going to http://natd.org/.

President Frisbie (center) congratulates 2005 Annual Meeting Program Chairs Xiaohong Gao (left) and Julie Noble (right) at the NCME Breakfast.

VAN DER LINDEN RECEIVES NCME AWARD FOR CAREER CONTRIBUTIONS TO EDUCATIONAL MEASUREMENT

By Richard M. Luecht, University of North Carolina at Greensboro

The 2005 NCME Award for Career Contributions to Educational Measurement was presented to Wim J. van der Linden in recognition of his many eloquent technical contributions to educational measurement, his mentorship of students and colleagues, his scholarship, and his service to the field. His pioneering work on optimal [automated] test assembly has led to a number of scientific breakthroughs that have revolutionized the testing industry. He has also made significant contributions in the areas of test equating; applying of statistical decision theory to selection, placement, and classification decisions; standard setting; and modeling response-time data. Dr. van der Linden has authored countless articles and book chapters and co-edited some of the leading measurement handbooks covering item response theory and computerized adaptive testing. He has also served on many technical advisory committees throughout the world, been the President of the Psychometric Society, held numerous editorial appointments, and served NCME in many important ways throughout his career. He deserves this honor for a career that has had widespread positive and creative impact on the field of educational measurement.

Career Award winner Wim van der Linden (right) receives congratulations from President Dave Frisbie (left).

MY CAREER IN EDUCATIONAL MEASUREMENT

By Wim J. van der Linden, University of Twente

As a student, I never thought of a career in educational measurement. I was more interested in research methods and statistics. In fact, the only class on test theory I ever took as part of my graduate training was my very last class. But two events were decisive in my early career. The first was a visit by Melvin R. Novick to the Netherlands in 1974 to teach a one-week course on Bayesian statistics, in which he covered the contents of a new book he wrote with Paul Jackson, *Statistical Methods for Educational and Psychological Research*. This was an unusual book, not only because it was strictly Bayesian but also because most of its examples and applications were on test theory. The second event was an ETS summer course on item response theory by Frederic M. Lord in 1977, in which he taught the first version of his manuscript for his 1980 book, *Applications of Item Response Theory to Practical Testing Problems*. These two giants in our field have deeply influenced my career, for which I am most grateful. Not many people have had the privilege of learning psychometric theory directly from Lord and Novick! From Novick are my Bayesian inclinations; from Lord I have inherited a deep interest in practical testing problems.

In order to explain what is so motivating about practical testing problems, I take the liberty to quote from the preface of my book *Linear Models for Optimal Test Design* (Springer, 2005), which will be out in a few weeks. There, I observe that the first hundred years of test theory have been a success story. But then I continue by observing that "…in spite of its enormous progress, though test theory is omnipresent, its results are used in a peculiar way. Any outsider entering the testing industry would expect to find a spin-off in the form of a well-developed technology that enables us to engineer tests rigorously to our specifications. Instead, test theory is mainly used for post hoc quality control, to weed out unsuccessful items, sometimes after they have been pretested, but not seldom after they have already been in use. Apparently, our primary mode of operation is not to create good tests, only to prevent bad tests. To draw a parallel with the natural sciences, it seems as if testing has led to the development of a new science but the spin-off in the form of a technology for engineering the test has not yet been realized."

The development of a technology of test design has been my main research topic for the last 20 years. The idea of designing a test by modeling its features as a set of constraints that have to be imposed on its composition has been fruitful and shown to work for an incredibly wide range of features. Another topic near to my heart is adaptive testing. The view of an adaptive test as a set of shadow tests (van der Linden & Reese, 1998) is fundamental; it has allowed me to apply all we know about the design of fixed tests to adaptive testing.

A more recent topic is response times. During the whole history of testing, the aspect of time has been ignored simply because we were not able to record response times on test items in a routine fashion. Since the arrival of the computer in testing, we are no longer innocent. All our current testing procedures−including procedures for such diverse topics as item analysis, item selection, test validation, standard setting, and test-score equating−have to be revised to deal better with the differences between the speed at which persons operate and the time intensities of the items they take.

I also have a more fundamental interest in test-score equating. A few years ago, the question of why we use a single transformation to equate the observed scores of test takers with different ability levels struck me. The use of different standard errors for test takers with different scores is now well established: We all know that the standard error we report has to reflect the differences between the observed-score distributions of different persons. But why do we still ignore these differences if we *equate* their scores? I am convinced we should rethink our current methods of test-score equating.

HARTZ AND ROUSSOS RECEIVE NCME ANNUAL AWARD

By James S. Roberts, University of Maryland

Drs. Sarah Hartz and Louis Roussos received the NCME 2005 Annual Award for an Outstanding Example of an Application of Educational Measurement Technology to a Specific Problem. Selection committee chair Jim Roberts presented the award to Hartz and Roussos at the annual meeting "for their development of the Fusion Model based Arpeggio Software System for Skills Diagnosis."

Dr. Hartz is currently a medical student at the University of Iowa, but formerly worked on this project with Dr. Roussos at the University of Illinois at Urbana-Champaign. Their project implemented a hierarchical Bayesian model and associated software for skills diagnosis that will soon be operational on several standardized tests administered by Educational Testing Service. The innovation in their work and the impact that it will soon have on skills assessment make this award well deserved.

PENFIELD RECEIVES 2005 JASON MILLMAN PROMISING MEASUREMENT SCHOLAR AWARD

By Susan M. Brookhart, Duquesne University and Brookhart Enterprises LLC

The Jason Millman Promising Measurement Scholar Award honors the lifetime work of Dr. Millman in the field of applied measurement and continues his own support of scholars just beginning their research careers. The Award recognizes an early career scholar whose research has the potential to make a major contribution to the applied measurement field. Committee Chair Jeffrey K. Smith presented the 2005 Jason Millman Award to Randall D. Penfield, University of Miami, at the 2005 NCME Annual Meeting Breakfast.

SUMMARY OF MY RESEARCH

By Randall D. Penfield, University of Miami

It is a tremendous honor to be the 2005 recipient of the Jason Millman Award, and to have the opportunity to be placed, if only for one day, among the ranks of previous recipients of the award – individuals whose work I admire greatly. My work in educational measurement began with dissertation research related to adjusting differential item functioning (DIF) effect size estimates for the impact of matching variable contamination. While this work was never published, it paved the road for subsequent published research in the area of DIF proposing: (a) the application of the cumulative common odds ratio as a DIF effect estimator for polytomous items, (b) nonparametric methods for the assessment of DIF across multiple groups, (c) a nonparametric estimator of nonuniform DIF effect, and (d) estimators of DIF effect variance across the items of mixed format tests. In addition, I have devoted a great deal of time to creating a windows-based computer program (DIFAS) that can compute a variety of nonparametric DIF statistics for dichotomous and polytomous items.

In between DIF projects, I found time to pursue research advancing methodology related to the assessment of content validity and parameter estimation of item response theory (IRT) models. With respect to content validity, I developed an asymmetric confidence interval for the mean of rating scale items that is not based on the assumption of normality associated with traditional confidence intervals for a population mean. This confidence interval has been shown to provide excellent coverage rates under a variety of conditions, and as such has been applied to the analysis of content validity ratings.

(continued on page 5)

(continued from page 4)
My research in the area of IRT has been based primarily on the estimation of examinee ability and the standard error of the ability estimator. This research has addressed several issues, including: (a) the effect of lack of fit on parameter estimation for the partial credit model, (b) the use of Bayesian methods in item selection of adaptive tests using polytomous item formats, (c) the use of a posterior-weighted approach to estimate the standard error of the ability estimator in adaptive tests, and (d) the uncovering of properties of information functions unique to items following dichotomous and polytomous Rasch models. My interest in teaching IRT led me to create a windows-based computer program (IRT-Lab) that can graphically display many of the concepts encountered in IRT, such as likelihood functions, information functions, and item characteristic curves. This program has now been used in approximately 20 countries world wide by hundreds of students, teachers, and researchers of IRT.

SKAGGS, KIM RECEIVE INAUGURAL HANSON AWARDS

By Richard J. Patz, R. J. Patz, Inc.

Bradley Hanson served the educational measurement community as a scholar, mentor, and open source software developer. He shared his extraordinary talents generously. For the inaugural presentation of the award established in Brad's memory, we were honored to have with us at the NCME breakfast his father Bruce and his brothers Scott and Todd. The Bradley Hanson Award is intended to recognize and support members of NCME as they advance projects of value to our community. In this inaugural year the committee was pleased to make two awards. Thanks to generous contributions to the award fund, we were able to make these two awards in the full $1000 amount.

In recognition of his contributions and in support of his project to create a psychometric software exchange website for the NCME community, The Committee was pleased to present a Bradley Hanson Award to Gary Skaggs, of Virginia Tech. In recognition of his accomplishments and in support of his project titled "A comprehensive review of IRT Command Language, with recommendations for future developments" The Committee was pleased to present a Bradley Hanson Award to Seonghoon Kim, of ACT. Dr. Kim was unable to attend the breakfast, and the award was accepted on his behalf by his doctoral dissertation advisor, Michael Kolen.

SUMMARY OF MY PROJECT PLAN FOR THE BRADLEY HANSON AWARD

By Gary Skaggs, Virginia Polytechnic Institute and State University

I am delighted and honored to receive the first Bradley Hanson Award. At the NCME breakfast, I had the pleasure of meeting Brad's father and brothers. They were thrilled to see NCME honor their lost loved one, and I felt privileged to be a part of that occasion. For this award, I proposed to begin to develop a website for exchanging psychometric software. Brad felt strongly about both having a software exchange website and making software available and free, and I am happy to have the opportunity to continue that vision. My own interest in software issues came out of research I have been working on recently as I noticed how dependent our field is on the availability of software or on the programming skill of the researcher in creating software.

This project is an outgrowth of my working with Brad on psychometric software issues. At the 2001 Annual Meeting, Brad presented the report from an NCME ad hoc committee he chaired on issues related to psychometric software. He asked me to be a discussant at that session. After that session, I surveyed several measurement journals on software use and published the findings in an article in *Educational Measurement: Issues and Practice* (Spring, 2004). One of the recommendations in that article was the creation of a psychometric software exchange website. This had also been recommended in the report of Brad's committee, and Brad had developed a prototype exchange on his website. Unfortunately, since his untimely death, this exchange is no longer available. I still firmly believe that such an exchange would be valuable to the measurement community and facilitate our research.

Specifically, I plan to use the funds from the award to: (1) obtain input from NCME members on the structure and function of the exchange; (2) meet with an NCME representative group (e.g. Board of Directors or appropriate committee) to finalize the design of the exchange, including submission procedures and criteria; and (3) begin to implement it. To obtain input from NCME members, I propose to create a brief survey and administer it via the NCME listserv. Based on the results of the survey, I will propose an initial design for the exchange and present it to the appropriate NCME group, probably at the 2006 Annual Meeting.

I have some initial thoughts on the website. First, although the site would provide links to commercially available software, a major focus of the site would be to house freeware, open-source applications, and subroutines and functions written by individual researchers. Second, software can be accessed and then downloaded through keywords (e.g. simulation, equating, etc.). Third, there should be some mechanism for software users to provide comments or recommendations to the research community. My hope is that this website will make our research efforts more efficient and productive.

Skaggs, G. (2004). Software use in psychometric research. *Educational Measurement: Issues and Practice, 23*(1), 28-33.

SUMMARY OF MY PROJECT – A COMPREHENSIVE EVALUATION OF THE COMPUTER PROGRAM ICL: DIAGNOSES, COMPARATIVE STUDIES, AND SUGGESTIONS FOR IMPROVEMENT

By Seonghoon Kim, ACT, Inc.

I am greatly honored to have received this award. I wish to thank Dr. Robert Brennan and Dr. Timothy Ansley at the University of Iowa for their sincere and eager nominations for this award. I would also like to acknowledge gratefully the constant guidance and encouragement I have received from Dr. Michael Kolen, my mentor, in conducting the project.

The goal of measurement theory is concerned with the justification of various measurement procedures and with the meaningfulness of their results. In item response theory (IRT), the goal is closely related to the statistical techniques for estimating item and proficiency parameters (i.e., calibrating) and developing calibration software to implement the techniques. The techniques are usually quite complicated and hardly applicable to practical problems without the help of calibration software.

It has been a matter of regret in the field of educational measurement that technical support and improvement of the open-source computer program *IRT Command Language* (ICL) stopped because of the sudden death of Dr. Bradley A. Hanson, the developer of the program. I believe that ICL is one of the best IRT calibration programs based on reliable frameworks in terms of psychometric, statistical and computing techniques. ICL is the only widely available IRT calibration computer program that is open source, meaning that all of the source code is available to the public. However, the sudden loss of the developer left it to the educational measurement community to make it more user-friendly and efficient, let alone to maintain it. Regrettably, practical applications of ICL are rarely
(continued on page 6)

(continued from page 5)
found in the literature, and there are gaps in the documentation of the theoretical details for ICL. This situation motivated the present project.

The main purpose of the project is to revive the usage of the computer program ICL by comparing its performance to the performance of other commercial IRT calibration software and revealing and documenting the rationale for algorithms used in ICL. To the best of my knowledge, areas that need detailed description of the algorithms include multiple-group estimation, bootstrapping item parameter estimates, and pretest item calibration. Of course, functions of ICL for implementing these areas need to be fully tested with a variety of real and simulated data. The following briefly summarizes the project:

1. Dissect, screen, and arrange source code of ICL, so that users can easily use the code for their own purposes.
2. Conduct a series of simulation studies to test the performance and behavior of ICL in comparison to other commercial calibration computer programs. Situations for simulation include those that need separate calibration, concurrent calibration, fixed-parameter calibration, and scale transformation for linking.
3. Document some algorithms built into ICL in technical papers that describe them in detail.
4. Based on the first three points above, make suggestions for further improvement of the performance of ICL.
5. Make suggestions about how to create a graphical user interface (GUI) version of ICL, so that users can use it more easily and more effectively.

SINHARAY RECEIVES CASCALLAR AWARD FOR AN OUTSTANDING PAPER BY AN EARLY CAREER SCHOLAR

By Linda Cook, Educational Testing Service

The Alicia Cascallar Award for an Outstanding Paper by an Early Career Scholar was awarded for the first time at the 2005 Annual Meeting of NCME. Alicia Cascallar is best remembered for her prolific work in the area of differential item functioning. The Alicia Cascallar award has been established to honor Alicia's professional commitment and accomplishments and to continue her practice of mentoring and encouraging 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. This year's award recipient is Sandip Sinharay, a Research Scientist at Educational Testing Service. Dr. Sinharay received the award for his innovative and important work in the area of Bayesian item fit analysis for item response theory models.

SUMMARY OF MY RESEARCH

By Sandip Sinharay, Educational Testing Service

It is deeply gratifying to receive an award honoring Alicia Cascallar, who was an excellent researcher as well as a wonderful person. A significant portion of my research focuses on assessing model fit, which is an area in psychometrics in need of more investigation. This paper applies a popular Bayesian model checking technique to examine item fit—an important aspect of model fit—for simple item response theory (IRT) models.

Item fit plots, which compare the observed and predicted proportion correct scores of examinees with different raw scores, are found to provide graphical evidence of misfit, if it is present, of the IRT model employed. These plots are among the handful of statistically rigorous model fit plots that can be found in current educational measurement literature. This paper also suggests a way to obtain (Bayesian) p-values for the item fit statistics of Orlando and Thissen (2000); the p-values numerically summarize the information in the above-mentioned item fit plots. As the paper describes in detail, the suggested techniques overcome the limitations of competing item fit techniques (Orlando & Thissen, 2000; Glas & Suarez-Falcon, 2003).

A number of detailed simulation studies and an actual data application demonstrate the effectiveness of the suggested item fit diagnostics. The suggested techniques seem to have adequate power and a reasonable Type I error rate under a variety of situations. Moreover, results are often similar to what is found in existing literature. As test consumers become more demanding, practitioners of educational measurement will need to employ more complicated statistical models for scoring examinees. The techniques suggested in this paper can be generalized in a straightforward manner to assess the fit of those models, as exemplified in another research article of mine that will appear in the *Journal of Educational and Behavioral Statistics.*

MICHAELIDES, KIM RECEIVE 2005 BRENDA H. LOYD OUTSTANDING DISSERTATION AWARD

By Lisa F. Smith, Kean University

The 2005 Brenda H. Loyd Award for outstanding dissertation work in the field of educational measurement was shared by two recipients. Dr. Michalis P. Michaelides of Stanford University's School of Education was recognized for his dissertation entitled, "Effects of Common-Item Selection on the Accuracy of Item Response Theory Test Equating With Nonequivalent Groups." Dr. Michaelides' dissertation advisor was Dr. Edward Haertel. Dr. Seonghoon Kim of the University of Iowa was recognized for his dissertation entitled, "Unidimensional IRT Scale Linking Procedures for Mixed-Format Tests and Their Robustness to Multidimensionality." Dr. Kim's advisor was Dr. Michael Kolen. The criteria used by the Dissertation 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 writing. The awards were presented at the NCME Breakfast by Lisa F. Smith, chair of the committee.

SUMMARY OF MY DISSERTATION RESEARCH

By Michalis P. Michaelides, The College Board

Entitled "Effects of common-item selection on the accuracy of item response theory test equating with nonequivalent groups," my dissertation comprised three studies on statistical problems in item response theory (IRT) test equating related to the selection of items to be used as common across alternate test forms.

The first study showed that the treatment of a few common items that behaved in unexpected ways across administrations (i.e. the outliers flagged by the delta-plot method) could have substantial influence on equated score summary statistics. Item responses from four statewide assessments were analyzed using four IRT equating methods – Stocking & Lord (1983), Haebara (1980), mean/sigma (Marco, 1977), and mean/mean (Loyd & Hoover, 1980). Dichotomous items were calibrated with both a three- and a one-parameter logistic model; Samejima's (1969) graded response model was used to calibrate the polytomous items. In two of the four assessments, mean scores, annual gains, and proportions above a cut score differed considerably depending on whether as few as two or three outlying items were included in or excluded from the equating pool. Factors such as the number of outlying items, their type (dichotomously or polytomously scored), their level of difficulty, the direction and the amount of their change across administrations, and the IRT model and equating transformation fitted to the data were discussed with regards to their influence on equated summary statistics.

(continued on page 7)

(continued from page 6)
In the second study, the Mantel-Haenszel statistic (Mantel & Haenszel, 1959), widely used in studies for identifying differential item functioning (Holland & Thayer, 1988), was proposed as an alternative to the delta-plot method. A procedure was developed and applied in the context of equating a statewide assessment for flagging common items that behave differentially across examinee cohorts. The proposed procedure has the advantage of conditioning on ability when comparing the performance of two administration cohorts on an item. There are schemes for interpreting the effect size of differential performance, for both dichotomous and polytomous items, which can inform the decision to retain those items in the common-item pool or to discard them. Some test design limitations that preclude the application of this procedure in a test equating framework were also presented.

Since common-item parameters are obtained from responses of samples of examinees, there is variability in the estimation of an IRT equating transformation, which is quantified by the standard error of equating. Common items are chosen according to certain content and statistical specifications, and are currently treated as fixed for purposes of deriving an equating function. However, other items that conform to the same specifications could have been selected to equate test forms, thus there is additional error variance due to the sampling of common items. In the third study, the selection of common items was considered random, and the amount of error due to the sampling of common items was quantified by an analytic formula and by a computational bootstrap procedure using data from a statewide assessment program. The two approaches produced nearly identical results when distributional assumptions for the analytic formulation held. Compared to other sources of sampling and measurement error, the relative size of the common-item sampling error was small with respect to individual scores, but loomed larger for group-level score interpretations.

References:

Haebara, T. (1980). Equating logistic ability scales by a weighted least squares method. *Japanese Psychological Research, 22*, 144-149.

Holland, P. W., & Thayer, D. T. (1988). Differential Item Performance and the Mantel-Haenszel Procedure. In H. Wainer, and H. I. Brown (Eds.), *Test Validity* (pp. 129-145). Hillsdale, NJ: Lawrence Erlbaum Associates.

Loyd, B. H., & Hoover, H. D. (1980). Vertical equating using the Rasch model. *Journal of Educational Measurement, 17*(3), 179-193.

Marco, G. L. (1977). Item characteristic curve solutions to three intractable testing problems. *Journal of Educational Measurement, 14*(2), 139-160.

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute, 22,* 719-748.

Samejima, F. (1969). *Estimation of latent ability using a response pattern of graded scores* (Psychometric Monograph No. 17). Iowa City, IA: Psychometric Society.

Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory. *Journal of Educational Measurement, 7*, 201-210.

CDs of 2005 NCME Annual Meeting Sessions.
Recordings of selected sessions at the 2005 Annual Meeting are available to all NCME members. Members may order copies by going to the tape company website http://www.softconference.com/250411 and selecting NCME Individual Audio CDs. CDs are $12 or $24 US, depending on length.

NCME editors at the 2005 Annual Meeting. Left to right: Michael Kolen, Journal of Educational Measurement; Steve Ferrara, Educational Measurement: Issues and Practice; Sue Brookhart, NCME Newsletter; David Miller, Webmaster; Deborah Harris, ITEMS.

SUMMARY OF MY DISSERTATION RESEARCH – UNIDIMENSIONAL IRT SCALE LINKING PROCEDURES FOR MIXED-FORMAT TESTS AND THEIR ROBUSTNESS TO MULTIDIMENSIONALITY

By Seonghoon Kim, ACT, Inc.

The use of item response theory (IRT) in testing applications has grown considerably over the last few decades. Meanwhile, in both classroom and large-scale assessments, there has been a steady increase in the use of mixed-format tests, which contain a mixture of different item formats, such as multiple-choice (MC) items and constructed-response (CR) items. When constructs measured by different formats of items are claimed to be almost identical, distinct unidimensional IRT models can be used to analyze the different formats on a mixed-format test.

The primary purpose of my dissertation was to investigate procedures for placing IRT parameter estimates from separate administrations of mixed-format test forms on a common scale. Four scale linking procedures, originally developed under dichotomous IRT models, were extended to mixed-format tests containing items of different formats. The four procedures included two moment procedures (mean/mean and mean/sigma) and two characteristic curve procedures (Haebara and Stocking-Lord). Two simulation studies were conducted to evaluate the relative linking accuracy of the four "unidimensional" linking procedures and to examine their robustness to test multidimensionality, in comparison to concurrent calibration.

The first simulation study generated data under the condition of unidimensionality holding across MC items and CR items. Based on the quantification of format effect as the correlation between two dominant constructs measured by MC items and CR items, the second study simulated multidimensional data reflecting the format effect. Three factors of mixed-format test type, nonequivalence in linking, and calibration type were considered for the first study, and additionally a format effect factor was included for the second study. In both studies, the common-item nonequivalent groups design was used and the computer program MULTILOG was used for test calibration.

Among linking procedures, the characteristic curve procedures outperformed the moment procedures, regardless of the presence of format effects. There was a negligible difference in performance between the two characteristic curve procedures. In general, the concurrent calibration procedure outperformed the four linking procedures in linking accuracy and robustness to format effects. However, the performance of the concurrent calibration procedure was only slightly better than that of the characteristic curve procedures via simultaneous calibration across formats. Although the concurrent calibration procedure and the two characteristic curve procedures showed some evidence of being robust to severe format effects (correlation of 0.5), the evidence did not seem to be consistent across test types. This study might have favored the concurrent calibration procedure, because the spread for the new form scale was not estimated but given with its parameters in concurrent calibration using MULTILOG.

NOMINATIONS FOR EDITOR SOUGHT

By Terry Ackerman, University of North Carolina-Greensboro

Nominations are sought for two NCME editorships this year: the *Newsletter* and *ITEMS*. Nominations will be screened by the NCME Publications Committee. The committee offers a slate to the president who, in turn, makes a recommendation for appointment to the NCME Board of Directors. This appointment process is designed to provide for a smooth transition between incoming and outgoing editors. If you are interested in either position, if you would like to nominate a colleague, or if you would like additional information, please contact Terry Ackerman, Publications Committee Chair (taackerm@uncg.edu or phone 336-334-3474).

NCME NEWSLETTER EDITOR

The Publications Committee is soliciting nominations for the editor of the NCME *Newsletter*. The *Newsletter* is published electronically four times a year. It includes announcements and brief descriptions of current activities, both internal and external to NCME, that would be of interest to the membership. It is intended to provide the broad diversity of NCME members with timely information about important current events in educational measurement practice and research. The *Newsletter* also is intended to inform members about the activities of the Board of Directors and committees.

The *Newsletter* editor must have an extensive understanding of the measurement field and of NCME as an organization. The editor must be skilled at composition and editing and have a deep appreciation for timeliness. The term of the *Newsletter* editor will be from January 2006 to December 2008.

ITEMS EDITOR

The Publications Committee is soliciting nominations for the editor of the NCME publication *ITEMS*: The Instructional Topics in Educational Measurement Series. The goal of *ITEMS* is to improve the understanding of educational measurement principles by providing brief instructional units on timely topics in the field, modules developed for use by college faculty and students as well as by workshop leaders and participants..

Instructional modules appear in the NCME publication *Educational Issues: Measurement and Practice (EM:IP)* and are designed to be learner-oriented and consist of an abstract, tutorial content, exercises, and annotated references. The teaching aids accompanying most modules are designed to support the use of the instructional modules in teaching and workshop settings by providing supplemental student exercises, references, test items, and figures or masters for transparencies.

The *ITEMS* Editor must be willing to solicit manuscript authors and have the organizational skills to track multiple modules through their various stages of completion. The editor's term will run from January 2006 to December 2008.

Photo credits. The photographs in this Newsletter were taken by Doug Becker, ACT, Inc., at the NCME Annual Meeting in Montreal, April 12-14, 2005.

PHOTOS FROM THE ANNUAL MEETING

The Directors' Table at the NCME Breakfast. Left to right: Linda Cook, Jeri Benson, Wayne Camara, Terry Ackerman, Suzanne Lane, Duncan MacQuarrie, Jim Impara, Dave Frisbie.

Vice President Daniel Eignor at the NCME Breakfast.

Thomas Haladyna, David Frisbie, and Steven Downing at the NCME Reception.

CALLS FOR NOMINATIONS FOR THE 2006 NCME AWARDS

Abbreviated versions of the Calls for Nominations for the 2006 NCME Awards appear below. Those who wish to nominate candidates for any of these awards should check the NCME website (www.ncme.org) for more details. For examples of award-winning work by this year's awardees, see the stories on pages 3 through 7 of this Newsletter.

CALL FOR THE NCME AWARD FOR CAREER CONTRIBUTIONS TO EDUCATIONAL MEASUREMENT

NCME members are encouraged to nominate individuals for the NCME Career Contributions 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 influential contributions may include theoretical or technical developments, service to professional organizations, conceptualizations of educational measurement that have enhanced public understanding of measurement programs, applications of theory that have influenced the nature of educational tests and measurements, or innovative ideas that have significantly affected measurement practices. Award recipients receive a check for $1,000 and a citation at the NCME Annual Breakfast. In addition, recipients are invited to provide an invited address at the next year's NCME Annual Meeting. To be considered by the Award Committee, 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. Please submit seven (7) copies of all materials by December 15, 2005 to : Dr. Robert Mislevy, University of Maryland, EDMS, Benjamin 1230-C, College Park, MD 20742 [rmislevy@umd.edu; 301-405-1933, voice; 301-314-9245, fax]

CALL FOR THE NCME AWARD FOR AWARD FOR TECHNICAL OR SCIENTIFIC CONTRIBUTIONS TO THE FIELD OF EDUCATIONAL MEASUREMENT

In the year 2006, NCME will honor technical or scientific contributions to the field of educational measurement in 2003, 2004 or 2005. Selection criteria for the award will include quality, innovation, and importance of the contribution. Self nominations and nominations of others are encouraged. Individuals or groups are eligible for this award. Nominees need not be NCME members. A nomination consists of 6 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. Please submit six (6) copies of all materials by December 12, 2005 to: Dr. Barbara G. Dodd, University of Texas at Austin, Department of Educational Psychology, 1 University Station D5800, Austin, TX 78712-0383 [bg.dodd@mail.utexas.edu; 512-471-0188, voice; 512-471-1288, fax]

CALL FOR THE JASON MILLMAN PROMISING MEASUREMENT SCHOLAR AWARD

The Jason Millman Promising Scholar Award honors the lifetime work of Dr. Millman and continues his support of scholars who are just beginning their research careers. In addition to recognition by NCME, the successful candidate will receive $1000. Only one candidate will be chosen to receive the award each year. To be eligible for the award, the candidate must have: received the doctorate within the last five years; two (2) or more unique papers either accepted for presentation at an NCME annual meeting or published in NCME publications within the last five years; and the support of his/her professional colleagues that his/her work represents a significant contribution to the field of applied measurement. Applications or nominations must include the following items: (1) letter of nomination from a professional colleague who is a member in good standing of NCME; (2) at least two letters of recommendation (from persons other than the nominator) that speak to 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 the reasons for which the candidate's work represents a significant contribution to the field of applied measurement; (3) 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 (the candidate must be the first author on all multiple-author papers and provide a statement that defines his/her contributions to the paper); (4) candidate's current curriculum vita ; and (5) a letter from the candidate outlining his/her career goals and how his/her work contributes significantly to the field of measurement. Deadline for submission is November 4, 2005. One (1) copy is required for materials submitted electronically. Six (6) copies are required for materials submitted as hard copy, submitted on the same date. If more than one mode of delivery is used for the submission, the candidate must notify the Committee chair of the modes and expected date(s) of arrival. Submit materials to Dr. Karen Mitchell, SRI International, 100 Wilson Blvd, Suite 2800, Arlington, VA 22209-2268 [karen.mitchell@sri.com; 703-247-8576]

CALL FOR THE BRADLEY HANSON AWARD FOR CONTRIBUTIONS TO EDUCATIONAL MEASUREMENT

The Bradley Hanson Award program honors Brad's contributions to the field of educational measurement and the goals embodied in his work as a scholar, practitioner, mentor, and developer of open source scientific software. The program annually awards a total of $1000 to a nominee or nominees in support of projects that promise to directly make a significant contribution to the field of educational measurement, and/or that promise to make a significant contribution to the development of new professionals in the field. To be eligible for the award a candidate must (1) be a member of NCME or be a graduate student under the supervision of an NCME member; and (2) be working on a project that promises to make a significant contribution to the field of educational measurement and/or a significant contribution to the development of new professionals in the field. Nominations, which describe both a project and a candidate recipient, will be evaluated by three criteria: (1) the importance of the project to be supported, (2) the importance of the financial award to the success of the project; and (3) the qualifications of the candidate and his or her ability to ensure the success of the project. Applications/nominations for the award must include the following: (1) letter of nomination from a member in good standing of NCME (self nominations/applications are welcome), describing both the candidate and the project to be supported, addressing specifically the selection criteria detailed above; (2) at least one additional letter of recommendation (from person(s) other than the nominator) addressing the qualifications of the candidate and the importance of the project; and (3) candidate's curriculum vita. Please submit seven (7) copies of all materials by November 1, 2005 to: Dr. Deborah Harris, ACT, Inc., 500 ACT Drive, P.O. Box 168, Iowa City, IA 52243, Attention: Bradley Hanson Award [Deborah.harris@act.org]

CALL FOR THE ALICIA CASCALLAR AWARD FOR AN OUTSTANDING PAPER BY AN EARLY CAREER SCHOLAR

Alicia Cascallar, who published most of her work as Alicia P. Schmitt, is best remembered for her prolific work in the area of differential item functioning (DIF). 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 and encouraging promising new scholars in the area of educational measurement. An award of $500, a citation, and a waiver of NCME conference fees will be provided as partial support for an early career member of NCME to travel to the annual meeting. The award will be given for the most outstanding paper by an early career scholar that is accepted for presentation at the Annual Meeting. To be eligible for this award the individual must have his or her proposal accepted for the 2006 NCME Annual Meeting and it must include a research paper that is either presented in a paper session or as part of a symposium or panel discussion. The author(s) must be an early career member of NCME (received their doctoral degree within 5 years of the annual meeting). Papers will be evaluated for their scientific merit, clarity and completeness, the extent to which the material is redundant with previous publications and presentations, and the relevance of the work to practitioners in the field. After notification of acceptance to the Annual Meeting, authors will submit 6 copies of their completed research papers, and a 3-5 page executive summary of the research and its relevance to the Award Chair for review. For multi-authored papers, the first author must meet the above eligibility criteria, and will receive the cash award and citation. All other co-authors must provide a statement indicating that the first author was responsible for at least 75% of the effort. Nominees for the award can be either self-nominated or nominated by some other person. Candidates must submit an electronic copy of their paper. Nominators and other endorsers of the paper will submit their letters either by mail or e-mail to the Committee Chair. Submit materials by November 1, 2005 to: Dr. Neil J. Dorans (Attention: Alicia Cascallar Award), Center for Statistical Theory and Practice--MS 12T, Educational Testing Service, Rosedale Road, Princeton, NJ 08541 [ndorans@ets.org]

CALL FOR THE BRENDA H. LOYD OUTSTANDING DISSERTATION AWARD

The National Council on Measurement in Education (NCME) is seeking nominations for the ninth annual Brenda H. Loyd Award for an outstanding dissertation in the field of educational measurement. Nominations will be accepted for dissertations completed between July 1, 2003, and June 30, 2005. 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 $1,000 and a commemorative plaque. In addition, the advisor or committee chair for the award-winning dissertation will receive a letter of congratulations. An honorable mention award may also be given; its recipient will be recognized with a certificate. To nominate a dissertation, the following items should be submitted to the Chair of the Brenda H. Loyd Dissertation Award Committee: (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 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 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. Please submit seven (7) copies of all materials by November 14, 2005, to: Dr. Cheryl D. Cardell, Academic Affairs, The University of Texas at Arlington, 701 S. Nedderman Dr., Suite 216, Arlington, TX 76019-0156 [cardell@uta.edu; 817-272-2737]

The NCME Recruitment of Educational Measurement Professionals Committee has updated the Graduate Program Descriptions posted on the NCME web site (www.ncme.org) under the OPPORTUNITIES tab. The listing contains information on the graduate programs in educational measurement and related areas for various institutions listed alphabetically by state. These listings can be used to provide prospective students with listings of graduate programs. Prospective employers can also use the information to contact measurement programs regarding job opportunities. If you have any more updates for your institution, think your institution should also be included in the listing or have any other related comments, please contact Ye Tong at 319-335-5581 or ye-tong@uiowa.edu.

Call for Papers

Educational Measurement: Issues and Practice

Educational Measurement: Issues and Practice (EM:IP) is seeking manuscripts. *EM:IP* publishes papers on a range of topics relevant to psychometricians, test developers, legislators, educational policy analysts and K-12 and post-secondary educators. *EM:IP* articles have ranged from technical papers addressing such topics as differential item functioning and validity theory to commentaries on the impact of testing on curriculum and instruction.

EM:IP's primary purpose is to promote a better understanding of, and reasoned debate on, timely measurement issues of practical importance to educators and the public. *EM:IP* is a vehicle for improved communication among NCME members and between NCME members and others concerned with educational measurement issues and practices. *EM:IP* articles are intended to illuminate issues in educational measurement with the goal of informing educational measurement practice. It is not as technical as its companion, the *Journal of Educational Measurement.*

EM:IP is seeking manuscripts that address measurement issues of concern to practitioners and academics, applications of measurement techniques in educational settings, and exemplary practices. Acceptable manuscripts could also deal with novel measurement techniques to assess educational objectives, controversial measurement issues, surveys of practices and changes in practice, and public critiques of testing and test use. Detailed information on the manuscript review process can be found at http://www.ncme.org or www.blackwellpublishing.com/emip