NCME national council on measurement in education

VOLUME 20, NUMBER 3 NEWSLETTER SEPTEMBER 2012

FROM THE PRESIDENT

Gregory J. Cizek, University of North Carolina at Chapel Hill

Hello Colleagues:

Thanks to all of you who have worked so hard over the past few months on some substantial and important tasks that are critical to our profession and indeed to the broader educational, scientific, and research communities and the general public.

2013 NCME Annual Meeting Program

Among those tasks is the production of the NCME Annual Meeting program. First, thanks to training program co-chairs Leslie Keng (leslie.keng@pearson.com) and Ye Tong (ye.tong@pearson.com). More information on the training program workshops will be forthcoming in the next newsletter.

Thanks also to annual meeting program co-chairs, Kimberly O'Malley (kimberly.omalley@pearson.com) and Elaine Rodeck (erodeck@mcc.ca), who are organizing an innovative, cutting-edge program for the 2013 annual meeting in San Francisco. The program is organized around the theme of "Building on the Past, Reaching for the Future: Innovative Ideas to Design NextGen Assessments." It will feature contemporary content, increased use of multimedia in sessions, and debates on vital issues of the day. The co-chairs have identified five key areas for which proposals were solicited:

1. Measuring Educator Effectiveness and Development
2. Connecting Assessments and Learning
3. Performance Assessments and Automated Scoring
4. Meeting the Assessment Needs of all Students
5. Psychometric and Statistical Methods

As of this writing, a record number of proposal submissions has been received (n = 554) for individual and coordinated sessions. Out of the 554 submitted proposals, the majority (n = 388, 70%) were related to the Psychometric and Statistical Methods area; 47 (8.5%) addressed Connecting Assessments and Learning topic; another 47 (8.5%) focused on Performance Assessments and Automated Scoring topic; 39 (7.0%) centered on Meeting the Assessment Needs of all Students; and 33 (6.0%) emphasized Measuring Educator Effectiveness and Development. Each of the five key areas has been coordinated by, respectively, Amy Hendrickson (ahendrickson@collegeboard.org), Claudia Flowers (claudiaflowers@uncc.edu), Mo Zhang (mzhang@ets.org), Heather Buzick (hbuzick@ets.org), and Jonathan Weeks (jweeks@ets.org).

I want to also thank all NCME members who participated in the proposal review process. It is one of the most important roles that a member can play in service to the association, as it represents the primary quality control mechanism by which the rigor and relevance of the program is assured.

Revision of the 1999 Standards for Educational and Psychological Testing

In 2006, the Management Committee for the *Standards for Educational and Psychological Testing* (*Standards)* approached the sponsoring organizations (AERA, APA, and NCME) and asked whether a revision of the 1999 *Standards* should be undertaken. The organizations agreed that the time was right to revise the 1999 *Standards*. In 2007, Co-chairs for the Committee to Revise the *Standards for Educational and Psychological Testing* were appointed; the Committee membership was finalized in 2008. The Committee has made regular progress reports on the revision of the Standards at the Annual Meetings of AERA, APA, and NCME.

The subsequent revision process has involved the efforts of many NCME members who have participated in drafting new sections of the *Standards*, reviewing drafts of the revised *Standards*, and providing input and feedback to the Management Committee at many key junctures over the past several years. A public call for comments was also made, and nearly 400 recommendations from 71 organizations were received. NCME provided a nine-page outline of concerns. The NCME Standards and Test Use Committee (STUC) has provided feedback to the Committee to Revise the *Standards for Educational and Psychological Testing* on a 2011 draft of the *Standards*. A second draft version of the *Standards* that accommodated input from the STUC and the other two sponsoring organizations was completed by the Committee to Revise the *Standards for Educational and Psychological Testing* earlier this year, and has recently undergone both legal review and editing. This version is now ready for a final review by the sponsoring organizations.

The revision process is now in its final stages. The NCME Standards and Test Use Committee (STUC) will review the latest draft of the full *Standards* and provide the NCME Board with its recommendations later this fall. Finally, the full NCME Board will carefully review both the entire set of revised *Standard* and the STUC recommendations in anticipation of a vote on NCME's endorsement of the *Standards* at their first meeting during the Annual Meeting in April 2013.

As is perhaps obvious from the comprehensive and conscientious work that has taken place so far, numerous NCME members have contributed greatly to the revision of the *Standards*. Thanks to everyone who has provided their expertise and input on what is arguably the single-most important lever for promoting sound professional practices. Ultimately, the revised *Standards* will help ensure that the information provided by tests is accurate, dependable, meaningful, and useful for the students, clients, educators, researchers, clinicians, policy makers, and others served by the results generated from the ever-improving tools we develop, administer, interpret, and evaluate.

GREETINGS FROM THE EDITOR

Susan Davis-Becker, Alpine Testing Solutions

In this issue we feature the second column from NCME president, Gregory Cizek, who discusses the preparation for the 2013 annual meeting as well as some major initiatives within our organization. In addition, much of the content of this Newsletter is inspired by the upcoming 75th anniversary of NCME. Neal Kingston provides us a look back at 1938 and the measurement world at that time. For the graduate student column, Jerome Clauser shares the insights on the past and future of mental measurement from several notable members of our organization. Joni Lakin, on behalf of the NCME Recruitment Committee discusses strategies for recruitment of the next generation of measurement professionals. We have also continued our member perspectives on the future of NCME in this issue with contributions from Dianne Henderson Montero, Leslie Keng and Laurie Davis, and Dubravka Svetina. Our Member Spotlight features Edynn Sato of WestEd. This is followed by a number of important announcements and calls for award nominations. Finally, this issue concludes with a feature about the life and work of Nancy Burton, who passed away in August.

Thank you to all who contributed to this issue!

IT'S 1938: A LOOK BACK AT THE FIRST YEAR OF NCME

Neal Kingston, University of Kansas

Test Theory

It's 1938, and although the basic structure of classical test theory is well established, it has been only seven years since Thurstone published the first book on test theory and two years since Guilford published the second book. Kuder and Richardson just published their article on reliability in 1937.

It's 1938, and there is no Rasch model, no item response theory, and no generalizability theory.

Computing

It's 1938, and you use statistics like the Kelley D and the new KR-21 because more exact statistics take too long to calculate. But if you are real lucky you have an expensive mechanical calculator that simultaneously can compute sums and sums of squares!

It's 1938, and there is no Bayesian posterior distributions, no marginal maximum likelihood estimation, and no Markov Chain Monte Carlo; performing a factor analysis can take months.

Tests and Testing Companies

It's 1938. The World Book Company became the first publisher of group-administered intelligence tests 20 years earlier and added the Stanford Achievement Test to its line of products 15 years ago and the Metropolitan Achievement Test six years ago.

It's 1938, and the Psychological Corporation was founded 17 years ago by James McKeen Cattell and two of his former students, Edward Lee Thorndike and Robert Sessions Woodworth. Just as the name implied, they published tests for psychologists.

It's 1938. The California Test Bureau was founded 12 years earlier in Los Angeles. Ethel and Willis Clark sent out 25 penny post cards to school districts advertising the Los Angeles Diagnostic Tests in the Fundamentals of Arithmetic. One year later, Kansas City was their first customer.

It's 1938, and three years earlier Lindquist and his colleagues published the Iowa Every Pupil Test of Basic Skills, which soon was marketed by Riverside Publishing.

It's 1938, and there is no ACT, no AIR, no Data Recognition Corporation, no ETS, no Measurement Incorporated, no Measured Progress, no Pearson Educational Measurement, and no Scantron.

It's 1938, and it has been only 23 years since the first large-scale assessment that included selected response items – the Kansas Silent Reading Test – and only 21 years since multiple-choice testing had its first great success with the Army Alpha.

Testing Standards

It's 1938, and there are no joint testing standards, no code of fair practices, and no peer review of state testing programs. But a young professor at Rutgers University, Oscar Buros, realized that if we do not police our profession those with less knowledge will choose to do so; he was about to publish the *1938 Mental Measurements Yearbook*, with learned reviews of more than 400 tests.

Graduate Programs

It's 1938, and there are not very many places to study educational measurement in the United States. You could study with E.L. Thorndike, Irving Lorge, and R.L. Thorndike at Teachers College; L.L. Thurstone, Karl Holzinger, and Harold Gulliksen at the University of Chicago; E.F. Lindquist at the University of Iowa; Ralph Tyler at Ohio State University; or Truman Kelley and Lewis Terman at Stanford (though Kelley moved to Harvard later in 1938). Anne Anastasi was a psychology instructor at Barnard College and thus did not teach graduate courses.

It's 1938, and Ledyard Tucker and Lee Cronbach are in graduate school in Chicago. Fred Lord has not yet started graduate school.

Organizations

It's 1938, and the American Psychological Association has about 600 members. Founded in 1892, many APA past presidents were known in the field of measurement (James McKeen Cattell, 1895; E.L. Thorndike, 1912; Robert Woodworth, 1915; Robert Yerkes, 1917; Lewis Terman, 1923; L.L. Thurstone, 1933, to name a few).

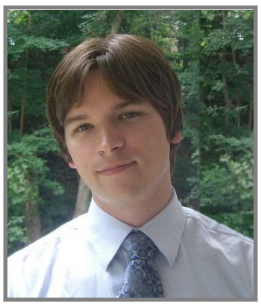
The Psychometric Society is two years old.

It's February 1938, and 45 college teachers and 12 persons from state departments of education or public schools adopted a constitution for the National Association of Teachers of Educational Measurement. Earl Bennet South is elected as the first president. In 1943, the organization would change its name to National Council on Measurements Used in Education, and in 1960 the name would be changed to the National Council on Measurement in Education.

NCME

That was 1938, and next April the National Council on Measurement in Education is celebrating its 75th anniversary. Test theory, computing, tests and testing companies, graduate programs, people, and even the name have changed. But the importance of what we do has not, and you have a major role in NCME's future. Many special events are planned for our conference. We hope to see ALL of you there!

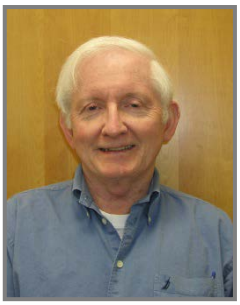
GRADUATE STUDENT CORNER: DISCUSSIONS ON THE PAST AND FUTURE OF MENTAL MEASUREMENT



Jerome Clauser, University of Massachusetts - Amherst

As graduate students begin a new academic calendar year, it is an ideal time to reflect on our past and plan for our future in the field of mental measurement. With that in mind, this column will be devoted to an interview conducted with three of the most notable and accomplished thinkers in our field: Drs. Ronald Hambleton, Michael Kane, and Wim van der Linden.

Ron Hambleton is a Distinguished University Professor at the University of Massachusetts Amherst. During his tenure at UMass he has written more than 600 publications including some of the foundational papers on criterion-referenced testing, item response theory, and international testing. Michael Kane is the Samuel J. Messick Chair in Test Validity at ETS. During a distinguished career he has published more than 75 articles and book chapters on test validity including the seminal chapter on validation in the fourth edition of *Educational Measurement*. Wim van der Linden is the Chief Research Scientist at CTB/McGraw-Hill and is responsible for leading the organization's research agenda. During his career at CTB and the University of Twente in the Netherlands, he has been a consistent innovator in the field and has published groundbreaking work in the area of adaptive testing, response time analysis, and automated test assembly. He is currently the NCME president-elect.



Above: (Left) Dr. Ronald Hambleton, Distinguished University Professor at the University of Massachusetts Amherst; **(Center)** Dr. Michael Kane, Samuel J. Messick Chair in Test Validity at ETS; **(Right)** Dr. Wim van der Linden, Chief Research Scientist at CTB/McGraw-Hill.

I personally admire each of these men and am extremely grateful to have had the opportunity to conduct this interview. I believe their responses reflect the depth of knowledge, wisdom, and experience that has characterized each of their careers.

What do you feel has been the most significant or unexpected change in the field over the course of your career?

Hambleton: Since 1969 when I began my career as an Assistant Professor at the University of Massachusetts, I have seen at least three significant changes in assessment theory and practice. The first was the transition from a primary focus on norm-referenced testing (NRT) in the 1950s and 1960s to a focus since 1970 on both criterion-referenced testing as well as NRT for addressing assessment practices. Looking at candidates in relation to a well-defined domain of content is a main component of student assessment today. The second has been the transition from paper-and-pencil assessments to assessments administered at a computer. At the computer, we have many new test designs (with "adaptive" designs being the most popular), new item types to assess important skills including those using audio and visual components and free response, and improved approaches for score reporting. The third has been the transition from classical to modern test theory (i.e., item response theory) and practice. While IRT models require that several strong assumptions be satisfied, when they can be, these IRT models provide the basis for better approaches to test development, item banking, computer-based testing, and test score reporting.

Kane: I had a professor in college who maintained that the goal of research is not to move from confusion to certainty, or even clarity, but rather, to move from confusion to a higher level of confusion. Think about it; the cosmos is much more mysterious now (with curved space, black holes, and dark matter) than it was just over 400 years ago, when Galileo built his first telescope. Progress is more about the depth and sophistication of our questions than it is about the certainty of our answers. I came into the field in the early 1970s, when criterion-referenced testing was being proposed as an alternative to the then standard norm-referenced approach. IRT was in its infancy, generalizability was being formulated, and validity came in three flavors. Things were relatively simple. In the last 40 years, we have uncovered more and more complexity. We still have an interest in rank-ordering individuals on unidimensional traits, but we recognize that for many testing applications, the variables

of interest are not unidimensional, and the intended uses go beyond rank ordering. We have faced the need to link the score scales for multiple test forms, and to set consistent performance standards across forms and grade levels. We recognize that item scores and test scores often have different meanings and implications for different groups. In order to address these issues, we have developed a range of increasingly sophisticated psychometric models, but these "answers" tend to generate deeper questions about what test scores mean and how they can best be used.

van der Linden: It is always easy to overlook the obvious when you answer questions like this. I remember a survey in which medical researchers were asked what they considered as the most significant change in the history of medicine. The common answer was not penicillin, the discovery of stem cells, or something spectacular as robotic surgery, but the flush toilet. The most significant change in educational measurement I have experienced is without any doubt the increasing importance of computers. Testing companies were among the early adopters of computers for their data processing. And several of them already owned mainframes when I began my career in the early 1970s. But back then, computers were treated as isolated wonders of electronic ingenuity; it was just impossible to foresee how common and pervasive they would become. Today, our field would be entirely desperate without them. We need them to bank our items, run IRT analyses, assemble and deliver tests, score test takers, and make tests adaptive. But there is more to come, for instance, automated item generation, the use of technologically enhanced test items, and embedding testing into instructional systems. And in our research, we rely more and more on Monte Carlo methods and the availability of open-source software (R). If you had asked me for more than one significant change, I would definitely have mentioned the huge developments in IRT (now increasingly applied outside of educational testing too) and the internationalization of our field. But computers are the absolute winners.

How do you foresee the field changing over the next decade?

Hambleton: I expect to see a continuation of research initiatives already started on many of the topics mentioned above: computer-based testing, development and validation of new item types, automation of test item scoring, and test score reports to improve the clarity and utility of score reports for policy-makers, educators, teachers, parents, and students. I think too, that the research of persons such as Rick Luecht on "assessment engineering" to generate templates to obtain test items matching desired content and statistical specifications is very important, represents the next generation of item development, and will further the construction of valid assessments, at lower costs.

Kane: Changes in testing over the next decade will continue to be shaped by technological and psychometric developments, but they will also be shaped by social, political, and economic trends, all of which are difficult to foresee. A trend that is already with us, and that is, I think, welcome, is a shift in emphasis away from reliance on generic psychometric models (e.g., of reliability, validity) toward analyses that focus on particular kinds of applications and contexts. For example, I think that estimates of standard errors and analyses of the impact of these random errors (and systematic errors) in specific applications are likely to prove more useful than estimates of generic reliability coefficients. I expect that we will get better at evaluating the salient issues in different testing applications.

van der Linden: I might be wrong, but for the next decade or so I don't foresee any revolution or major change of paradigm. Instead, I expect further technical sophistication, both of our research and applications. More of the same thus, but introduced at a much faster speed and higher level of complexity. Am I happy with this? Yes and no. I'm certainly supportive of technical perfection and will keep trying to contribute to it. But the real danger exists in the neglect of our conceptual basis. I'm sometimes shocked by the blind applications of techniques I see. The NCME annual meetings have definitely shown a trend toward more papers with minor technical tweaks. The same holds for our journal articles. If you ask their authors for a motivation, you hardly get an answer. Some of our research areas, I believe, would definitely benefit from a conceptual cleanup. We need technical progress to find solutions to practical problems. But each good solution begins with a correct conceptualization.

What advice would you give to graduate students as they prepare for this future?

Hambleton: Probably by the time you read this answer below you have already committed to a doctoral program. I will skip the topic of graduate training program selection here. But with the broadening of our field in areas such as psychometric methods, statistical methods, and cognitive psychology, I doubt that any graduate programs can do it all, even the best ones. Therefore, make sure that you expand on your university training by attending regional, national, and even international meetings to hear presentations from technical leaders, and network with persons doing interesting work. If at all possible, try to present your own research at these meetings to gain experience and to begin to build your reputation as a serious scholar. Job hunting will be more successful for you. Summer internships are another activity where you can gain experiences that are different from those you can gain working in the university context, and are valuable in the job hunting process. I would also encourage you to read on a regular basis the professional journals and research report series published by many universities and testing companies. Graduate students need to know what is being studied currently, who the main scholars on various topics are, and what these scholars consider to be the relevant ideas and follow-up studies. Don't make the mistake either of thinking

once you graduate, you know it all. Continue to attend training workshops on topics of interest, learn from your colleagues and students, attend conferences, get involved in applied assessment projects, and read the professional journals!

Kane: I would advise graduate students not to specialize too much. As the technical aspects of testing have become increasingly sophisticated, it is easy to get lost in the methodological details. It is, of course, important to be familiar with a wide range of models and methods (in the sense of knowing what they can do and how they do it) and to master some of these in detail, but it is also important to understand how different stakeholders see the problems to be solved. The issues in licensure testing are quite different from those in employment testing, both of which are distinct from those in educational placement testing. The fact that we have a shiny, new hammer does not make every issue a nail.

van der Linden: It has always struck me how many of my generation actually retired when they graduated. You hardly ever see a book on their desk; they have never subscribed to any serious journals. Given all new developments, clearly the safest road to perdition! Of course, most of us have extremely busy professional lives. But we should exploit every opportunity to hone our minds and skills. What helps me personally is a small collection of books (some ten titles on mathematics, statistics, and test theory), which I reread periodically. Some of them I've now worked through more than a dozen times. You would expect a diminishing return, but exactly the opposite is true. It is amazing how much you learn when you re-acquaint yourself with an earlier text. I've always one of them within reach (right now a linear algebra text), which I grab whenever I've 15 minutes left. You can do this with a book you already know, not with a completely new text. So here is my advice: Get your degree if you think you need it. But don't graduate, always remain a student.

INNOVATIVE RECRUITMENT EFFORTS FOR THE NEXT GENERATION OF MEASUREMENT PROFESSIONALS

Joni M. Lakin, Auburn University, and Members of the NCME Recruitment Committee

We've heard this story so often that we're beginning to think it's universal—"I didn't choose educational measurement, I just stumbled into it."[1] Sharing the story of our "accidental" careers has become the mainstay of conference cocktail hours at (we're guessing) every NCME conference the past 75 years.

These stories leave us with an important question: What can we do to help others *thoughtfully* choose educational measurement as a career? This was the topic of a recent article by Sara Finney and Dena Pastor (2012) of James Madison University, who shared their experiences and recommendations for how best to recruit undergraduate students into the field. This article has the potential to motivate more local efforts to recruit undergraduates into educational measurement. In fact, local efforts seem essential if a "personalized touch" is, as Finney and Pastor claim, the best way to recruit new members.

The recruitment of new contributors to the field of measurement is also part of the mission of the NCME Recruitment Committee. Like most NCME members, we all love what we do and want to help others discover the field and contribute to a vibrant and diverse measurement community.

One of our goals this year is to create materials—both online and off—to support the recruitment of new individuals to our profession. These materials would serve both as a stand-alone recruitment effort and to create additional resources for students whose interest is piqued by other recruitment efforts (perhaps by one of Sara and Dena's talks), but want more information. These resources might also be used locally by professionals for additional recruitment activities.

However, we don't want to work on this project in isolation. We know that the committee members, as well as Sara and Dena, are just a few of many thoughtful educational measurement professionals who have developed strategies to increase the visibility of our field. Therefore, we want to put the call out for your stories, materials, and tips on how to recruit individuals from a variety of backgrounds into educational measurement. In addition to recruitment talks for undergraduate psychology students, we feel we should consider a variety of strategies and target audiences in our efforts. Here are just some of the questions we are considering:

1. How can we most efficiently (and appealingly) provide recruitment information to undergraduates when personal visits are not always feasible?
2. Are there ways to get the message about careers in measurement to high school counselors and students?

[1] It may indeed be universal given the findings of the APA Task Force for Increasing the Number of Quantitative Psychologists (2009; http://www.apa.org/research/tools/quantitative/quanttask-force-report.pdf) cited by Finney and Pastor (2012).

3. What could we add to the "Prospective Students" and "Career Center" of our new NCME website to provide detailed and engaging information about our work?
4. What measurement-focused curricula should we advocate for undergraduate students?
    a. Is it realistic to have educational measurement majors or minors?
    b. Could we lay out some elective paths for undergraduates interested in graduate measurement programs?

These are just a few ideas, and we welcome any contributions you have to our efforts to create recruitment resources. If you have materials, stories, or other tips you'd like to share, please email them to ncme.recruitment@gmail.com.

References

Finney, S. J., & Pastor, D. P. (2012). Attracting students to the field of measurement. *Educational Measurement: Issues and Practice, 31*(2), 44–49.

THE FUTURE OF NCME: MEMBER PERSPECTIVES, PART 2

In the last issue of the NCME Newsletter, we introduced a new series featuring member perspectives on the future of NCME. In this current issue, we present additional perspectives from Dianne Henderson Montero, Leslie Keng and Laurie Davis, and Dubravka Svetina.

Protecting the Public from Unsound Measurement Practice in Education

Dianne Henderson Montero, Educational Testing Service

In my career as a measurement professional, I've been working with other experts and colleagues at testing companies designing, developing, and producing assessments to meet a number of different purposes for a wide range of customers and clients. While the purposes underlying design and technical details of the assessments may vary, the most important quality of any assessment is the accuracy of the resulting scores. Once reported, these scores are used to make decisions or judgments about individuals or groups of people. So we expend a great deal of effort on making sure that the scores are accurate, then understanding and explaining the results. It is through these experiences that I have come to the opinion that the one way that NCME could advance the science and practice of measurement in education in the next 25 years would be to have a clearer and more direct focus on protecting the public. As a professional organization, I believe that NCME needs to consider our role and obligation in becoming the recognized authority in measurement science and take clear steps to protect the public from unsound measurement practice in education. In this essay I offer three suggestions for consideration and debate by the membership:

1. Active and Visible Education of the Public
2. Enforcement of the Standards
3. Evaluation of Member Competence

All three suggestions, while departures from the current focus of the organization, fit within the stated mission of NCME: to clearly advance the science and practice of measurement in education. The following paragraphs describe these suggestions in more detail.

<u>Active and Visible Education of the Public</u>

Today there is no recognized, independent professional authority actively educating the public, policy makers, and users of educational assessments on the current state of measurement science and the limitations associated with this science. Measurement professionals are faced with increasing pressure to produce assessments that go beyond current capabilities or that use approaches with little to no foundational research. To address these challenges we need to educate the public—including potential customers, clients, stakeholders, and policy makers—and the content of this information needs to come from an organization with recognized authority on the subject matter. Without a strong single authority advocating for what measurement science can do, the education of these various entities rests with the measurement professionals involved and the companies and organizations that employ them. Unfortunately, the public often views the advice provided by these measurement professionals with some amount of suspicion, wondering if they are being unnecessarily cautious or simply more concerned about maximizing revenues. This same information presented by an independent professional organization would help alleviate these suspicions. As a focus for the next 25 years, NCME can and should become the independent authority on measurement science in education, advocating for sound and ethical measurement practices.

<u>Enforcement of the Standards</u>

In addition, there is no professional body certifying that all educational assessments delivered to individuals or groups are based on sound measurement science and principles and meet the minimum requirements described by the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999). Currently, there are a number of assessment products and services that can be purchased by the public that vary a great deal in quality. Some are constructed to be consistent with research findings and the *Standards*. Others are not. While there are a number of testing companies that follow the *Standards* and have procedures to ensure that their professional staff is following them, these procedures are not independently verified. Consequently, the public must rely entirely on the internal checks and balances within each of the organizations. Institutions such as the Buros Center for Testing publish reference information and evaluations of commercially available tests, but do not cover custom products like state assessments, or those products that have insufficient information about technical quality. Therefore, as a focus for the next 25 years, NCME can and should consider providing ways for the public to quickly ascertain whether a product meets minimum quality standards as recognized by an independent entity. This could take the form of a partnership or joint venture with existing organizations, or NCME could expand its mandate and begin to certify products and services directly. NCME should also consider making available to the public an accessible, searchable database of certified assessments. These steps would provide the public with the assurance that the product they are purchasing meets minimum standards and would reduce the possibility of time and money lost on poor quality products.

Evaluation of Member Competence

Further, there is no organization with the obligation to protect the public from under-qualified people involved in the development of educational assessments of individuals or groups, and/or in the evaluation of educational programs. Currently there are no minimum standards that a person practicing as a measurement professional must meet, no ongoing evaluation of competence, and no accountability to an independent authority for their practice. This means that essentially anyone can design, construct, or use a standardized test, or any form of assessment, without ever having taken a measurement course. They need not show any evidence of membership in a professional association such as NCME, APA, or AERA; they do not have to demonstrate competency in any aspect of the work that they perform; nor are there any requirements for keeping up-to-date with the latest research and best practices. We rely heavily on self-regulation, hoping that people will not go beyond their level of competence. Unfortunately this leaves the public vulnerable and potentially exposed to procedures that do not follow best practice, as well as errors of judgment and omission that could have a significant impact on individuals or groups. For example, anyone can identify themselves as a psychometrician and conduct highly technical work such as scaling items using an item response theory (IRT) model without knowledge of IRT, the underlying assumptions associated with the model, or an understanding of the software being used. To ensure the protection of the public, I propose that as a focus for the next 25 years, NCME can and should develop ways of evaluating the competence of our own membership, particularly for those providing services directly to the public. Recognizing that the complexities associated with this suggestion deserve a detailed evaluation of the merits and challenges that go beyond the scope of this essay, I propose that NCME commission a task force to evaluate the various approaches that could be used for evaluating the understanding of basic measurement science and methods of demonstrating ongoing competence in specific specialty areas.

Let me conclude this essay by asking each of you to consider the following: to what extent does NCME have an obligation to protect the public, and how should we meet this obligation? I think that we do have an obligation, and I have offered three suggestions that I believe will help to protect the public from unsound measurement practices. By offering this vision of the future of NCME, I hope that this will stimulate the desired conversation among the membership and current and future leadership.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Expanding the "M" in NCME

Leslie Keng and Laurie L. Davis, Pearson

In the previous issue of this newsletter, Dr. Chad Buckendahl wrote an insightful article titled "Defining the 'E' in NCME". In this article, we have decided to focus on the "M" in NCME. The primary purpose is to share our perspectives on how our role as measurement experts and professionals will need to evolve and expand as the field of educational measurement changes in the 21st century.

How the Field Is Changing

In recent years, there has been considerable demand for educational assessments to go beyond simply evaluating a student's knowledge, skill, and understanding of a particular curriculum. This demand is evident in educational policies at both state and federal levels as well as the expanding areas of research in educational measurement that go beyond the traditional psychometric topics. Next generation assessments must not only do more, but must do so engagingly and effectively. Moreover, their outcomes should be easily consumable by various stakeholders.

What does the field want next generation assessments to do? Here is a sample of questions that next generation assessments are expected to help answer.

- Is my student on track to succeed in a future grade or course?
- Is my child ready for opportunities after high school, such as college or careers?
- Did my student make enough progress since the previous assessment?
- Is my child's teacher effective?
- What do my student's test results tell me about the instruction he or she needs to improve learning in the next course/grade level, or in the current course/grade level?
- My student is still learning English. Can you tell me whether he or she is learning the curriculum despite the language barrier?
- My child is a special education student. How can I find out how much he or she has learned and what tailored instruction is needed?

Simply creating a reliable and valid assessment—while still vital and essential—is no longer the only goal. The field wants assessments that can engage all types of test takers by using the latest technology and innovative delivery methods while maintaining the psychometric integrity and defensibility of the assessments. Examples of these requirements include:

- Technology-enhanced test items
- Gaming application to assessments
- Delivery of assessments on various digital devices, from desktops, laptops to tablets and even smart phones
- Linguistically accommodated test items
- Tests and items that are accessible by students with various types of physical or cognitive challenges

Finally, results from next generation assessments provided back to students, parents, teachers, and the general public need to be easy to understand, visually appealing, and accessible in different modes and formats. In addition, the development of next generation assessments must involve a collaborative effort between people with varying expertise, from numerous organizations and agencies, across different states and even countries.

How Measurement Professionals Need to Change

What do all of these new expectations for assessments mean to both aspiring professionals and veterans in the measurement field? In short, a "next generation psychometrician" must likewise go beyond understanding the traditional topics found in a measurement textbook, such as reliability, validity, scaling, and equating. Knowledge of these technical topics is still essential. However, next generation psychometricians must also look to expand their knowledge and skill sets so that they can:

- Understand the "big picture" – knowing how decisions will impact the assessment program not only psychometrically, but also instructionally, politically, and legally
- Explain psychometric concepts – providing value by consulting, guiding, advising, and educating assessment stakeholders
- Keep up with the latest trends – staying on top of the latest developments in technological innovations, methods of delivering assessments to different types of test-takers, and ways of summarizing and displaying data for various users
- Generate innovative yet defensible solutions – suggesting creative methods to solving practical challenges while maintaining psychometric integrity
- "Play nice" with others – working together with various personnel and groups to develop assessments that meet the "big picture" goals of the assessment program

It is a daunting task, yet we are not left to our own devices in this endeavor. In fact, an overwhelmed measurement professional needs to look no further than his or her NCME membership for "next generation" development opportunities. The theme of "Building on the Past, Reaching for the Future: Innovative Ideas for NextGen Assessments" for the upcoming 2013 annual meeting shows NCME's commitment of helping its membership directly address the demands of the changing field. This is evident in the conference's call for proposals, which encourages members to submit proposals for sessions with topics that go beyond the traditional "psychometric and statistical methods" category. Participants planning to attend the conference can

expect to see a program that includes a plethora of sessions in the areas of educator effectiveness and development, connecting assessments and learning, performance assessments and automated scoring, and meeting the assessment needs of all students.

Additional resources and learning opportunities are also available to measurement professionals from the various testing organizations and research institutions, many of which are trying to stay on top of demands of the rapidly evolving field of educational measurement through research and development initiatives. One such example is the work being conducted by the Research and Innovation Center at Pearson. The interested reader can view a series of videos showcasing the type of projects at the Center: http://www.pearsonassessments.com/pai/ai/Products/NextGeneration/videoseries.htm

Winston Churchill once said, "The pessimist sees difficulty in every opportunity. The optimist sees the opportunity in every difficulty." The demands of next generation assessments may seem intimidating. However, if the measurement community can seize this opportunity and devise creative and effective solutions to address the various demands, we may be standing at the precipice of an exciting breakthrough era for the field of measurement.

The Future of NCME: The Role of Professional Organizations

Dubravka Svetina, Indiana University

As the NCME 75th anniversary approaches, I started to reflect on the role professional organizations play in the larger context of intellectual community of any field. This reflection gave me an opportunity to ponder what our organizational goals are and the actions its members can take in order to achieve them. With that in mind, I would like to contemplate two aspects of our mission, vision, and goals statement.

'Promote awareness of measurement in education as a field of study and work to encourage entry into the field and interdisciplinary collaboration.'

Promoting awareness of the field of measurement could be done in a variety of contexts. This summer, a timely issue of summer 2012 NCME's publication *Educational Measurement: Issues and Practice* (EM:IP) speaks to several such contexts where this and other goals of our organizations can be attained. I encourage members to read the entire issue, as each article truly opens a discussion about how to achieve our goals. Let me mention a couple of them. In their article on attracting students to the measurement field, Finney and Pastor (2012) provide advice on how to inform (and recruit) others about the field of measurement, while recognizing the existent challenges. Once we recruit students, one such challenge is to strengthen their commitment to the field of measurement. As the authors suggest, we ought to offer experiences that will contribute to the development of one's identity as a measurement professional; while this may not be an easy task, it is nonetheless an important part of the mission statement.

In a somewhat different direction, in their article on teaching measurement courses, Bandalos and Kopp (2012) provide a summary of current practices of those who teach such courses, topics that are covered, as well as offer some useful activities to include promoting student learning. One of the highlights of the article was the authors' recognition that there is a need for measurement literacy, similar to that of statistical literacy (p. 8). We should welcome opportunities to promote awareness of measurement concepts and issues. We should take ownership in contributing to the growth of measurement literacy; for example, we could do it as part of our teaching responsibilities, intellectual exchanges at the annual meeting, or development of innovative interactive exhibitions about testing (e.g., see Allalouf & Alderoqui-Pinus, 2012). The opportunities to aid to this goal are ample; we just need to be open to work towards carrying forward our mission. The articles in EM: IP's summer 2012 issue remind us of just that.

'Disseminate knowledge about educational measurement…'

In the guest editorial of the previously mentioned EM:IP issue, Allalouf and Sireci (2012) address the very challenge of dissemination of measurement concepts to the public (p. 1). I believe that the dissemination of knowledge about educational measurement has come a long way in the last 75 years, and with innovative approaches, we may face current and future challenges adequately. Allow me to comment on three approaches in which we should partake in order to accomplish this goal.

First, in examining the annual meeting programs over the past decade or so, I observed that the workshop offerings increased in both the numbers and 'diversity' of topics. Even as late as the early 2000s, only a handful number of workshops were offered; a limited selection of topics would typically include reliability, equating, tips for graduate students, and standard setting. These topics are still important and relevant, as continuous demand for them suggests. In addition, the educational community faces a number of new challenges in various aspects of measurement practice and assessment. Credit is due in large part to the planning committee and workshop/training chairs, who organize workshop offerings that reflect the current climate of educational community. Challenges of innovation in assessment, assessment of the skills, strengths, and weaknesses of our

students in the 21st century, technical and practitioner's challenges present in the age of accountability, and legal reforms have been reflected in both the annual meeting program and the workshops in the recent years.

Second, I echo Randy Penfield (2012), who previously discussed the importance of dissemination of measurement-related concepts and ideas. I, too, use ITEMS modules in my class and adopt them to create class activities. I find that articles like these provide useful information for a variety of audiences: students, practitioners, scholars, as well as policy makers. ITEMS are quite efficient in disseminating various measurement issues to anyone interested to know more about them.

Lastly, developments in technology and popularity of social media have also contributed to the goal of dissemination. A very recent example is the utilization of online environments, in particular the introduction of webinars. Currently, only two such webinars are available on the NCME's website; however, it is conceivable that a large number of sessions like this would become over time a rich resource depository.

It is my hope that activities such as workshops, ITEMS, webinars, and resources of similar continue to be offered and grow for many more years to come because they are valuable and efficient ways to achieve this goal. However, a caveat needs to be made: given the complexity of measurement topics, they cannot be mastered in one session or by reading one article. Rather, we should use these outlets as a gateway for further learning.

References

Allalouf, A., & Sireci, S. G. (2012). Guest editorial: Dissemination of measurement concepts and knowledge to the public. *Educational Measurement: Issues and Practice, 31*(2), 1.

Allalouf, A., & Alderoqui-Pinus, D. (2012). Museum exhibition on testing and measurement: Scientific principles, social impact of testing, and dialogue with the public. *Educational Measurement: Issues and Practice, 31*(2), 14-20.

Bandalos, D. L., & Kopp, J. P. (2012). Teaching introductory measurement: Suggestions for what to include and how to motivate students. *Educational Measurement: Issues and Practice, 31*(2), 8-13.

Finney, S. J., & Pastor, D. A. (2012). Attracting students to the field of measurement. *Educational Measurement: Issues and Practice, 31*(2), 1-7.

Penfield, R. (2012). Extending our impact: Innovative dissemination for a broader audience. *NCME Newsletter, 20*(2), 11-12.

SPOTLIGHT ON THE PEOPLE WHO MAKE OUR ORGANIZATION GREAT – EDYNN SATO, WESTED

This month our member spotlight features Edynn Sato, Associate Director of the Assessment and Standards Development Services Program at WestEd.

How did you get into the field?

It seems to me that I was fortunate to be at the right place, at the right time, having what ended up being a life-changing conversation. I was about to complete my Masters in Educational Psychology and was having a conversation with one of my professors, Dr. Eva Baker, about what I was planning to do next. What I was planning to do was to try to find a job in which I could focus on developing instructional materials, applying what I had learned about how children learn and what motivates children to learn. Long story short, as a result of my conversation with Eva, I had an "aha moment," and I realized (and really appreciated) the critical role of assessment in effective instructional practice. I realized that measures of what students know and can do, ones that yield accurate outcomes that can be validly interpreted, are essential to being able to provide students the appropriate resources and supports they need to thrive and succeed. "The rest is history"—I applied for the doctoral program and got a graduate student research assistant position at the Center for Research on Evaluation, Standards, and Student Testing at UCLA. And that was my entre into the field.

If you weren't doing this, what would you do?

When I was accepted to UCLA for my Masters, I had actually applied for the teaching credential program. I was planning to become an elementary school teacher. It seems the powers that be thought otherwise, but if I weren't doing what I'm currently doing, I'd likely be teaching kindergarten or first-grade students and be on the implementation end of the assessments that I've been involved in designing/developing.

What advice would you have for graduate students who want to get into this field?

I recently had a conversation about this with a graduate student who did a summer internship with me. My advice is, in addition to delving into the substantive and technical content of our field, to gain an understanding of the contexts to which measurement is applied, or in which it occurs, and the uses for the information it yields. And, don't forget to test assumptions, because our context is complex (student characteristics, content characteristics, conditions, interactions, etc.). I believe that this will better ensure meaningful outcomes, help us avoid unintended consequences, and possibly have a more positive impact on educational practices more broadly.

What do you like to do for fun outside of work?

I enjoy spending time with friends, hiking, biking, eating(!), going to museums, and reading. Though I'm doing very poorly at it, I'm trying to learn how to play the guitar. And, I also volunteer in Golden Gate Park.

What would you say has been one of the biggest innovations in psychometrics in the last decade or two?

There have been a good number of valuable innovations in psychometrics in recent decades, but if I were to name one, I would say that it is in the area of "innovative items and tests". We seem to be moving forward in our measurement of knowledge and construction of instruments and procedures with an approach informed by an increasingly better understanding of how students learn, including students from diverse backgrounds. With more refined theories of learning, as well as with (the beginnings of) leveraging the technology that is available to us, innovative items/test formats provide us a mechanism that can be sensitive to individual differences/needs, and can enable students to more fully demonstrate what they know and can do.

When you go to conferences, how do you pick what sessions to attend?

For the topics that are most relevant to what I do and in which I am most interested, I typically select sessions that fall into one of the following three categories: (1) sessions with more established presenters focused on conceptual work or technical issues, so that I can gain clarity or greater insight or depth on the content of the presentations; (2) sessions with more established presenters focused on applied research and outcomes, so that I can gain exposure to the contexts and more practical issues relevant to the content of the presentations; and (3) sessions with "early career" presenters, so that I can be exposed to possibly newer or different perspectives or approaches to various issues of interest.

Who has been a significant influence in your professional life?

It's difficult for me to name an individual because I have been very fortunate to have benefitted from the mentorship and professional generosity of so many people. However, Dr. Eva Baker, as I mentioned in a previous response, was very influential in starting me along this professional path. She established and encouraged a culture in CRESST that supported dynamic discussion, out-of-the-box thinking, reasoned risk-taking, and a principled approach to the work in which we engaged. I was fortunate to have colleagues and mentors at CRESST who influenced my professional development in such positive ways. And, it would be remiss of me to not mention my current colleagues and mentors, particularly those at WestEd—I am grateful to be enjoying a similar type of culture of support and engagement in the work that I do under Dr. Stanley Rabinowitz's mentorship and support.

NCME Member Li Cai Named As Presidential Outstanding Early Career Scientist

The following is an excerpt from a press release issued by CRESST at UCLA on July 23, 2012. The full press release can be found here:
http://www.cse.ucla.edu/downloads/files/LiCaiNewsReleaseFINAL.7.24B.12.pdf

President Obama announced that UCLA Associate Professor Li Cai was named as a Presidential Early Career Scientist for 2012. A professor in UCLA's Graduate School of Education & Information Studies and Co-Director of the National Center for Research on Evaluation, Standards, and Student Testing, Professor Cai was the sole recipient from the U.S. Department of Education. He was honored for his early contributions to improved measurement methods, particularly in the area of statistical computing.

The Presidential Early Career Award for Scientists and Engineers is the "highest honor bestowed by the United States Government on science and engineering professionals in the early stages of their independent research careers," according to a White House news release. Ninety-six researchers were named ranging from the National Aeronautics and Space Administration to the Department of Health and Human Services.

"Discoveries in science and technology not only strengthen our economy, they inspire us as a people," President Obama said. "The impressive accomplishments of today's awardees so early in their careers promise even greater advances in the years ahead."

"I am deeply honored to be recognized by the President and my fellow scientists so early in my professional life," said Li Cai. "I couldn't have possibly come this far without the support of the school and my colleagues at CRESST. I look forward to the challenging and important work before us in the field of education."

"It's an understatement to say that our school is delighted by President Obama's recognition of Li Cai's early contributions to research and teaching," said Megan Franke, Interim Dean of UCLA's Graduate School of Education & Information Studies. "Professor Cai exemplifies the dedication and commitment needed to consistently improve education in our nation."

President Clinton established the Presidential Early Career Scientist Awards in 1996 to recognize individuals for their pursuit of innovative research at the frontiers of science and technology and their commitment to community service.

"Li Cai is a brilliant scholar and his recognition as the only recipient from the Department of Education brings honor to his field," said CRESST Director and GSE&IS Professor Eva Baker. "In his role as new Co-Director of CRESST," added Baker, "we look forward to his leadership in meeting the challenges of scholarship in the future."

ANNOUNCEMENTS: NCME

CALL FOR NOMINATIONS FOR THE ROBERT L. LINN DISTINGUISHED ADDRESS AWARD

AERA Division D welcomes nominations for the 2013 Robert L. Linn Distinguished Address Award. The Award recognizes the important contributions of Robert L. Linn to educational measurement and assessment policy. This annual award honors a scholar whose work bridges educational measurement and some other significant area of research (e.g., assessment policy, learning theory, curriculum and instruction) and has resulted in a widespread positive impact on the field of educational measurement. These contributions may include theoretical or technical developments, conceptualizations of educational measurement issues that have enhanced public understanding of these issues, or innovative ideas that improve the validity and effectiveness of educational assessments.

The following criteria will be used in the award selection process:

1. The nominee should have a substantial publication record in educational measurement. A substantial number of the publications should reflect contributions to another area of research (e.g., assessment policy, learning theory, curriculum and instruction).
2. The nominee's work must be judged to be of high quality and to have had significant impact from the perspective of both measurement experts and scholars or practitioners in the second field.

The awardee will receive a cash award of $1,000 at the Division D business meeting and will be invited by Division D to give an address at the annual meeting of the American Educational Research Association. In addition, each year immediately following the address, the written version of the presentation will be submitted to an appropriate assessment or measurement journal.

Description of Nomination Materials

A nomination must include three items:

1. A 1-page summary of the nature and significance of the nominee's work and a description of how it bridges educational measurement and some other significant area of research or practice,
2. A copy of the nominee's vita, and
3. Copies of 1 or 2 articles that exemplify the nominee's bridging contributions.

Self-nominations are permitted.

Application Procedure

Submit the complete nomination packet electronically by January 27, 2013, to: Ronald K. Hambleton, Committee Chair (rkh@educ.umass.edu).

CALL FOR NOMINATIONS FOR THE 2013 AERA DIVISION D EARLY CAREER AWARD IN MEASUREMENT AND RESEARCH METHODOLOGY

Division D of AERA welcomes nominations for the 2013 Early Career Award in Measurement and Research Methodology. This annual award recognizes emerging scholars in the field of educational research and methodology. The award rotates annually among the three sections of Division D. The 2013 award will be in Quantitative Methods and Statistical Theory (Section 2). The winner will be announced and honored at the 2013 Division D luncheon and business meeting with a plaque and a $1,000 award.

Eligibility

To be eligible for the 2013 Early Career Award in Quantitative Methods and Statistical Theory, an applicant must have received the doctorate after April 1, 2007. The nominee must also have two or more unique papers as sole or first author either accepted for presentation at the AERA annual meeting or published in AERA journals since 2007 that address educational research and methodology issues with a focus on statistical methods.

Application Procedure

A complete nomination will consist of the following four items:

1. A letter of nomination from a professional colleague who is an AERA Division D member.
2. An additional letter of support that addresses the nominee's contribution to the field and the reasons why the nominee's work represents a significant contribution to the field.
3. Two papers presented at any of the last six AERA annual meetings or published in an AERA journal. The research must address educational research or methodological issues. The AERA annual meeting papers may be a revised version for publication. Candidates must be the first or sole author on these papers.
4. Nominee's curriculum vita.

Submit the complete nomination electronically by November 30, 2012, to: Marilyn S. Thompson, Early Career Award Committee Chair, School of Social and Family Dynamics, Arizona State University, m.thompson@asu.edu

OTHER CONFERENCES OF INTEREST

Big Issues in Testing: Improving Admissions and Learning in Higher Education

The Buros Center for Testing is proud to celebrate UNL's addition to the Big 10 conference by sponsoring a conference on testing issues in higher education, March 28-29, 2013. The conference will highlight research and insights regarding the current state of admissions and learning assessment at both the undergraduate and graduate level. Issues to be discussed include the extent to which testing and assessment produces desired result or effect, influence of the context in which assessment occurs, implications for special and minority populations, lessons higher education can learn from K-12 assessment for accountability, and challenges to thinking about in the future of testing and assessment for admissions and learning. For more information, please visit the conference web site: http://buros.org/big-issues-testing

IN MEMORY OF NANCY BURTON

Nancy W. Burton died at her home in Ringoes, NJ on August 19. She was born in New Haven, CT and spent most of her early years in the Denver, CO area before moving to New Jersey in 1981.

Nancy received her Ph.D. in Educational Research and Evaluation Methodology from the University of Colorado, Boulder in 1972. She worked for the National Assessment of Educational Progress (NAEP) in Denver until 1979, when she received a two-year post-doctoral fellowship at the L.L. Thurstone Psychometric Lab at the University of North Carolina at Chapel Hill. In 1981, she went to work for Educational Testing Service of Princeton, NJ. Nancy worked at ETS for 25 years in several positions—including Measurement Statistician, Program Director, and Senior Research Scientist—before retiring in 2006.

Nancy was an active member of NCME for over 30 years. Her contributions to educational research focused on the validity of assessments with a special emphasis on fairness for gender, ethnic, and non-native English speakers. One of Nancy's major professional accomplishments was an analysis of the construct of verbal reasoning from the point of view of current cognitive theory. While at ETS, Nancy directed an integrated program of research on the validity of the GRE general test. She also directed studies of the validity of SAT item types, studies of minority student success in AP courses, a review of studies predicting long-term success in college, and a study of the construct validity of the SAT II Writing measure. Nancy was also integral to the development of the College Board's Admitted Student Evaluation Service, a general purpose validity study system delivered over the World Wide Web.

Daughter of the late John H. Walter, Nancy is survived by her mother, Mildred Walter, her daughter, Liz Burton, two brothers, and five nieces and nephews.

To get the NCME Newsletter four times a year (March, June, September, and December) go to
http://ncme.org/publications/newsletter/

Susan Davis-Becker
Alpine Testing Solutions
6120 Loma Circle
Lincoln, NE 68516

Phone: 402-483-5898
e-mail: susan.davisbecker@alpinetesting.com