NCME national council on measurement in education

Volume 18, Number 4 NEWSLETTER December 2010

FROM THE PRESIDENT: THE STATE OF OUR ORGANIZATION

Wayne J. Camara, The College Board

Colleagues:

In my first two columns I chose to try to provide historical references to national and state standards (see June 2010 newsletter www.ncme.org/pubs/pdf/vol_18_num_2_v3.pdf) and an analysis of the various standards and guidelines that address educational assessment (see September 2010 newsletter www.ncme.org/pubs/pdf/vol_18_num_2_v3.pdf). An observant reader may have even noticed that I attempted to change photos to correspond with the seasons.

In this column I will attempt to update the membership on NCME initiatives and organizational issues that we have addressed in the past few months. First, I want to recognize the commitment and effort of so many NCME members who have taken on key leadership positions to improve and advance our mission. As past president, Terry Ackerman has continued to lead many important initiatives for NCME. You may recall that he and Michael Rodriguez spearheaded the effort to provide training workshops to measurement professionals in less developed nations. This year he is assisting our training and development committee (Amy Hendrickson, chair and Luz Bay, member assigned to this project) in expanding this effort and also piloting a workshop during the Annual Meeting. In addition, Terry and Plumer Lovelace (Executive Director) are concluding negotiations with AERA on a contract renewal for the Annual Meeting.

Vice President and in-coming President, Linda Cook has also been very active in developing several initiatives and helping to increase NCME's organizational effectiveness. For example, this year NCME and AERA Division D will collaborate in sponsoring a mentoring lunching. In addition, Linda has revised the process for members to volunteer for committee service to ensure we have the most current list of volunteers, and proposed a new process to recognize members for their service and longevity. I am also amazed at the amount of work and responsibility assumed by our Budget and Finance Committee. This committee was only established a few years ago, and Jerry Melican is completing his term as only the second chair (following the initial chair David Frisbie). Jerry and this committee are responsible for activities that include: developing and implementing an investment policy for NCME finances, obtaining insurance coverage for NCME, developing and monitoring NCME's budget and related policies, and conducting a financial analysis and recommendations for any new NCME initiative or activity before Board action is taken. This year, NCME again decided to avoid any increase in dues or conference registration fees, but as expenses increase each year, we will need to consider such increases in the near future.

NCME Board members have two major responsibilities – to serve on the Board and vote on policy and actions and to serve as liaison to one major area. I would like to highlight the work of our two outgoing Board members this year:

Susan Loomis has headed our publications area and worked closely with Kris Walton in establishing the duties and position of website editor, selecting a new editor for JEM and EMIP, overseeing all publication issues, and establishing a new NCME Edited Book Series (more later). Kadriye Ercikan has led the membership recognition area. In just this past year, Kadriye has established a new invited session at the Annual Meeting for award recipients to describe their work which has resulted in an NCME Award. In addition, she has also been instrumental in developing our member recognition process and new ways to highlight award recipients (on the webpage and also in a new brochure that will be distributed at our breakfast). First-year Board members Bruno Zumbo and Deborah Harris are working with Susan and Kadriye in transitioning duties. Of course, Sherry Rose-Bond who oversees our professional area (outreach, recruitment, diversity, and graduate student issues) and Michael Rodriguez who has responsibilities for NCME's involvement with standards have also been actively engaged in many of the new initiatives I touch on below.

Jennifer Kobrin holds a new position as Secretary for NCME. This is a one-year term position appointed by the President with option for renewal. This position is responsible for maintaining the governance roster and responsible for ensuring all committees are appropriately staffed with overlapping terms. The secretary works closely with the incoming President to

ensure members are appointed to fill vacancies and that there is institutional history in terms of appointments. In addition, Jennifer has also attended each Board meeting and call and is responsible for developing minutes and action items that remind us all of what we promised to accomplish between meetings.

This year for the first time we have asked committee chairs to attend at least one of our two-day Board meetings. NCME has 20 standing committees, and chairs of most committees attended our June meeting in Madison, WI to present their goals and strategies for the coming year, based on the strategic plan and organizational goals that had been adopted by the Board. In October, we held another Board meeting in Washington DC which several committee chairs attended to report on major initiatives that are underway. Delegating greater responsibility for implementing operational processes, identifying important new actions or efforts, and collaborating with Board members in discussing and directing NCME is an important responsibility that committee chairs subsume. During the past year we have attempted to involve committee chairs more directly in our operations and policy development, and we have many very positive results to report. Most of the initiatives I describe below are only in their infancy this year, but I am confident that many of these efforts will develop into major new services and benefits for members of NCME. We need NCME members to get involved in our governance and we need to give them permission and authority to act. Delegating down the organizational chain is important for the future of NCME because a Board that meets three times a year simply does not have the wisdom and the time to handle the multitude of activities for an organization like NCME.

STANDARDS, POLICY, AND ASSESSMENT

- Greg Cizek is the chair of our Standards and Test Use Committee and is leading NCME's effort to submit comments on the revised draft. Steve Ferrara is our liaison to the Joint Committee and David Frisbie our liaison to the management committee. The revised Testing Standards will be released for public comment the week of January 10th (see the separate announcement in this issue). I hope you will review this draft and submit your comments and recommendations by the April deadline – it is the only draft that will be released for comments (see http://www.teststandards.net). I also will be writing to NCME members in February with details on an open forum on the draft Test Standards that will be held at the Annual Meeting. Individual members may register in advance to provide comments on this draft during the session.

- The Board approved the formation of a new standing committee – The Committee on Assessment Policy. The co-chairs are: Scott Marion (Center for Assessment) and Kristen Huff (soon to be moving over to the NY State Regents). The goal is to identify initiatives where NCME can inform policymakers of research-based findings and best practices in the use of assessments. The committee will devote the next 12 months to refining its charge and mission and will kick off activities with an invited session at NCME. The purpose of the invited session is to engage distinguished measurement experts in addressing a single important policy issue – this year the focus will be on how to establish comparability across different state consortium assessments: limitations and methods. The committee expects to then take the papers and develop a policy paper that NCME members could provide to state and federal policymakers.

- Last month, Linda Cook (President – Elect), Phoebe Winter and Jason Nicholas (co-chairs of the Outreach Committee) and I attended a meeting with a subset of leaders from CCSSO SCASS groups on accountability and large-scale assessment. CCSSO graciously funded our participation in order to develop collaboration with NCME. One proposal would be to have NCME appoint a liaison or member to each SCASS that addresses technical and scientific issues in assessment and measurement, such as the two groups we met with last week. Membership costs would be waived and CCSSO would pay for hotel and meals if NCME paid for travel. This would be one mechanism to involve more technical expertise in assessment issues at the state level. In addition, we have established a joint CCSSO-NCME invited session that will be held at both annual meetings. This year, we will have the leaders of both state-led common core assessment consortia, a state testing director from each consortium, and technical experts participate in a panel discussion at NCME and the Large Scale Assessment Conference.

NCME PUBLICATIONS

- EDITED BOOKS – The Board approved the development of an NCME edited book series. I am working with Susan Loomis and Bruno Zumbo (Board liaisons) to develop the final requirements, initiate discussions with 3-4 publishers and solicit nominations for an 'editorial board' that will work with the publisher to implement the book series. This action was actually spearheaded by Mark Gierl (Chair of Publications Committee), Jacqueline Leighton, and Greg Cizek who formed an ad hoc committee that provided recommendations to the Board. They recommended, and the Board approved, creation of an independent editorial or advisory board that would have authority to identify book titles, editors and authors (after consultation with the publisher) and subsequently manage book projects with the selected volume editors. The goal is to have royalties shared 50/50 between the editors and NCME. We hope to have a publisher selected and the first volume chosen by the Annual Meeting.

- NCME also named its first website editor – John Willse (UNCG) has agreed to oversee the webpage and our efforts to increase content and functionality. Also, for the first time, we started distributing the NCME newsletters to members via email rather than simply posting them on the webpage. I was disappointed to discover that the spring issue of our newsletter was only retrieved by 250 members when our membership is approximately 2,000. Also, please welcome Brian Clauser who is the incoming editor of JEM and be sure to send a thank you to Jim Carlson our outgoing editor for his successful term.

A NCME FOUNDATION

- Several non-profit membership organizations (e.g., APA, SIOP) have established and grown very successful foundations. Ann Fitzpatrick and Linda Cook have agreed to investigate what it might take to establish an NCME foundation. Such a foundation provides a mechanism for members and organizations to contribute (generally, tax-deductible) for various initiatives and for NCME to explore new areas that contribute to our mission. We could establish a fund to memorialize a distinguished member and later use the funds for a range of purposes that are in line with our mission (e.g., student travel grants, dissertation grants). Members may also decide to bestow some portion of their estate to NCME to extend their profession in a manner that has yet to be addressed.

ANNUAL MEETING NEWS

- NCME's program will be held at the Westin Hotel in New Orleans. We have several invited sessions that complement the conference theme of innovation in assessment and measurement. We will be listing the NCME breakfast separately from the Presidential address in the program even though these events take place in the same room at adjoining times. We hope that perhaps a few more individuals will attend the address if they believe they can sleep in a little longer and still get seating without breakfast. I hope more experienced breakfast attendees will encourage graduate students and new members to attend the breakfast and enjoy a New Orleans surprise!

- One major accomplishment is that past presidents, Board members, and committee chairs will not be scheduled for any AERA Division D sessions during our Business meeting and Presidential Address.

MEMBERSHIP

- By now you have received notice to renew your membership in NCME. When you do, you will find that we are asking a few more questions than usual in order to develop a more comprehensive membership dB. We want to know members' areas of research or applied practice, their memberships, organizational affiliations, employer and job title, and interest in volunteering for NCME roles (or serving on technical advisory or research committees). Next, we plan to update our member directory on the web so you can search for members by affiliation, state, and area(s) of expertise. We will also ask members if they are willing to allow us to share their area of expertise with federal and state government agencies when they are searching for experts.

- We are moving to a web-only membership application system and are also looking to introduce a new process for members to volunteer for NCME governance groups.

Finally, I would like to share a few personal reflections about my first seven months as President. NCME truly thrives when members become involved, and I am grateful to have had so many members both initiate and implement important new efforts for NCME. The work that has been accomplished this year has been completed by committee members and chairs – their involvement and initiative have been important to move us forward. Involvement does require time and if you are overcommitted it may be better to delay service until your professional calendar can accommodate the additional commitment. However, service is rewarding. If you have the time your efforts will be rewarded and you will be advancing the effectiveness of NCME as the prime organization in educational measurement and assessment. I am grateful to the members who have taken on such responsibility in the past and especially this year.

A NOTE FROM THE EDITOR

Thanos Patelis, The College Board

In this issue our president, Wayne Camara, The College Board, provides an update of what our organization has been undertaking. We have the last submission of Dubravka Svetina from Arizona State University in the graduate student column. We want to thank Dubravka for her effort and contributions. Thank you, Duvravka! Next, we are pleased to offer the results of our interview with Ida Lawrence, Educational Testing Service. We are also pleased to have an overview of the Common Core effort and the assessments by the consortia submitted by Pat Forgione and Nancy Doorey, Center for K-12 Assessment & Performance Management at ETS. I am also pleased to offer a new addition to the newsletter entitled, *What's New*. This is a column dedicated to highlight new publications, events, and occasions that are of interest to the NCME membership. Finally, we have various calls and announcements. As always, please drop me an email with suggestions. Sincerely and at your service, Thanos.

GRADUATE STUDENT CORNER: SUMMER INTERNSHIPS: START EARLY

Dubravka Svetina, Arizona State University

Dear Graduate Students,

I hope that your end of semester is a productive and positive one. For this column, I thought I'd share with you some information about internships. If you are a member of a listserv in our field (such as NCME, AERA Div D, SEMNET listserv or others), you likely have started to receive emails regarding applications for summer internships. Since many deadlines for internships are due soon (e.g., early February), I thought it might be a good idea to start thinking about the application process.

The goal of this column is to give you a starting point to support your interest in a summer internship – whether it's now or in the future. In my own experience interning at the CTB/McGraw-Hill a couple of summers ago, I found many beneficial experiences. The main reason that I wanted to have an internship was to experience the work that goes on in a testing company. At the end of the summer, the list of experiences that I had was quite lengthy. I learned about the company, its infrastructure, and in particular about its Research & Development (R&D) agenda. As an intern, I attended a number of professional development experiences, including presentations by researchers in our field from within and outside the company. All of these experiences were in addition to gaining skills and working closely with two mentors on a main project. The project that I undertook that summer (and continued to work on afterwards) evolved into a presentation at the annual meeting and eventually into a journal article. From talking with colleagues who have interned in the past in a number of different companies, I found that their experience was similarly positive.

If you are considering doing an internship, I hope that this column offers you some useful information. Below I summarize some of the internship opportunities that I have heard of recently. Please note that this list is not exhaustive, but only representative of what I've seen and a good start. You may also go to the NCME web site for a listing.

Several things should be noted about the application process. One is that the requirements may differ by the internship program, including the years in graduate school, degree requirements (e.g., do you need to have earned a Master's degree prior to the internship), coursework taken (e.g., do you need to have taken classes in measurement or statistics), and the like. Application documentation may also differ in terms of what an applicant needs to provide along with the application. Often, a statement of interest is required in which you discuss the reasons that you are interested in a position, the skills you bring to the table, as well as the contributions you can make to the position. Additionally, transcripts, curriculum vitae, and/or written samples may be required.

The application might also ask for letters of recommendation. If that is the case, you might consider whom you should ask to write the letters (e.g., your professors, mentors, current supervisors). It might be helpful to provide your references with your application materials that you submit. This way, your references can support your area of interest and reinforce your potential in contributing to the position and projects assigned. You should give the people preparing your recommendation letter enough time to complete the letter (a month is not a bad idea, but two weeks should be a minimum). This aspect alone makes starting early a good idea. It is good to provide your references with your CV, letter of interest, format of the letter, and due date.

Most of the internships offer stipends and might have housing arrangements for applicants living outside the area. Many run for about 8-10 weeks in the summer, although dates might differ. The following description of various opportunities is provided next (information is taken from companies' websites or advertisements). Here's a partial list of internship opportunities. Again, you may want to go to the NCME web site for additional ones.

ACT

ACT sponsors a Summer Internship Program, offered annually to outstanding doctoral students interested in careers related to assessment and educational studies. The eight-week program provides interns with practical experience through completion of a project, seminars, and direct interaction with professional staff responsible for research and development of testing programs and other educational services.

Internships in psychometrics and statistics as well as applied research are available. Areas include Psychometrics and Statistics (analysis of real or simulated data in areas such as equating, cognitive diagnostics, computer-based testing, validity, reliability, test theory, and score reporting), Education and Workforce Research (analysis of school and workplace programs, policy issues, and/or intervention strategies as they relate to college readiness, retention, remediation, and workforce training and success); Industrial-Organizational Psychology (analysis related to employers' personnel practices and their use of ACT's products and services); Career and Vocational Psychology (analysis related to career interest, values and skill assessments, evaluation of career interventions, computer-assisted career guidance, and college and career readiness).

For more information, visit http://www.act.org/humanresources/jobs/intern.html.

CTB

CTB, a division of McGraw-Hill Education, is a leading global provider of educational assessment and reference solutions guiding Educators toward targeted instruction that is making a positive impact on student achievement. The CTB summer internship program allows doctoral students an opportunity to conduct research and gain practical experience in educational measurement. Interns will collaborate with a CTB Research Scientist on a project, attend research seminars, and learn about various operational aspects of testing programs. A goal of the internship program is to facilitate interest among graduate students in pursuing a career in educational measurement.

Research Focus: Technical issues will be investigated in an area such as scaling, IRT estimation, equating, computer-based testing, or diagnostic assessment using real or simulated data under the direction of a CTB Senior Research Scientist.

For more information, visit https://mh.taleo.net/careersection/3/jobdetail.ftl?lang=en&job=10000000CS.

ETS

The goal of the ETS R&D Fellowship and Internship programs is to promote quality and distinction in educational measurement and related fields through support of significant research by early-career scientists and graduate students and exposure to methodologies within the ETS environment. These programs encourage research in areas such as educational measurement, psychometrics, validity, natural language processing and computational linguistics, cognitive psychology, learning theory, linguistics, speech recognition and processing, teaching and classroom research, statistics, international large scale assessments, and assessment design and development.

For more information, visit http://www.ets.org/research/fellowships/summer/.

HumRRO

HumRRO's internship is available to graduate students currently enrolled in full-time masters or Ph.D. accredited programs. The traditional summer internship at HumRRO is for 3 months; however, a 6-12 month internship will be considered. Our internship is a paid, full-time opportunity. An intern will serve as a research assistant/associate on one or more projects. Typical tasks include literature review, synthesis, and analysis; data collection, entry, and analysis; survey and other instrument development and administration; and documentation of findings.

For more information, visit http://www.humrro.org/corpsite/internship.

Measured Progress

Measured Progress grants several internships annually for students interested in psychometrics or educational testing studies. The program is designed to assist students in gaining better knowledge and understanding of daily psychometric activities at Measured Progress. During the intensive eight-week program, the intern works closely with our team of psychometricians.

For more information, visit http://www.measuredprogress.com/assessments/research/psychometrics/internships.aspx

Pearson

The Pearson Higher Education Summer Internship Program runs every summer beginning in late May and ending in early August. In addition to the valuable experience the student will gain from working in the corporate environment of a global publishing company, he/she will work closely with a mentor, participate with other interns in workshops geared toward career development, and make networking connections for possible employment after graduation.

For more information, visit http://www.pearsoned.com/careers/jobs.htm (keyword internship)

RAND

RAND is a nonprofit institution that helps improve policy and decision-making through research and analysis. As such, it provides a distinctive environment for graduate students. RAND's primary activity is research. The environment provides a rich variety of ongoing studies and an experienced staff of professional researchers with whom associates interact. RAND projects are typically interdisciplinary. Some research projects bring together economists, psychologists, statisticians, and health professionals, for example, whereas others bring together engineers, operations researchers, and students of organizational behavior.

RAND's Graduate Student Summer Associate Program introduces outstanding graduate students to RAND, an institution that conducts research on a wide range of national security problems and domestic and international social policy issues. Students receive a stipend and are given the opportunity to conduct research that can be completed during the three months they are at RAND.

For more information, visit http://www.rand.org/about/edu_op/fellowships/gsap/

Other companies that might be of interest for internship or fellowship opportunities include: American Council on Education, GED® Testing Service, National Conference of Bar Examiners, the College Board, Law School Admissions Council (LSAC), and AERA/NSF.

It's never too early to start thinking about doing an internship. It's always good to talk to your mentors, who themselves might have had an internship, as well as your colleagues who had interned in the past. Ask them about their own experience, and see if they can help you answer questions you might have. Also, ask them to review your application materials. As with any project, manuscript, or presentation, getting feedback from multiple people can be beneficial and may strengthen your application.

If you decide to apply for summer internships, I wish you best of luck and hope that your experience will be as positive as mine.

SPOTLIGHT ON THE PEOPLE WHO MAKE OUR ORGANIZATION GREAT – IDA LAWRENCE

For this issue, we are fortunate to receive some insights and reflections from Dr. Ida Lawrence. Below are her responses to our questions. Many thanks to her for her responsiveness and time. It's a great opportunity to get to know one of our members.

1. How did you get into the field?

My undergraduate major was in psychology and I really enjoyed the experimental classes. When it came time to decide on graduate school, I was not sure what to do. It turned out that my close friend's father was on the faculty at Teachers College, in measurement, and she told me to take a look at that field. I ended up attending NYU.

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

I have no idea. I had minored in Art History, so maybe I would have done something along those lines.

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

I would say they should try to go to an excellent program to get strong training and to make sure to take lots of courses in psychology in addition to the quantitative courses. The profession needs people who have technical skills in measurement and also have strong background in cognitive science, for example. I would also say, meet as many professionals as you can so you can figure out what you want to do (e.g., teach at a university, work for a testing company, and so on).

4. When not teaching and researching, what do you do or like doing?

I like to exercise and I like to cook. My husband and I like to go on long walks. I like to read, mainly nonfiction.

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

I would say design. I say this because I think it brings together so many facets of test design and really forces practitioners to define the construct they are intending to measure. This careful definition is sometimes overlooked and Evidence Centered Design really forces the issue.

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

I try to get to the invited addresses at conferences like AERA and NCME. I also try to go to sessions where graduate students are presenting so I can think about recruitment possibilities because I am always looking for talent to bring into ETS.

7. Who has been a significant influence in your professional life?

I would have to say that the most significant influence came from my advisor, Elazar Pedhazur. I say this because I had the luxury of being his last graduate student and he spent an enormous amount of time working with me and teaching me how to conduct and evaluate research. He also really loved the measurement field and it was kind of infectious, I think, and really spurred me to want to stay in the field. I have been with ETS for 25 years and I still love going to work. So, I consider myself very fortunate.

8. What have you most enjoyed about your career?

As I said, I have been at ETS for 25 years. Never, in my wildest imagination, did I expect to stay with one organization for this long. But I was really lucky to be exposed to a variety of opportunities, so my job has radically changed over the years. I entered ETS as a junior-level psychometrician and I worked in a technical area for over a decade. I had the opportunity to manage increasingly larger numbers of staff and I discovered that I really enjoy leadership roles. To have the opportunity to lead the Research and Development organization at ETS has been a real thrill and a real honor. I work very hard to make sure that the organization is well served for today and going forward. I consider it my responsibility to make sure we bring in and nurture the best people possible. This part of the job is very satisfying, I also love learning about the research projects and trying to help stimulate innovation. I think that my technical training helps out here, because I am able to see the larger picture and the smaller picture at the same time. All in all, I love my job because it is so varied and it really keeps me thinking.

NEW ASSESSMENTS FOR THE COMMON CORE STATE STANDARDS

Pascal D. Forgione, Jr. and Nancy A. Doorey, Center for K-12 Assessment & Performance Management at ETS

A Unique Moment in Time

2011 promises to be a dynamic and challenging year not only for the US economy, which we hope continues its recovery, but also for the field of measurement and statistics. The intellectual commodity of the future will be individuals' imagination, Thomas Friedman, author of the best-seller *The World is Flat*, told the opening session of a conference on next generation assessments last March. A shift from silos of information with limited connectedness to a global flow accessible to anyone with access to the Web, he argued, is the biggest event to change how human beings interact since Guttenberg invented the printing press in 1450. "The global

economic playing field is being leveled," Friedman was warned in an interview with a prominent Indian entrepreneur, "and you, Americans, are not ready!"[1]

And, for our own field of educational measurement and assessment, we should ask ourselves, are we ready to advance the technical science and methods that will be required to successfully implement the aggressive education reform agendas being advanced by national and state education leaders? The leader of a major testing company recently extended a bold invitation to the measurement community, "IF YOU WORK IN EDUCATIONAL ASSESSMENT, CONGRATULATIONS: You have been invited by history to help shape the future of public education, and to do so in ways that strengthen your country, improve its communities, and enhance the lives of your 300 million compatriots."[2] Kurt Landgraf of ETS described the current context as "a unique moment of history." The federal government's "Race to the Top" education reform initiative, combined with states' adoption of common academic standards, Landgraf points out, puts us on the verge of creating a new generation of comprehensive assessment systems that will replace the state-by-state mix of tests. An ETS colleague, Steve Lazer, also points out that "advances in cognitive science, task design, psychometrics, and natural language processing combined with the wide availability of technology make it possible to assess a more meaningful array of skills and knowledge than ever before."[3]

So why should we in NCME and the education measurement community care? Because education matters! Assessment policies and practices touch everyone in the education system, from students and parents to educators and government leaders. They serve as the yardstick for our performance in education and a predictor of our nation's economic and social future. Further, there are few moments in history when our society is presented with an "inflection point," or what Friedman describes as a development that will create vastly expanded opportunities to connect with the ideas of others, collaborate and create value.

A new paradigm and new era of assessment development is becoming possible because of the unique opportunity to collaborate and to connect made possible under the Race to the Top (RTTT) initiative that has brought forward unprecedented resources to stimulate bold proposals for improving the way our schools operate and the quality of tools available to our students and educators. The adoption of Common Core State Standards (CCSS) in language arts and mathematics by 43 states and the District of Columbia in 2010 has set a strong foundation for successful interstate collaboration.[4] Then, the voluntary banding together of large numbers of states into two State Comprehensive Assessment Consortia presents a unique opportunity to enhance our technical capacity to create much higher quality yet affordable assessments for the future that will measure student progress against these new CCSS.[5] Both the Partnership for Assessment of Readiness for College and Careers (PARCC) and the SMARTER Balanced Assessment Consortium (SBAC) have received more than $175 million each to design, develop and pilot test the next generation of K-12 assessments over the next four years. This opportunity will place much importance on the availability of timely research and psychometrics guidance to support these new visions of assessment and data systems that are intended to enhance quality teaching and student learning across the states. The active and critical guidance of NCME researchers, with our technical knowledge and skills, will be needed in the important conversations over the next four years that will be led by PARCC and SBAC.

These new comprehensive assessment systems are required to incorporate a number of new reform policy purposes in their designs that will challenge current assessment methods and technologies, namely:

- Go beyond the current measurement of status/proficiency and measure individual student growth, for all students across the achievement spectrum;

- Measure the extent to which each student is on track, at each grade level tested, toward college- or career-readiness by the time of high school completion;

- Provide information that is useful in informing:
    - Teaching, learning, and program improvement;
    - Determinations of school effectiveness;
    - Determinations of principal and teacher effectiveness for use in evaluations and the provision of support to teachers and principals; and

[1] "A Fresh Start: Creating the Next Generation Assessment Systems", Special Supplement published in Education Week Edition, April 28, 2010, Center for K-12 Assessment & Performance Management at ETS, "Education Urgency in a Flat World", page 14.
[2] "A Fresh Start", "An Invitation to the Future", page 16.
[3] Ibid.
[4] Two websites with information about the Common Core State Standards are: Common Core State Standards Initiative: www.corestandards.org and Education Week's Common Standards Watch: http://blogs.edweek.org/edweek/curriculum/2010/11/common-standards_watch_south_d.html.
[5] Next Generation Assessment Systems Proposed under the Race to the Top Program, Center for K-12 Assessment & Performance Management at ETS, July 2010.

- Determinations of individual student college- and career-readiness, such as determinations made for high school exit decisions, college course placement in credit-bearing classes, or college entrance.[6]

So what do the new State Assessment Consortia propose to advance the field of measurement and to meet the new emerging policy demands in the high-stakes educational accountability environment across the Nation?

State Comprehensive Assessment Consortia

The Center for K–12 Assessment & Performance Management at ETS has created a guide that describes the two assessment systems to be built by the Consortia, stimulates discussion about the key features of each assessment design and identifies a set of significant and high-leverage measurement and technical challenges within the designs of these new assessment systems.[7] In addition, a PowerPoint that provides an overview of the key features of each assessment design is available at www.k12center.org/publications.html.[8]

Below is an overview of each state comprehensive assessment system design. For each an overview will be provided followed by (a) the systems components, (b) resources, tools, and capacity building description, (c) technology, and (d) the timeline.

I. Partnership for Assessment of Readiness for College and Careers (PARCC)[9]

- MEMBERSHIP: 24 states and the District of Columbia serving more than 30 million K–12 students, approximately 62 percent of the nation's K–12 students
- PROCUREMENT STATE[10]: Florida
- GOVERNING STATES[11]: Arizona, Arkansas, the District of Columbia, Florida, Georgia, Illinois, Indiana, Louisiana, Maryland, Massachusetts, New York, Rhode Island, Tennessee
- PARTICIPATING STATES[12]: Alabama, California, Colorado, Delaware, Kentucky, Mississippi, New Jersey, North Dakota, Ohio, Oklahoma, Pennsylvania, South Carolina
- PROJECT MANAGEMENT PARTNER: Achieve
- HIGHER ED PARTNERSHIPS: More than 200 two- and four-year institutions, which typically receive 90 percent of all students across the PARCC Consortium states who enter college within two years of graduating from high school, will use the assessments as an indicator of readiness for credit-bearing entry-level courses
- AWARD: $186 million

(The previous information was accurate as of January 2011.)

The purpose of the PARCC system is to increase the rates at which students graduate from high school prepared for success in college and the workplace. To reach this goal, PARCC intends the assessments to help educators improve teacher, school and system effectiveness. This will be accomplished by providing a wider variety of data that is useful for the purposes of analyzing effectiveness, calibrating interventions, holding school professionals accountable for student outcomes, supporting strategic management of human resources, and identifying mid-year professional development and support needs for educators. This, in turn, is intended to lead to higher levels of teacher and administrator effectiveness and faster rates of student and school improvement.

(A) System Components

Summative Assessments for Accountability
The PARCC assessment system will be composed of a series of summative assessments given across and at the end of the school year, as well as aligned formative assessment resources for classroom use (see Illustration 1: PARCC).

Through-Course Assessments. These components are intended to focus instruction throughout the year on critical skills and concepts that are then assessed closer to the time of instruction, allowing for mid-year corrections. They will be given near the end of the first, second, and third quarters of the school year.[13]

---

[6] Notice Inviting Applications, U.S. Department of Education, April 9, 2010.
[7] Coming Together to Raise Achievement: New Assessments for the Common Core State Standards", Center for K-12 Assessment & Performance Management at ETS, January 2011, 17 pages. (At the printer)
[8] "*Let's Talk About It!* PowerPoint Presentation on the New Common Core State Standards Assessment Systems", Center for K-12 Assessment & Performance Management at ETS, November 2010.
[9] This summary of the PARCC assessment systems has been approved by the Consortium.
[10] Procurement states are the fiscal agents.
[11] Governing states case decision-making votes on test design and policy.
[12] Participating states consult on test design and policy, but have o decision-making authority and must participate in pilot and field testing.

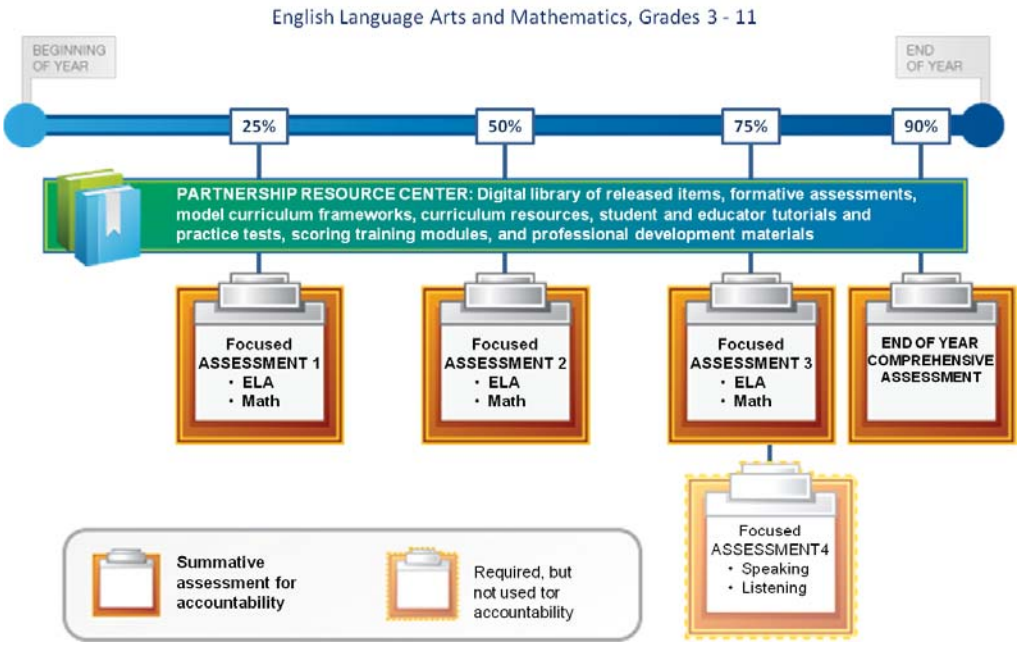

Illustration 1: PARCC

These assessments will be given primarily on computers or other digital devices and primarily scored by computer, enabling rapid return of results. Multiple types of items will be used, including computer-enhanced items and performance tasks, to measure the full range of knowledge and skills called for in the Common Core State Standards. Results are expected to be reported within two weeks of assessment.

English Language Arts

- **First quarter and Second quarter**: In a single session/class period, students will complete one or two focused literacy tasks that involve reading texts, drawing evidence from them, forming conclusions, and writing an analysis.
- **Third quarter: Part 1**: Over several sessions/class periods, students will complete a longer written task in which they will conduct electronic searches (within a predefined set of digital sources), evaluate the quality of the sources, and compose an essay or research paper using evidence from them. At each grade level, the sources will represent a range of reading/text complexity levels to enable students at higher and lower ranges of performance to demonstrate their skills.
- **Third quarter** Part **2**: (Required, but will not be used for accountability purposes) Students will present their written task (above) to their classmates and respond to questions. Teachers will score the student's speaking and listening skills using a standardized rubric, and may use the scores within the determination of student grades.

Mathematics[14]

- First quarter and Second quarter: In a single session/class period, students will complete one to three tasks that assess one or two of the essential topics identified in the CCSS for mathematics.
- Third quarter: These items will require students to apply key mathematical concepts and processes to multi-step problems. These include the ability to use technological tools to solve complex problems of the types encountered in everyday life, work, and decision-making.

End-of-Year (EOY) Comprehensive Assessment. The end-of-year assessments in English/literacy and mathematics will sample all of the standards for the grade level. These assessments will be taken online during the last few weeks of the school year, and will be entirely computer-scored. Each test will be composed of 40–65 questions per subject across a range of cognitive demand.

---

[13] Specific timing of the focused assessments will be determined after a deeper analysis of the sequencing of critical skills across the school year.

[14] In consultation with higher education faculty, PARCC is currently looking into the best approach for the high school assessments (e.g., an end-of-course model or a cumulative end-of-domain model). It is also looking into building the high school mathematics assessments as modules, such that they can be assembled into either an end-of-course (i.e., Algebra I, Algebra II, Geometry) format or integrated course (Math I, II, III) format.

- English/literacy: This assessment will focus on reading comprehension, vocabulary, and editing for grammar, usage, and language conventions.
- Mathematics: This component will assess conceptual understanding, procedural fluency, and problem solving. Technology will be used within items to enable students to, for example, create equations, graph functions, draw lines of symmetry, or create bar graphs.

Major Features

Scoring. PARCC states will adopt a common set of performance standards and scoring rubrics so results will be comparable across states. For the focused assessments, a combination of computer and distributed human scoring (either teacher or vendor) will be used. To monitor the quality and reliability of scoring, 10–20 percent of randomly selected items for grades 3 through high school will be scored a second time by humans. In addition, in high school (due to the higher stakes when used to determine college course placement), an additional 10–20 percent will be scored again. The end-of-year component will utilize 100 percent computer scoring. The Partnership plans to press for advances in automated scoring, including the use of artificial intelligence. When paper forms are used for younger students or students with disabilities, responses will be scanned for electronic or human scoring.

The Partnership will develop a technology platform to support efficient, distributed human scoring. Member states will have the option of utilizing trained teachers (who will not score their own students' work), vendor services, or a combination thereof. In all states, all teachers will have access to the online training modules for scoring so they can more deeply understand the assessments and score classroom assignments in a consistent manner. These scoring and administration plans may change as a result of the research conducted during the development phase.

Measuring Growth. In addition to the required assessment for grades of 3 through 8 and once in high school, PARCC will develop assessments for grades 9 and 10 to support measurement of annual growth. While a specific analytic approach for calculating growth has not yet been determined, the objective will be to describe each student's relative growth, expected growth given the student's prior achievement, and the extent to which that student is "on track" toward college and career readiness. The growth measure will be reported at the student, classroom, subgroup, school, district, state, and Partnership levels.

Accountability. The Partnership plans to use the results from the focused and end-of-year components in each subject to calculate annual combined scores for each student. A number of technical and psychometric challenges will be investigated during the development phase to determine if and how the scores from these multiple components can be aggregated to yield valid, reliable, and legally defensible scores. Both proficiency and growth data will be produced by the system for use, as needed, in accountability systems.

Reporting. An online Interactive Data Tool will provide teachers, parents, and administrators with access to results after each assessment, and will include various tools for displaying data, creating customized reports, and comparing the performance of similar schools. In addition, parents will be mailed printed reports after each assessment. For administrators, the system will include tools to help identify the individual professional development needs of teachers, as well as grade-level and school-level needs.

(B) Resources, Tools and Capacity Building

The Partnership Resource Center. This web-based platform is designed to be a continually expanding collection of resources for teachers, students, administrators, and parents. The resources, some of which will be available prior to 2014-15 to allow users to gain familiarity with the PARCC system, will include the following:

- Model curricular frameworks and exemplar lesson plans: PARCC will provide support to state efforts to build these resources and will provide the digital platform for sharing them across states.
- Released test items and performance tasks: Teachers will be able to use these within the flow of instruction to check student understanding. States may contribute existing state-owned items or tasks aligned to the CCSS. Within a few years, all performance tasks used in the focused summative assessments will be added, along with student performance data, scoring rubrics, and sample responses for each item. The item bank will include capabilities for sharing, improving, analyzing, comparing, ranking, and accrediting items, as well as formative and interim assessments.
- Educator training materials: Designed to help teachers understand the assessment system, implement the assessments, and interpret and use the results.
- Online practice tests for educators and students: These will allow teachers, students, and parents to become familiar with the assessments.

- An item development portal and tools: Teachers can develop their own innovative, computer-scored assessment items and share them with others via the item bank.
- Optional performance tasks for grades K–2: Teachers and schools can use these "ready-to-use" formative tasks to monitor students' performance and progress. The tasks will consist of developmentally appropriate measures such as observations, checklists, running records, and on-demand performance events and may include the use of technology innovations, such as touch screens.

The Interactive Data Tool. See Reporting, above.

Sharing State-Developed Tools. Formative and diagnostic tools being developed by member states and districts may be added, including a diagnostic reading tool (New York City), classroom reading and math diagnostic assessments (Pennsylvania), and an adaptive assessment platform (Tennessee).

Capacity Building. To help educators use the new assessment system well, the Partnership will:
- Build a leadership cadre of content experts within each state;
- Develop training tools to help educators implement the assessment system;
- Develop a sequence of online training modules for educators to learn to score, interpret, and use the assessment results; and
- Share advice on effective ways in which educators can understand and address the curricular and instructional implications of the Common Core State Standards and the Partnership's assessments.

(C) Technology

Technology is a critical component for all aspects of the PARCC assessment system, from test delivery, administration, scoring, and reporting to delivery of professional development and model lesson plans. The Partnership plans to require that all of the technology created with the support of federal RTTT resources be open source, and any pre-existing technology employed in the system be either open source or documented in a fully transparent way. PARCC received a supplementary $10 million award to support development of a highly robust and stable system and to accelerate advances in technology-enhanced items and scoring engines.

(D) Timeline

2010–2011	Development and approval by member states of common policies and procedures
2011–2012	Item and task development, piloting of components
2011–2012	Development of professional development resources and online platform
2012–2014	Field testing
2014–2015	New summative assessments in use
Summer 2015	Setting of achievement standards

II. SMARTER Balanced Assessment Consortium (SBAC)[15]

- MEMBERSHIP: 31 states serving more than 20 million K–12 students, representing about 45 percent of the nation's K–12 students
- PROCUREMENT STATE: Washington
- GOVERNING STATES: Connecticut, Hawaii, Idaho, Kansas, Maine, Michigan, Missouri, Montana, Nevada, New Hampshire, New Mexico, North Carolina, Oregon, Utah, Vermont, Washington, West Virginia, Wisconsin
- ADVISORY STATES: Alabama, Colorado, Delaware, Iowa, Kentucky, New Jersey, North Dakota, Ohio, Oklahoma, Pennsylvania, South Carolina, South Dakota, Wyoming
- PROJECT MANAGEMENT PARTNER: WestEd
- HIGHER ED PARTNERSHIPS: More than 170 two- and four-year institutions in these 31 states have committed to participate in the Consortium, help design the new assessments, and use the assessments as an indicator of readiness for credit-bearing entry-level courses in lieu of existing placement tests. These participating institutions typically receive 74 percent of all students in SBAC Consortium states who begin college within two years of graduating from high school
- AWARD: $176 million

(The previous information was accurate as of January 2011.)

[15] This summary is the work of the K – 12 Center and has not been formally approved by the SBAC.

The design of the SMARTER Balanced Assessment Consortium is intended to strategically "balance" summative, interim, and formative assessment through an integrated system of standards, curriculum, assessment, instruction, and teacher development, while providing accurate year-to-year indicators of students' progress toward college- and career-readiness.

The assessments and curricular materials will rely on research-based learning progressions, which further define how students acquire the knowledge and skills called for in the standards. Summative assessments will include both extended performance tasks and a computer-adaptive end-of-year assessment to assess the full range of the Common Core State Standards. In addition, SBAC will provide a suite of optional interim and formative tools and resources. These include: computer-adaptive interim assessments using non-secure tasks and items of types similar to those used in the summative assessments that provide teachers with instructionally useful information about each student's progress during the year; formative tools and strategies for more regular classroom use; and professional development resources in the formative assessment process and use of assessment data of all types to adjust and improve instruction.

Technology will be leveraged in this design in several ways: adaptive testing will be used to enhance the precision of scores across the full achievement spectrum; technology-enhanced test items will expand the range of skills that can be assessed; online professional development resources and research-supported instructional tools will support improved instruction and school leadership; and, through use of an interoperable electronic platform, the Consortium will support both standardized and customized reports that can be targeted to a range of audiences for tracking and analyzing progress.

A guiding principle for the SBAC Consortium is "responsible flexibility." SBAC will make it possible for states to customize system components while also ensuring comparability of student scores across all participating states on the summative assessments.

(A) System Components

Summative Assessments for Accountability

Assessments will be developed for English language arts and mathematics for grades 3 through 8 and 11, with assessments for grades 9 and 10 available for states that choose to use them (see Illustration 2: SBAC).

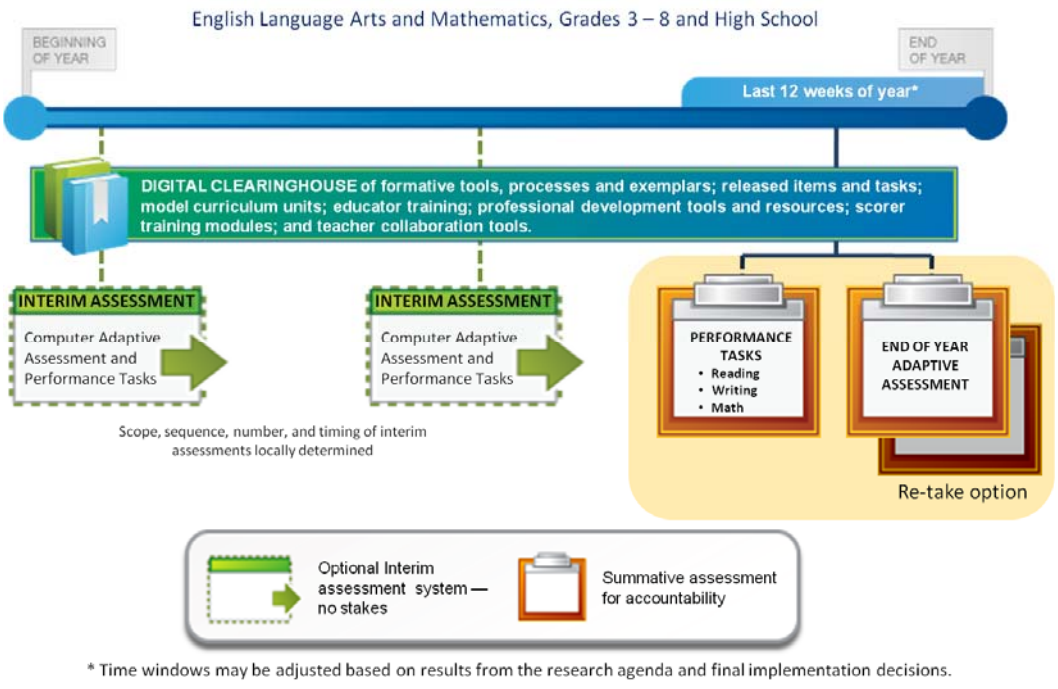

Illustration 2: SBAC

Although all assessments are to eventually be delivered via computer, the Consortium expects to offer a paper-and-pencil option for three years to support states that do not yet have sufficient technology infrastructure to make a complete transition at the outset.

Taken during the final 12 weeks of the school year[16], the summative assessments for each grade and subject will have two major components: performance tasks and a comprehensive end-of-year computer adaptive assessment, as described below. All of these assessments will provide students with information regarding their achievement, growth, and progress toward college- and career-readiness by the end of high school.

Performance Tasks. Students will complete one task in reading, one in writing, and two in mathematics annually, during a Consortium-defined testing window. Each task will be delivered via computer and will require one to two class periods to complete. These tasks will evaluate aspects of the Common Core State Standards that are difficult or not possible to assess through more traditional items. They will involve student-initiated planning, management of information and ideas, interaction with other materials and/or people, and production of an extended response such as an oral presentation, exhibit, product development, or an extended written piece. A combination of machine and teacher/human scoring will be used depending on the task-specific availability of adequate technology for valid and reliable scoring.

End of Year (EOY) Comprehensive Adaptive Assessment. The EOY assessment will consist of approximately 40–65 questions per content area presented within a computer-adaptive assessment. It will include selected-response, constructed-response, and technology-enhanced items.

- The EOY component includes a retake option. Students who are approved to do so may take the assessment a second time, but will see a new set of items. The student's highest score would be used to determine annual achievement and annual growth relative to staying on track to college- and career-readiness.

The Consortium will also conduct studies to determine whether distributed summative assessments (a series of tests taken across the school year) are sufficiently valid, reliable, and comparable to the above EOY assessments to be offered as an alternative to the current EOY assessment. The distributed model would likely be developed based on content clusters and would allow students to demonstrate mastery in specific units of content throughout a grade level or course.

Major Features

Scoring. Performance tasks will have some components that are scored by computer and others that require human scoring. A Consortium priority is the strategic involvement of teachers in the development of items and scoring guides and in the scoring of constructed-response items (10 percent teacher scored) and performance tasks (33 percent teacher scored), although no teacher would score his/her student's responses. An online system will be developed to allow efficient distributed human scoring and monitoring of the accuracy of each reader.

For the EOY assessment, selected-response and technology-enhanced items will be computer- scored and preliminary results from this set of items will be reported immediately. Additional items that can be reliably scored using artificial intelligence engines will be electronically scored, with 10 percent back-read by humans to verify the accuracy of the engine. Final combined scores for these EOY assessments are expected to be delivered within two weeks. The Consortium plans to leverage advances in both electronic item types and electronic scoring to support its design, and will invest in the development of a training system for human scorers.

Measuring Growth. The Consortium intends to build vertical scales across the grade 3 through 11 span in English language arts and mathematics, which can then be used as the basis for growth measures evaluating the individual's progress toward college- and career-readiness across the years, but will also build horizontal (grade specific) scales. Both the summative assessment results and the interim assessment results will be reportable on these vertical and horizontal scales. The Consortium proposes to allow states flexibility in selecting the specific growth model that is appropriate for their state. The Consortium will conduct studies of the characteristics of different models when used in conjunction with the data from the summative assessments to inform subsequent decisions.

Accountability. Student scores from both the performance tasks (one in reading, one in writing, and two in math per year) and the EOY comprehensive assessment will be combined for the annual summative score. Research will be conducted to inform decisions concerning the aggregation and weighting of the results from these two components.

Reporting. A common electronic platform will be developed to manage assessment data and provide sophisticated data reporting and analysis tools for customized reports. Students, teachers, parents, and administrators will be given security settings to access appropriate data only. Student scores on the performance tasks will be reported separately, as well as in combination with the EOY comprehensive assessment. Student performance levels will be explained with examples to aid interpretation. Reports will provide item-level information for clusters of items, provided that this is found to yield valid and reliable information. In addition to summative results, scores from the interim assessments throughout the school year will be available in the same reporting suite and report more detailed information concerning progress toward that grade level's standards. This system also will include links to model curriculum and instruction resources and assessment professional

[16] Time windows may be adjusted based on results from the research agenda and final implementation decisions.

development resources. The reporting tool will be customizable, allowing each state to "brand" the reporting in a manner consistent with other state-level reports.

(B) Resources, Tools, and Capacity Building

Optional Interim Assessments. These optional computer adaptive assessments can be self-administered several times each year (to be determined by states/locals) and will provide near-immediate results on the same scale as the summative assessment. The item types will mirror those on the summative comprehensive assessment.

Two modes of test administration will be available, both of which can be given multiple times per year at the discretion of the state, district, or school. One version mirrors the length and scope of the EOY summative assessment and yields a scale score that can be used as a growth or achievement metric. A shorter "cluster assessment" version also will be available that assesses, at a deeper level, a smaller set of standards based on defined learning progressions, thereby providing more detailed diagnostic feedback. The items will be stored in a non-secure item bank and can be grouped into customized clusters based on state or local curricula, and can be administered before, during, or near the end of instruction. Reports of student results will link teachers to related formative tools and strategies for their students and professional development resources targeted at individual student areas of weakness.

Comprehensive Electronic Platform. The SMARTER Balanced Assessment System will be built around a comprehensive electronic platform that contains an expanding collection of resources for teachers, administrators, students, and parents, including:

- A System Portal ─ This portal will serve as the single point of entry for educators, students, parents, and policymakers to all components of the system. In addition to the features described below, the portal will provide access to the assessment delivery platform, the distributed hand-scoring platform, virtual classrooms, and issue-focused chat rooms.
- The Educator Dashboard ─ A secure online portal will allow educators to:
    - download, view, and analyze assessment reports, scoring rubrics, and longitudinal data;
    - generate custom reports;
    - access model curricula that are aligned to the Common Core State Standards;
    - access research-based instructional strategies and interventions related to specific assessment results for individuals or subgroups;
    - access vetted instructional units, formative tools, and sample performance tasks; and
    - network with teachers to share information and resources and discuss curriculum, instruction, and assessment.
- Formative Tools, Processes, and Practices Clearinghouse ─ To be developed for grades 3–8 and high school, this bank of resources will include:
    - formative assessment tools and strategies, including the use of performance tasks to solicit formative information, and rubrics that can be used by teachers on-demand to support teaching and learning;
    - assessments created by teachers using these tools and instruments which can be administered as computer adaptive assessments, teacher-administered performance tasks, or classroom exercises; and
    - research-based instructional tools and processes.
- Item Development/Scoring Application ─ Online training modules will be available for both development of assessment items and tasks and for scoring of items and tasks. For those educators who successfully complete the training, item authoring and scoring software will become accessible.
- Reporting Suite ─ See "Reporting" above.
- Feedback/Evaluation Tools ─ These tools will support regular surveying of system users and vetting of submitted materials.

(C) Technology

The Consortium plans to develop an open-source technology platform for this assessment and learning system using a combination of existing and newly developed open-source software and proprietary software. Upon completion of the system development, a public license defining this as free, open-source software will be created.

(D) Timeline

2011 Development of formative tools, processes and practices underway
Specifications for summative and interim assessments developed
2012 Summative and interim item development completed
Interim item pool becomes available for use
2013 Field testing of items for adaptive summative assessment completed

2014 Preliminary achievement standards proposed and other policy definitions adopted
2015 Operational summative assessment administered
Final achievement standards verified and adopted

Finding Solutions, Moving Forward

The assessment system designs developed by the two Consortia have some attributes in common but also differ sufficiently to allow for "dueling banjos" during the design, development, and piloting phases ─ a period of four years during which we can learn a great deal that will inform these two systems and the future of the field of assessment more broadly.

PARCC places emphasis on the use of focused, computer-delivered tasks (also called through-course assessments) that assess key skills and concepts at key points across the school year, closer to the time of instruction. SBAC places a priority on the use of adaptive interim and summative assessments to gain greater score precision and student engagement. Both will stimulate the development of a broader range of computer-delivered and scored item types and more informative reports.

Embedded in these designs, however, are a number of important and unresolved psychometric challenges. How far can we push the frontiers of measurement during this development phase? Can we find better solutions to address two priorities that stand in tension: first, the need for highly precise and reliable data for high-stakes decisions; and second, the need for assessments that require students to apply knowledge and skills to solve complex, real-world problems? Moreover, can we use this federally funded initiative to develop an information infrastructure that will allow states, districts, and schools to significantly accelerate the rate at which we can access and analyze data to propel continuous system improvement?

Over the next several years, The Center for K–12 Assessment & Performance Management at ETS (the Center) will be bringing together some of the best minds in the country to debate and identify promising approaches to the challenges regarding assessment and performance management systems and will be sharing the products from these activities publicly in order to stimulate and contribute to broad advances in the field. We will begin in 2011 with two challenges that are central to the consortia designs: a) the development and use of through-course summative assessments and b) the use of technology-enhanced tasks and scoring engines.

Through-Course Summative Assessments

A through-course design is "one in which an assessment system component or set of components is administered periodically during the academic year" (Race to the Top Notice of Invitation to Apply, USED, 2010). This format allows for the assessment of skills and concepts that were recently taught, enabling mid-course corrections, and may include items designed to inform the next unit of instruction. The PARCC Consortium has chosen to include three focused assessments in mathematics and in English language arts each school year. These will likely be given near the end of the first, second, and third quarters, will consist of a small number of tasks, and will be augmented by a comprehensive end-of-year assessment. The SBAC plans to develop the model shown on pages 9–11 of this guide, which does not contain through-course assessments. However, they plan to investigate an alternative model in which the end-of-year assessment is replaced by a series of computer-adaptive assessments, each of which assesses a cluster of standards, and are aggregated with the annual performance tasks.

This through-course assessment format raises a number of measurement challenges. First and foremost, the determination of the subset of skills and concepts to be emphasized in these through-course assessments requires that we identify within each content area and grade level the "keystone" topics or cognitive targets for which deep mastery is necessary for ─ and highly predictive of ─ readiness for college and career. What are those skills and concepts, at what level of mastery, and in what sequence, if one such sequence exists? Studies will need to be carried out to gain deeper understanding than we currently have to support these decisions.

Designing these components such that they can be placed onto a common scale and equated from year to year may require new approaches. Policy decisions concerning the weighting of the individual components into a composite annual score will need to be informed by data from field tests to ensure that the final composite scores are legally defensible for use in high-stakes decisions concerning individuals.

The goal of assessing clusters of highly important competencies throughout the year is an important and worthwhile one. Finding solutions to the measurement challenges would help our schools identify individual student needs for intervention or acceleration throughout the year. The Race to the Top Assessment Program, therefore, will likely stimulate important advances in the measurement field.

Technology-Enhanced Tasks and Scoring Engines

The demand for the transition from paper-and-pencil to online testing has grown as parents and teachers demand faster return of results and states seek to reduce the costs ─ in dollars and personnel time ─ associated with printing, shipping, and securely handling millions of printed test booklets.

Computer-based assessment can also bring benefits to the quality of the assessment itself. It creates the opportunity for an expanded set of built-in accommodations for students with disabilities and English language learners, such as text-to-speech utilities, text enlargement, and language translations. Many skills and competencies that are fundamental to college and career readiness, and are called for within the Common Core State Standards, cannot be assessed on paper, such as online searching (in contained digital libraries) and the use of word processing and data analysis software. New interactive computer tasks allow us to gain information about both the content knowledge of the student and the processes used by the student in solving complex problems. Here, the more challenging electronic and online games can provide some insight into what may be possible to incorporate into simulations that require the application of knowledge and skills to real-world problems.

There are clearly new frontiers to be reached in the area of automated scoring, as well. A stated goal of the PARCC consortium is to drive innovation in this area such that all items and tasks can be scored very quickly by computer. Artificial intelligence engines exist that score the large majority of student essays, at least as reliably as humans, and "send back" those essays that are so unique or creative as to require human scoring. However, as we look to assess writing in the context of science, English literature, or history, as called for in the Common Core State Standards, new advances are needed to produce reliable sub-scores for both writing and the content area constructs assessed.

The development of interactive computer tasks and their automated scoring engines is challenging, particularly when used within high-stakes assessments. To illustrate, the National Board of Medical Examiners has been working on interactive medical case simulations and automated scoring methodologies for use in their licensure examination for nearly 30 years, and using them in operational exams for more than a decade. Examinees currently complete nine simulations, each of which takes about 25 minutes. However, equating from year to year is conducted based on the roughly 480 selected response items in each assessment. If results from the new K–12 assessment systems are to be used for comparable high-stakes decisions regarding individuals, such as the awarding of high school diplomas, similarly high thresholds for psychometric quality must be met. This is clearly a time for aggressive research and development if we are to realize and use the benefits of such new assessment items.

Looking Beyond 2015

Some of the challenges discussed above may be solved within the next two or three years, in time for application within the initial roll-out of PARCC and/or SBAC. Others will require more time to develop prototypes, pilot, and prepare for use within high-stakes assessments.. New technologies will be developed, which will impact the technical and financial feasibility of some approaches, and open the doors to new approaches.

This, then, points out the need to think of the next four years as only the beginning ─ the development of a strong foundation ─ for a new, robust data and research platform from which we can leverage technologies to accelerate advances in K–12 education and to enhance student learning for all children. The Common Core State Standards and aligned common assessments create the opportunity to shift from norm-based educational decision-making ─ identification of what works for most students, most of the time ─ to more nuanced and personalized educational decision-making. The data platforms of 2015 should allow us to answer questions such as: What would work best for this student based on accumulated information from large numbers of students with similar prior achievement patterns? What are the most effective professional development activities for the unique needs of this particular teacher as identified by the progress of his/her students?

To realize this vision, we will need to standardize data formats such that anonymized student, teacher, and assessment data can be collected in real time and organized to support timely analysis and research. Educational researchers and measurement experts have a critical role to play now, during the development phase of the Consortia work. If, five years from now, we are to be able to tap the wisdom of the largest number of experts in identifying "what works" to improve learning in more and more nuanced and timely ways, then the new assessment and data systems will need to be designed with this purpose in mind.

This is truly a unique moment in the history of American public education and for the field of assessment. And, as earlier advocated by Thomas Friedman and Kurt Landgraf in the opening section of this article, we have an important opportunity to advance the field of educational measurement and to positively impact educational standards and aligned assessments in the new digital world of the 21st Century. We, as a community of educational researchers and measurement professionals, must utilize this opportunity to make significant advances in the quality, timeliness and usefulness of our assessments and to create the information infrastructure that will enable us to accelerate our learning. In doing so, we will help our country regain its former position as first in the world in educational attainment and equity.

The Center will be working with nationally recognized measurement experts from across the country to explore possible solutions and will share the resulting ideas and recommendations through webinars and our website. To sign up for notices as these resources are made available, go to www.k12center.org.

WHAT'S NEW?

This is a new column to the NCME Newsletter. It is an opportunity for the membership to share new things of interest with each other. Please send any new items that are of interest to our members (e.g., new or forthcoming books, members who have changed jobs, or new applications) by email to the editor (tpatelis@collegeboard.org). I hope you find this informative.

New/Recent Books:

- ***ASSESSMENT LAW IN EDUCATION*** **by S.E. Phillips**

The impetus for the book originated with an invitation from the Iowa Testing Program and the College of Education at The University of Iowa in the fall of 2006 to teach a seminar on Legal Issues in Educational Testing for graduate students in the Educational Measurement and Statistics Program. Teaching the seminar provided an opportunity to revise and reorganize materials the author had used previously for similar courses at Michigan State University and Thomas M. Cooley Law School. The author is grateful to Dr. Michael Kolen, Dr. Robert Forsyth, Dr. Tim Ansley and Dean Sandra Damico at The University of Iowa for their support in arranging the visit to the University and making the book possible.

The invitation also included an offer of a graduate assistant, Michelle Croft, who was ABD in measurement and statistics and a second-year law student. As a member of law review, Michelle was experienced in legal research and was interested in pursuing a dissertation topic combining psychometrics and law. Michelle attended the seminar class and asked the author to serve as her Dissertation Co-Director. Michelle provided valuable legal research, cite checking, legal case materials, multiple-choice and essay review items, glossary entries and other organizers for the book. Several professional colleagues also reviewed drafts of chapters for the book and provided helpful comments and editorial suggestions. These professionals represented a variety of expertise, employment and experience in the areas of psychometrics and law.

In the Fall of 2010, The University of Iowa invited the author back to teach the *Legal Issues in Educational Testing* graduate seminar class again. The author was able to field test a draft version of the book while teaching the course and appreciated the many helpful editorial suggestions and comments provided by the seminar students. Both the professor and the students benefited from having the course content collected in a single volume.

The book (published in January, 2011 by www.SEPHILLIPS.DOKSHOP.COM) is different from traditional textbooks/reference books because the author was professionally involved in several of the legal cases discussed in detail in many of the chapters. This firsthand knowledge allowed the author to present in-depth discussions of issues, data and evidence not usually available in the reported court decisions. However, because legal cases are adversarial and expert witnesses represent only one of the parties (plaintiff or defendant), such information also provides a more comprehensive treatment of one side of the case relative to the other side of the case.

Nonetheless, because the author believed such information would be useful to testing programs in (1) understanding the nature of potential legal challenges and (2) developing the necessary documentation and data for defending such challenges, the author included this information, where appropriate, with a notation indicating that the material was adapted from the author's expert declaration or expert witness report filed in the case. The author hopes that the educational value of this material outweighs any limitations attributable to a lack of neutrality.

The cases discussed in the book for which the author served as a consultant and/or expert witness, listed chronologically with the party represented in bold, include the following: *Pandazides v. Va. Bd. of Educ.* (1991), *Golden v. Ala. Dep't of Educ.* (1994), *Fairfield Sch. Dist.* (1996), *Rene v. Reed* (1999), *Chicago Sch. Reform Bd. v. Substance, Inc.* (1999), *GI Forum v. Tex. Educ. Agency* (2000), *Cal. Dep't of Educ. v. San Francisco Unified Sch. Dist.* (2000), *Williams v. California* (2003), *Doe v. Nat'l Bd. of Medical Exam'rs* (2006), *Kidd (Chapman) v. Cal. Dep't of Educ.* (2006), *Valenzuela v. O'Connell* (2007) and *Coachella Valley v. California* (2007). Many of the valuable lessons learned during this work are recounted in the book together with discussion of many other important testing cases, related federal laws and professional psychometric standards.

- ***HIGH STAKES TESTING IN EDUCATION: SCIENCE AND PRACTICE IN K-12 SETTINGS*** **edited by James A. Bovaird, Kurt F. Geisinger & Chad W. Buckendahl.**

This edited book is forthcoming in the spring of 2011 published by the American Psychological Association. The book offers a compilation of chapters written by many NCME members offering a variety of perspectives and information about testing in a several contexts and applications within K-12 environments.

- ***INTRODUCTION TO PSYCHOMETRIC THEORY*** **by Tenko Raykov and George A. Marcoulides.**

This recently published text (September 2010 by Routledge Academic) provides an introduction to educational and psychological testing and measurement theory that reflects many intellectual developments of the past two decades. The book introduces psychometric theory using a latent variable modeling (LVM) framework and emphasizes interval estimation throughout, so as to better prepare readers for studying more advanced topics later in their careers. Featuring numerous examples, it presents an applied approach to conducting testing and measurement in the behavioral, social, and educational sciences. The authors rely on LVM when discussing fundamental concepts such as exploratory and confirmatory factor analysis, test theory, generalizability theory, reliability and validity, interval estimation, nonlinear factor analysis, generalized linear modeling, and item response theory.

ANNOUNCEMENTS: NCME

DRAFT TESTING STANDARDS AVAILABLE FOR COMMENT JANUARY 2011

Following the completion of the 1999 Standards for Educational and Psychological Testing, AERA, APA, and NCME established a Management Committee overseeing the publication, the budget, and future revisions. Members of the management committee are: David Frisbie (NCME), Suzanne Lane (AERA) and Wayne Camara (Chair and APA). The Joint Committee for the Revision of the Test Standards, which has been charged with completing the revision, is preparing to release a draft in January, 2011. We wanted to share the announcement below with NCME members, as well as individuals and organizations involved with measurement and assessment issues. We also encourage you to share this announcement with other organizations and groups whom you believe would be interested in reviewing and commenting on the draft revision of the Test Standards. We hope individuals and organizations interested in assessment and measurement will review the draft and submit comments according to the procedures described in the announcement. The January 2011 draft will be the only draft open to the public for review and comment.

Wayne Camara, Suzanne Lane, and David Frisbie

COMMENTS ON REVISIONS TO STANDARDS FOR EDUCATIONAL AND PSYCHOLOGICAL TESTING

The Joint Committee charged with revising the Test Standards is anticipating release of a draft revision for comments during the week of January 10, 2011. The draft revision of the Standards for Educational and Psychological Testing (AERA, APA, NCME) will be posted at http://www.teststandards.net under the tab labeled 'REVISION' and then "Comment on the Revision". An on-line template will allow individuals and organizations to review the draft revision and to submit their recommendations and rationale for each recommendation. All comments and recommendations must be submitted to the Joint Committee through this mechanism and must be completed by approximately April 20, 2011. Please share this announcement with organizations and groups interested in professional and scientific issues concerning testing and assessment.

OTHER ANNOUNCEMENTS

The Seminar and Training Center of the Department of Measurement, Statistics and Evaluation at The University of Maryland is pleased to announce the very popular short course:

INTRODUCTION TO MULTILEVEL ANALYSIS METHODS:HIERARCHICAL LINEAR MODELING
March 16-18, 2011 (Wednesday, Thursday, Friday)
Instructor: Dr. Robert Croninger, University of Maryland, College Park (croninge@umd.edu)
For more information, please visit http://www.education.umd.edu/EDMS/ShortCourses/HLMworkshoppage.html

CASMA SUMMER EQUATING WORKSHOP
June 27 - July 1, 2011 --- Iowa City, IA

A more intensive and extensive workshop than a one-day session on equating will be offered by CASMA on June 27 - July 1, 2011. Two similar workshops were offered previously. They were well received and well-attended. For the 2011 workshop, attendance will be limited to 18 persons. We realize that these difficult economic times may impede attendance. If registration is too low to justify holding the workshop, fees will be returned. No workshop can replace a full-length course, but this particular workshop should provide participants with a good working knowledge of basic equating designs, statistical procedures, and applications.

Visit the workshop web page www.education.uiowa.edu/casma/equating_workshop.htm for details.
Should you have any procedural, housing, or registration questions, contact Jennifer Jones at 319-335-5439. Workshop content questions can be directed to Bob Brennan, University of Iowa, at 319-335-5405.

CALL for NOMINATIONS for NCME Newsletter Editor

The NCME Publication Committee is requesting nominations or self-nominations for the position of NCME Newsletter editor. Undertaking this kind of position allows one the opportunity to make an important contribution to our organization. The newsletter is a quarterly publication containing announcements and news about current events related to research and practice in educational measurement. The newsletter editor selects an advisory board and, together, they solicit ideas for the newsletter. The current editor of the NCME Newsletter is Thanos Patelis. His 3-year term will come to an end in December 2011.

If you are interested, please send your CV by Friday March 11, 2011to Mark Gierl, University of Alberta, (mark.gierl@ualberta.ca). After March 11, a list of candidates will be selected and presented to the President, who will rank order the top three candidates and present these candidates to the Board. The Board will then vote on each of the candidates thereby reaching a decision. The new editor will be appointed at the 2011 annual meeting in New Orleans, and begin his or her term on January 1, 2012.

Professional Development Opportunity

AERA is offering a number of unique professional development courses this year including one that may be of special interest to NCME members. *PDC13: Making an Impact with Your Research through Effective Presentations, Social Media, and Writing* will focus on improving your communications skills with in-class practice. **Saturday, April 9, 1- 5 pm,** fee: $95. Instructors are Ron Dietel, CRESST/UCLA; Barbara McKenna, Stanford University; and Paul Baker, Wisconsin Center for Education Research. More information: http://communicateresearch.wordpress.com/
Registration - http://communicateresearch.wordpress.com/register/
Syllabus - http://communicateresearch.files.wordpress.com/2011/01/communications-101.docx
Blog - http://communicateresearch.wordpress.com/ ; Ning - http://aera2010.ning.com/

Thanos Patelis
The College Board
45 Columbus Avenue
New York, NY 10023

Phone: 212.649.8435
Fax: 212.649.8427
e-mail: tpatelis@collegeboard.org