

FROM THE PRESIDENT

EDUCATIONAL MEASUREMENT REMAINS A PILLAR IN FEDERAL AND STATE EDUCATION POLICY

Richard J. Patz, ACT

Congress has passed and President Obama has signed the Every Student Succeeds Act (ESSA), replacing No Child Left Behind (NCLB), the 2001 legislation that has guided the federal role in K-12 education for nearly a decade and a half.

What's the difference between ESSA and NCLB? Logically, if each child is a student who either succeeds or is left behind, then no child is left behind if and only if every child succeeds. This parsing of monikers suggests a great deal of consistency between the newest reauthorization of the 1965 Elementary and Secondary Education Act and its predecessor.

Indeed, ESSA and NCLB do have a great deal in common, specifically with respect to educational testing. Nonetheless, significant changes that address new political realities and important public concerns have been introduced, and these changes will have a pronounced impact on educational measurement across the United States.

The new law, like its predecessor, governs the way federal funds are used to support K-12 education at state and local levels. These funds continue to flow to states annually, and states allocate funds to local education agencies, all of this by formulas that take into account factors such as poverty and education costs. The law continues to impose requirements for student testing and school accountability, setting the boundaries within which states operate their unique programs. Each state's program is to be written up in an implementation plan, submitted to the U.S. Department of Education, subjected to a peer-review process, and executed beginning in the 2017-18 school year.

ESSA continues many of NCLB's testing and accountability provisions. States must adopt challenging academic standards in reading or language arts, mathematics and science, and assess students annually in these subjects using tests that are aligned to their standards. Results are to be reported for whole populations and for key subgroups including English learners, students in special education, and those based on race, ethnicity, and family income. Progress is to be tracked and reported annually, with interventions required for low performing schools.

Despite their broad similarities, ESSA leaves some of NCLB's most controversial and problematic provisions behind. Stated most generally, ESSA provides more flexibility to states, limits the federal role in education policy, and includes provisions to encourage innovation in assessment policy and practice.

- States must establish targets for improving outcomes and closing achievement and graduation gaps, but they do this by establishing their own achievement goals.



- State academic standards are not submitted to the Secretary of Education for approval, and the Secretary is expressly prohibited from encouraging states to pick a particular set of standards.
- The requirement that states use student achievement outcomes in teacher evaluation has been dropped and federal mandates on teacher evaluation have been prohibited more generally.
- Computer adaptive assessments are allowed and specifically encouraged for assessing students on content above grade level.
- The use of nationally recognized assessments in high schools is allowed.
- States are given some flexibility to implement their own opt-out policies, and concerns about too much testing are addressed with provisions for auditing assessment systems, eliminating unnecessary tests, and reporting publicly the time spent on testing in schools.
- Innovation is encouraged through tryouts in as many as seven states of new assessments utilizing new ways of demonstrating mastery.

In these and other ways, ESSA has responded to the challenges evident in NCLB, some lessons learned through the recent era of Common Core assessments, and shifting public sentiment about the role of assessment in public education.

The reauthorization of NCLB was long overdue. In navigating our way forward it will be important to learn from both the successes and setbacks of the recent and distant past.

Before NCLB, state assessments were much more diverse. States focused their assessments on different grade levels, used different testing approaches including performance, portfolio and multiple choice formats, and leveraged both custom and publisher-created assessment elements at the state and district levels. The variety was greater, but the technical quality was uneven, a narrow conception of standardization limited accessibility, and the accountability systems used demographics in ways that were interpreted as lowering expectations for communities with the greatest needs.

During the NCLB era testing expanded dramatically and there was little oxygen (i.e., money) to support assessment approaches that did not scale up to very large volumes. Implementations across the states became less diverse, and accountability provisions became more controversial and problematic. Technical quality remained uneven and technical innovation was constrained.

As NCLB's accountability provisions proved increasingly unworkable, states have been operating under waivers and working together in new ways. Common Core State Standards were widely adopted, and common assessments have been built using a mix of new models for state collaboration and traditional models of publisher investment and state adoption.

There has been a healthy recognition of the economic imperative for higher academic standards, and an understanding that states could leverage high quality assessments built beyond their borders and outside of the direct control of their state agencies. There have been technical, policy, economic, and political challenges in this environment as well, a number of which are playing out in evolving state policies and also reflected in provisions of the new ESSA law.

The educational measurement profession has a great deal to offer in this dynamic policy landscape. We bring an understanding of the standards and best practices that are the foundation of sound educational assessments, the technical expertise necessary to ensure accurate and meaningful achievement data supports policy objectives, and more than a little practical wisdom about how assessment programs can be structured and implemented successfully.

Now is an opportune time for members of NCME to be engaging with colleagues working in K-12 education, offering assistance as peer reviewers, technical advisors, and/or as interested and engaged scholars. We have a wealth of

relevant research to share in addition to our expertise and experience, all of which is needed if we are to allow empirical evidence and solid science to guide us on the long path to an educational system in which every student succeeds.

Learn more about Every Student Succeeds Act: <http://www.ed.gov/essa>.

FROM THE EDITOR

Heather Buzick, Educational Testing Service



Happy New Year! In this issue, NCME President Rich Patz briefs us on the federal government’s reauthorization of the Elementary and Secondary Education Act. We remember NCME past president, Robert L. Linn (1981–1982), who passed away in early December, with tributes from Barbara Plake and Gerald Melican. Next, Jonathan Rollins adds to previously written perspectives of graduate students on the job search. Our spotlight member is Brian Gong from the Center for Assessment. For our special topic, Susan Davis-Becker and Andrew Wiley, both from Alpine Testing Solutions, write about communicating the positive aspects of assessment. Rose McCallin, chair of the NCME publications committee, has authored an explainer on bibliometrics for NCME’s publications. We have updates from the career contribution committee and the fund development committee, as well as announcements about the NCME Fun Run/Walk and NCME Sunrise Yoga. The final announcement pertains to the in-progress NCME elections. Polls close on January 15, 2016.

As his 1-year term comes to an end, I would like to thank Jonathan Rollins for serving as our Graduate Student Corner columnist.

I encourage you to send me articles, suggestions for content, and feedback on this or previous issues.

From the President Educational Measurement Remains A Pillar in Federal and State Education Policy	1
From the Editor.....	3
Remembering Robert Lee Linn	4
Tribute to Robert L. Linn, Past President of NCME	4
Graduate Student Corner: Considerations for the Job Search.....	5
Spotlight on the People Who Make Our Organization Great: Brian Gong.....	8
Explaining the Positive Side of Assessment: The Struggle Is Real	9
Publications Committee Update	11
Career Contribution Committee Update	17
Fund Development Committee Update	17
NCME Fitness Walk/Run Corner	19
NCME Sunrise Yoga	19
NCME Elections.....	20

REMEMBERING ROBERT LEE LINN

Barbara Plake, Professor Emeritus, University of Nebraska-Lincoln

December 8, 2015, was a very sad day for psychometrics, as that was the day Robert Linn passed away following a long illness. I had the great good fortune of knowing Bob for almost 40 years, starting in 1979 with my involvement with the National Council on Measurement in Education. I remember Bob back then as an active board member in CME and then as president in 1981. I even remember his presidential address (but can hardly remember any of the others). I remember his address because it was on selection and test bias and tackled head-on, in clear language and with data representations, important issues of race and ethnic bias in testing.

Tackling difficult and important topics in policy and research was a hallmark of Bob's work. He didn't shy away from controversial topics or hard work, including serving as vice-chairperson for the 1985 Standards for Educational and Psychological Testing and serving as editor of the third edition of *Educational Measurement*. He had a major impact on the field of educational measurement.

But even more, Bob was a wonderful human being. He was kind and thoughtful. He gave freely of his time and energy. He was active in many technical advisory committees. Later in his life, as his health became challenging, he continued his devotion to the measurement community with the support of his devoted wife Joyce, who accompanied him on his travel to professional meetings. He will be greatly missed.

TRIBUTE TO ROBERT L. LINN, PAST PRESIDENT OF NCME

Gerald Melican, Consultant and Newsletter Advisory Board Member

As the news of Dr. Robert Linn's passing spread across the profession the number of short testimonies began to aggregate. Dr. Linn's impact on education, psychology, psychometrics, and research were noted as might be discerned by his prodigious work. In every case that I saw, however, the writer mentioned Dr. Linn's personality traits; his work ethic, generosity, gentlemanliness, and willingness to share his knowledge. I did not know Dr. Linn well, but I had many opportunities to be in meetings and planning sessions when he was on the Psychometric Oversight Committee (POC) at AICPA and a consultant to the College Board during my tenures at each organization. As a midcareer professional, it was amazingly instructive to watch Dr. Linn in action. His demeanor was always collaborative and friendly yet the quiet force and the deep thinking that he brought to bear resulted in several powerful insights that guided our subsequent planning, research, and action. His knowledge of the *Standards* and best practices and his concern for test takers and other constituencies were always apparent in his comments.

One observation may clarify my comments. The POC was composed of 10 well-known and highly respected and accomplished individuals. The conversations were always robust with everyone contributing important thoughts based on deep pools of knowledge and experience. But several times I was struck by the quiet that ensued when, as the discussions were winding down, Dr. Linn leaned forward. Every person listened intently to what would turn out to be a synthesis of his thinking melded with the many comments that had been made throughout. The short remarks he made brought everything into focus, helping to establish our actions.

Dr. Linn had a lasting effect on me and one can only imagine the impact he had on so many people, especially those who knew him better. Future committees and organizations will miss his direct wisdom but through his scholarship and the many friends and colleagues he influenced the profession, test-takers and organizations will continue to benefit from his amazing legacy.



Photo taken during a 2002 AICPA Psychometric Oversight Committee meeting. Pictured: Sitting (left to right) Bob Linn, Jim Impara, and Fritz Drasgow. Standing (left to right): Bruce Biskin, Norm Hertz, Rose McCallin, Bob Brennan, Kara Schmitt, Linda Crocker, Steve Downing, and Ron Hambleton.

GRADUATE STUDENT CORNER

CONSIDERATIONS FOR THE JOB SEARCH

Jonathan D. Rollins III, the University of North Carolina at Greensboro

With such a wide variety of jobs in measurement, it is no easy task for graduate students to decide on the context in which we will start our careers. There are numerous opportunities among testing companies, academia, government, and the private-sector. This decision is complicated for many, as we often find that there are multiple topics and areas that interest us during our graduate school training. The gravity of the decision may seem strong, given the number of years we have spent in school and the desire to ensure that we will be satisfied in our career. Despite the complicated nature of trying to find good job fit, I suggest some approaches that can be used to facilitate the search.



To complement this column, I would like to encourage readers to review columns by former graduate students which also relate to entering the job market (i.e., Chris Orem in the [March 2011 issue](#) (Vol. 19, No. 2); Melinda Montgomery in the [December 2013 issue](#) (Vol. 21, No. 4); and Diane Talley in the [March 2014 issue](#) (Vol. 22, No. 1). In line with their pieces, I elaborate on how to prioritize components of what to consider when looking for jobs that best fit our interests and skill sets. First, I describe qualities to consider in defining an ideal job in measurement. Following this, I provide ways to prioritize the list and make comparisons. Finally, I suggest different strategies for managing complications during job searching.

Defining an Ideal Job Description

When reading through job descriptions, it can be difficult to narrow down the list to those which best fit us. As such, this often tends to be based on relative comparisons. After reading through many job descriptions, it may start to become clear which jobs are more desirable when compared to others. However, I would like to advocate the benefit in approaching the job market deductively as well. That is, writing a brief description of an ideal job prior to conducting any serious searches and comparing this description with those that are advertised. The appeal of this approach is that there are defined criteria to guide decisions regarding the extent to which a job is a good fit, and these criteria can be used to better inform the differences found during comparisons of the potential employment opportunities.

While not intended to be exhaustive, a suitable list of criteria can be brought to the forefront: location, schedule, work environment, skills required, salary/compensation (considering benefits and vacation time), future colleagues, employer size, and potential for advancement. The relationships between these criteria may be negligible in certain cases and not in others. First, it is vital to have a clear understanding of each of these factors. If that is not the case, it may help to define them individually. As an example, the criterion for a small company may be roughly 50 employees for one person, while this number may be 200 for another. Afterward, this list can be sorted by order of importance. It may be easier to start with the factors that are least important and would likely not influence any decisions substantially. Subsequently, it may be simpler to determine the order of the remaining factors by considering them using thought experiments. As an example, imagine working in particular job in the northeastern portion of the United States versus the Midwest. As another example, speculate as to the differences in working for one particular college/university that prioritizes factors in the tenure process differently compared to another college/university of interest. While sorting all the factors by importance may be difficult, it will be fairly indicative of any preferences and will allow any previous uncertainty to be better operationalized. Do this without considering restrictions, as these can be addressed later. This purpose in this first portion of the process is to create an archetype of an ideal job that is based on personal preference. In fact, there may be similarities between this list and the ones we created when initially searching for graduate programs.

Though, it would be remiss not to encourage collaboration and networking during this process. Networking is crucial for building professional relationships and learning about different varieties of work that are available. However, it should not be viewed as a process solely for seeking job options. Being able to speak with others who work in areas that align with our interests allows for future collaborations and scholarly inquiries. Moreover, these relationships are important for better understanding the scope of work involved in a typical day. While it may be obvious, seeking advice from faculty members in our graduate programs is crucial. They often have insights into our strengths that may not always be obvious to ourselves. Also speaking with other graduate students is a helpful starting place because they learn about our strengths when collaborating on projects. The combination of these supplemental resources may even serve to be more useful than the list of criteria provided above. Nonetheless, writing an ideal job description which is informed by input and feedback from others will undoubtedly yield a strong start to this process.

Compare and Contrast With Criteria

Having ideal job conditions in mind, ranked by order of significance, makes reading through job descriptions much simpler. It is easier to discern between jobs that fit very well in contrast to those that fit moderately. It may allow jobs to be more easily filtered (i.e., location, setting/field). For example, one may prefer certification/licensure settings over companies that work with achievement or aptitude assessments. Also, this is a helpful exercise in reflecting on how realistic our expectations happen to be. Being too restrictive on location and salary, for instance, may entail not finding any jobs at all. That being stated, limitations in finding jobs that fit well may be self-imposed. This is an important step to consider before making comparisons and contrasts with the criteria, because the criteria must first be reasonable given the current state of affairs in the field. If the criteria are too restrictive, then this exercise is helpful in finding ways to be more flexible with job opportunities that may not have been considered otherwise.

It may be useful to approach this process quantitatively as well. Comparisons could be made by ranking potential jobs based on a weighted composite reflecting the relative importance of your criteria, similar to what is sometimes done in informal product reviews, for example. The rankings can be used conjunctively with supplemental information or advice received from a peer, mentor, or professional colleague. An empirical approach is offered more as a recreational

consideration, however, because the weights would be subjective and based on a single case. In due course, the decision is ultimately guided by personal conviction and circumstances.

Managing Complications

Confines in pursuing an ideal job may be more prevalent in certain situations than others. For instance, there are likely more challenges in relocating whenever other people are drastically impacted by such a decision, including but not limited to children, spouses, and parents. On the other hand, an improved work schedule or salary may be more important, depending on the situation. This is the point at which adjustments can be made to the original list, and it also provides insight into areas in which compromise must be reached. In essence, this provides a systematic framework for addressing the best fitting jobs while also considering concessions that must be made during the process.

A major assumption made in this process has been that we have the time to develop the list and thoroughly think about this process in the midst of so many other activities occurring in our lives simultaneously. The importance of making time to completely think through this process yields the benefit of having a head-start on the process. In fact, the earlier this can be done in one's graduate program, the more benefit there is in knowing which skills to practice in order to develop expertise in specific areas. However, the difficulty in doing so is a limited knowledge of measurement during the initial stages of a graduate program, which will inhibit the extent to which career options are known and can be adequately understood.

Still, a potential complication that may arise is determining where our niche lies within the measurement field. We all wish to find an area in which we can be effective in making a significant impact. Most likely, the tasks we work on will not always be aligned with our primary interests. What we gain in this process is learning other skills and knowledge that may enhance our understanding of our primary interests and may even influence our research lines by incorporating new lines of inquiry. So what originally may be perceived as a complication may be what is needed to expand our comfort zones. In that case, portions of a job description that may not be perfectly aligned with our interests are best perceived as an opportunity to grow professionally and personally.

Conclusion

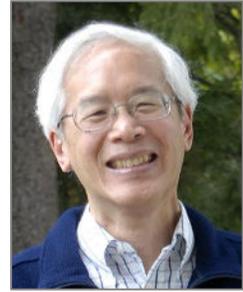
It is prudent to begin considering job fit as early as possible. Other practices may help in defining an ideal job and managing complications in the process. Personally, I found keeping a journal before and after starting my current year-long internship experience with the College Board to be very useful. In doing so, I have a better understanding of how realistic my expectations are, worries that do not come to fruition, and the many skills that I have been learning in the process. Keeping a journal of this process, no matter how detailed or brief it may be, will likely benefit all of us given the magnitude of the task.

It is encouraging to think of job searching deductively because this anchors many aspects of this process in operationalized ways. Equally helpful is carefully considering all opportunities and giving ample time to search so that ideal ones do not pass unnoticed. Multiple resources exist that ease this entire process. The NCME web site has [job postings in the Career Center](#), which can be found under the Resource Center. The AERA [Career Center](#) also contains job listings. More jobs in higher education can be found using the Chronicle of Higher Education's [Vitae web site](#). There likely exists a plurality of ideal jobs, and it is probably less likely that there is only one job that will provide an excellent fit with our interests and skills.

Author note: Jonathan Rollins is a Ph.D. student in the Department of Educational Research Methodology at the University of North Carolina at Greensboro. Some of his interests include IRT parameter estimation, equating/linking, and dimensionality.

SPOTLIGHT ON THE PEOPLE WHO MAKE OUR ORGANIZATION GREAT

Brian Gong, Senior Associate and past Executive Director of the Center for Assessment



How did you get into the field?

I've worked in assessment my whole career. However, I took a roundabout path to get into my current work, which involves providing technical assistance to states for their large-scale assessment and accountability systems. In graduate school, I was mostly interested in learning theory, and assessment was a tool to study and promote learning. I started working after graduate school at Educational Testing Service, working on several projects aimed at investigating how to integrate assessment with instruction. Today we would call those projects formative assessment and curriculum-embedded assessment. In one of those projects I was working with outstanding science teachers and educators in Delaware and Kentucky. When Kentucky started its systemic educational reform, I was intrigued by the idea that assessments could be part of larger policy-framed efforts to improve education intentionally, especially for economically disadvantaged students and schools, and I accepted an invitation to join the Kentucky Department of Education. At first I worked primarily in evaluation and validation efforts around writing portfolios, and then became more involved in the large-scale, on-demand assessment work, and eventually became the associate commissioner in charge of curriculum, assessment, and accountability. Throughout my work at the Department of Education, I was impressed with the expertise and passion that assessment vendors could bring to the endeavor. I was honored when Richard Hill, who was president of the primary vendor working on the Kentucky assessment and accountability program invited me to join with him in starting up a nonprofit consulting business to use the knowledge we had learned in Kentucky to help other states design and implement reliable and effective assessment and accountability systems. Over the past 17 years at the Center for Assessment, I've had the pleasure of working with great colleagues—academics, public employees, consultants, contractors—on really meaningful applied measurement problems in education that have affected millions of students and thousands of school systems.

If you weren't in this field, what would you do?

Although the content of my current work is assessment and accountability, I think the larger challenges are in situating the work in policy systems that are implemented by real people and that are intended ultimately to help students learn more and better. So I see my work as very human—we're developing human capacity along with technical tools and advanced theory and methods. If I weren't in the measurement field, I'd probably be working in teacher development through helping teachers develop their curricula and assessments, which is an area I worked in extensively during my graduate student days.

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

I would advise graduate students first to find a substantive area that they are passionate about and then develop some specific content expertise in it, especially understanding why and how to validate assessment interpretations. For me I was first passionate about learning theory, and it was then clear to me how and why assessment expertise was necessary. Every measurement graduate program will provide graduate students with excellent training in measurement methodologies. But you'll need to develop perspective and wisdom in how to apply those methodologies. That comes in large part through work opportunities where you can apprentice with people who are solving the types of problems you want to be able to solve. Summer internships that challenge you to use your measurement methodological skills as well as applied problem-solving, reasoning, and human-interaction skills may be especially valuable.

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

Outside of work, I especially like to spend time with my family. I have three granddaughters who live nearby and constantly remind me how much I need to keep learning. I do extensive volunteer work through my church. And I love being outdoors in the four seasons of New England, where I live. Over Thanksgiving, we went to the beach even though it was 45 degrees, and my granddaughters took off their boots and danced in the waves. And then we bundled up and had hot chocolate!

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

I have been impressed with the development of many new applications of large-scale testing, and the measurement field's response to those applications in grounding them in validity theory. For example, in state assessments there has been a proliferation of what I call *derivative scores*—scores generated from student test scores, such as growth scores, value-added scores, accountability scores. The attention to trying to establish frameworks, methods, and criteria for evaluating the reliability/precision and validity of those scores' interpretations has been largely a triumph for our field's theoretical and practical machineries. We're currently in the midst of trying to extend our conceptions and tools of validation and fairness to the new challenges of technology-mediated assessments for complex knowledge, skills, and performances for even more diverse populations in less standardized administration conditions. Being willing to tackle those challenges speaks well of our field.

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

I really enjoy going to conferences to learn and network. I find I pick four types of sessions: (a) I try to find sessions that help me with problems that I am working on or anticipate working on, especially if there are presenters I do not know personally so I can meet them after the session and set up a connection to correspond; (b) I like to go to sessions that present overviews of important topics, such as the NCME Career Award lectures or AERA president's addresses; (c) I often go to sessions to support my colleagues, especially those in early stages in their careers; and (d) I always try to select a session or two on a topic I know nothing about but which sounds interesting and important, just to keep from being too narrow.

Who has been a significant influence in your professional life?

I've been blessed by many colleagues who have mentored and encouraged me throughout my professional life. To acknowledge just a few: Professor Bob Calfee agreed to be my dissertation advisor at Stanford after my previous advisor suddenly left the school after I'd had my proposal approved. Bob taught me to work on getting conceptual clarity before working on methodology, although he was a superb methodologist. Bob also taught me the impact that being kind and generous can have on someone just getting started in a challenging field. Vic Bunderson offered me my first job and he and ETS President Greg Anrig set the challenge I'm still working on: putting assessment in the service of learning. My colleague in the Kentucky Department of Education Ed Reidy taught me that people need assessment to get good feedback to improve; assessments need good systems to be embodied; and good systems always have a policy/political side, so that for assessment to have a good impact, we have to pay a lot of careful attention to the context as well as the design. From Rich Hill, with whom I cofounded our consulting business, I learned the value of an answer that was correct to the first decimal place and cost one-tenth as much to produce as an answer correct to the fourth decimal place, how to tell when each was needed, and especially to value colleagues who could do both types of problem solving.

EXPLAINING THE POSITIVE SIDE OF ASSESSMENT: THE STRUGGLE IS REAL

Susan Davis-Becker & Andrew Wiley, Alpine Testing Solutions

We've all been there. In fact, if you've been working in the testing industry for any length of time, you've been there numerous times. That moment when you're chatting with someone at a party/social function/public event. And then they ask: "So, what do you do for a living?" You cringe thinking that as soon as you tell them you work in testing they will tell you how they have been scarred by a particular testing experience; perhaps even worse, their child has had such an experience. If you're like us, you've probably come up with one or several fictitious jobs that have less negative connotations—tax auditor, undertaker, professional hit man—rather than face that uncomfortable conversation where you are blamed for the personal impact that a test (that you have never worked on) has had on a person's life.



This negative perspective on testing is not just part of social conversations—such perspectives appear frequently in popular media and even serve as the source of humor (Foley, Dwyer, Chuah, & Rawls, 2013). Many of us started working in this industry not to argue the politics of testing; instead, we had an interest in the science behind the practice. However, we cannot escape the fact that the majority of the public views our work as something that limits individuals, disadvantages particular groups, and adds extreme pressure to our educational and professional lives. Therefore, it is important that we recognize the numerous benefits of using high-stakes standardized testing. We should all be advocates for the industry to help others understand why testing is a critical part of our society. In this article, we highlight a few of these key ideas and themes that represent the positive side of assessment in both education and professional credentialing.

Accountability and Assurance

This particular concept is likely the cornerstone of why professional credentialing helps everyone in our society. The main purpose of licensure programs (and many certification programs) is protecting the public. How does someone know their doctor is competent? That their lawyer understands the legal system well enough to offer sound advice? That a teacher has the knowledge to educate their child? Through the credentialing process, individuals can demonstrate that they have taken the appropriate steps and can demonstrate adequate knowledge, skills, and abilities to successfully serve in their profession. Without such safeguards, the public might be at risk of incompetent individuals working in these critically important professions. The system is not perfect as professionals are sometimes stripped of their license due to inappropriate practice. However, a properly developed credentialing exam helps reduce the number of such cases.

In education, the results of standardized tests are one of the few pieces of objective standardized information available to the educational community that helps them understand how their children and schools are performing in relation to the tested content. When used properly, the data can provide critical information for all parties and highlight teachers and schools that perform exceptionally well or exceptionally poorly. This data is essential for society to gain the assurance that the educational system is enabling students to learn and develop as expected.

Improving Education and Training

Education and training programs are designed around the goals of the program itself. For many such programs, the key goal for learners is to be able to demonstrate a level of knowledge, skills, and abilities on some type of measure that qualifies them for a next step in their lives (e.g., next grade level, career opportunity). Typically, the curricula for the program will at least in part be designed around the content of this measure. Critics of testing will point this out as teaching to the test and use it as an example of the perils of standardized testing. However, if the test is properly formulated and appropriately aligned to well-defined standards/learning objectives, this teaching to the test can actually be viewed as applying an appropriate focus on the essential knowledge, skills, and abilities.

In education, this goal is typically focused on standards that the state has determined are appropriate for students to learn. In professional credentialing, exam content is typically aligned to a foundational job/practice analysis. In turn, these expectations drive the downstream activities such as curricula design. Without this common set of expectations, educational programs may take the liberty of designing their own curricula, which may or may not reflect the comprehensive set of expectations for a given grade level/subject area or profession.

Reliable (Better) Decision Making

For all of the flaws that can be pointed to regarding standardized testing, in many cases testing is the only type of data available that can provide reliable and standardized information across a large group of students or professionals. In education, decisions about the future of students and the performance of teachers and schools need to be made. This uncomfortable truth is frequently ignored. Making those decisions is never going to be easy and inevitably, some people will not be happy with the decisions. Standardized assessments enable these decisions to be informed with reliable, objective data that should enhance our ability to make these difficult judgments.

The process of professional credentialing also requires making challenging important decisions. In the first two themes, we highlighted how the overall credentialing processes provide benefits through protecting the public as a

screening mechanism and then by setting expectations for educational programs. Using a standardized measure in this process provides the benefit of ensuring a fair mechanism to make decisions about who is ready to enter a professional field. Without a standardized measure such as a large-scale assessment, such decisions may be open to subjective influence or varying sets of expectations mandated by numerous organizations.

How We Can Still Do Better!

Although we have highlighted several key benefits of high-stakes testing in educational and professional credentialing programs, more work still needs to be done to improve the position of our work and increase the benefits of assessments. For example, while we can state that educational assessments can provide objective data to help schools evaluate how well their students and teachers are performing, we can all look at the process we use to provide this information and point to significant limitations. Collectively, many of us would agree that we need to find better, more user-friendly, ways to provide this information and allow test users to take full advantage of the data provided.

In professional credentialing, it is important that as a measurement community we help program leaders understand the importance of communication with their candidate population and professional community. Members of these communities often do not understand how such assessments are developed, administered/scored, and maintained. In turn, many misunderstandings exist about these programs, which logically leads to questions/challenges about the integrity of the decisions made based on the results. In some cases, programs have taken a proactive approach to being transparent about their processes and the professional communities have become more engaged and informed participants in the process as consumers of the exam-related information.

This article provides only a brief set of ideas on how and when standardized assessments can benefit society. Most importantly, the NCME community needs to step up and address this question through the collective wisdom of our members. We encourage all of our colleagues to think about opportunities within their day-to-day work to help with this effort and promote the positive benefit of the work we do.

References

Foley, B. P., Dwyer, A. C., Chuah, D., Rawls, A. (Producers), & Foley, B. P. (Director). (2013). *Testing in the movies and on television* [Motion picture]. United States: National Council on Measurement in Education.

PUBLICATIONS COMMITTEE UPDATE

Rose C. McCallin, Colorado Department of Regulatory Agencies

Wiley publishes the *Journal of Educational Measurement* (JEM) and *Educational Measurement: Issues and Practice* (EM:IP). Each year Wiley prepares a journal citation ranking bibliometrics report for JEM. The annual report provides a variety of data aimed at quantifying and communicating JEM's standing among scholarly journals included in the Social Sciences Citation Index® (SSCI) maintained by Thomson Reuters. If this topic sounds familiar, you may recall a recent editorial about the acceptance of *Educational Measurement: Issues and Practice* (EM:IP) for inclusion in the SSCI (Briggs, 2015).



Wiley sent NCME the 2014 bibliometrics report for JEM about the same time Briggs' editorial appeared in EM:IP, prompting this article about the various metrics NCME receives for JEM and will be receiving for EM:IP beginning next summer. The purpose of this article seemed simple enough: I wanted to share information in the 2014 report for JEM. I soon realized my goal was not at all simple. While there is a long history of efforts to quantify a journal's influence (e.g., see Garfield, 1955), the ownership of citation databases involves a maze of proprietary resources within the publishing industry. Still I felt compelled to move forward in my humble attempt. Please keep in mind the metrics presented and discussed herein represent only some among several measures that attempt to quantify a journal's scholarly influence.

The data in this article covers citations to/from journals in the Thomson Reuters SSCI. I limit the metrics to those produced for the Journal Citation Reports® (JCR) from the Web of Science™, a subscription citation index database operated by Thomson Reuters. The JCR database includes three Web of Science index groups: (a) Science Citation Index®, (b) SSCI, and (c) Arts & Humanities Citation Index®. Each index group contains a limited number of journal titles. Metrics are computed with reference to journals that are accepted in the respective index group. The SSCI index group covers 5,300 journal titles (Thomson Reuters, 2015). The JCR metrics consist of impact factor (IF2), 5-year impact factor (IF5), total citations, immediacy index, cited half-life, and Eigenfactor® (EF) and Article Influence™ (AI) scores. EF and AI scores originated from the Eigenfactor Project, an academic research project cofounded in 2007 by Carl Bergstrom and Jevin West at the University of Washington.

Table 1 provides the SSCI JCR metrics for *JEM* over the past 5 years. The lower portion of Table 1 gives the data needed to compute IF2 for 2012–2014.

Table 1. JCR Metrics for *JEM*: 2010-2014

Metric	2010	2011	2012	2013	2014
Impact factor (IF2)	1.049	1	0.660	0.867	0.922
IF2 without self-citations	—	—	0.447 ^a	0.689 ^b	0.804 ^c
5 yr impact factor (IF5)	1.5	1.297	1.038	1.064	1.025
Total citations to all journal content	865	849	937	972	938
Immediacy index	0.25	0	0.045	0.172	0.042
Cited half-life	10	10	10	10	10
Eigenfactor (EF)	0.00153	0.00216	0.00142	0.00187	0.00172
Article Influence (AI)	0.837	1.168	0.763	1.019	0.899
Data for IF2 calculations					
Citable articles published ^d	24	23	22	29	24
Cites in YYYY to articles published the previous 2 years					
	YYYY				
	2012	23	8		
	2013		27	12	
	2014		12	35	

^aIn 2012, 32% of cites to articles in 2010 and 2011 were self-citations. ^bIn 2013, 20% of cites to articles in 2011 and 2012 were self-citations.

^cIn 2014, 12% of cites to articles in 2012 and 2013 were self-citations. ^dWiley classifies articles as citable if they contain a significant list of references (significant is not defined), an abstract, or if Thomson Reuters determines the content is likely to be well cited. Editorials and meeting abstracts generally are not considered citable articles.

Impact Factors

Impact factors are normative attempts to quantify a journal’s influence. The impact factor is a measure of average citations over a defined period of time (typically 2 or 5 years). The IF2 measures the citation performance of articles in their 2nd and 3rd year of publication, while the IF5 measures the citation performance of articles published in the previous 5-year period. All IF metrics discussed below include self-cites to articles. The general calculations for the IF2 and IF5 are:

IF2 Target Year (TY): \sum of TY citations to *JEM* articles published in TY₋₁ year and TY₋₂ years \div \sum of citable articles published in *JEM* for TY₋₁ year and TY₋₂ years

IF5 Target Year (TY): \sum of TY citations to *JEM* articles published in TY₋₁ year ... TY₋₅years \div \sum of citable articles published in *JEM* for TY₋₁ year ... TY₋₅years

The numerator of the impact factor includes all citations made to *JEM* articles while the denominator consists of only citable articles published in *JEM*. Tables 2, 3, and 4 give the *JEM* IF2 calculations for 2012, 2013, and 2014, respectively.

Table 2. 2012 IF2 Calculation

Cites in 2012 to articles published in:	2010	23	Number of citable articles published in:	2010	24
	2011	8		2011	23
	Sum	31		Sum	47
IF2 calculation	Citations to recent articles		31	= 0.660	
	Number of citable articles		47		

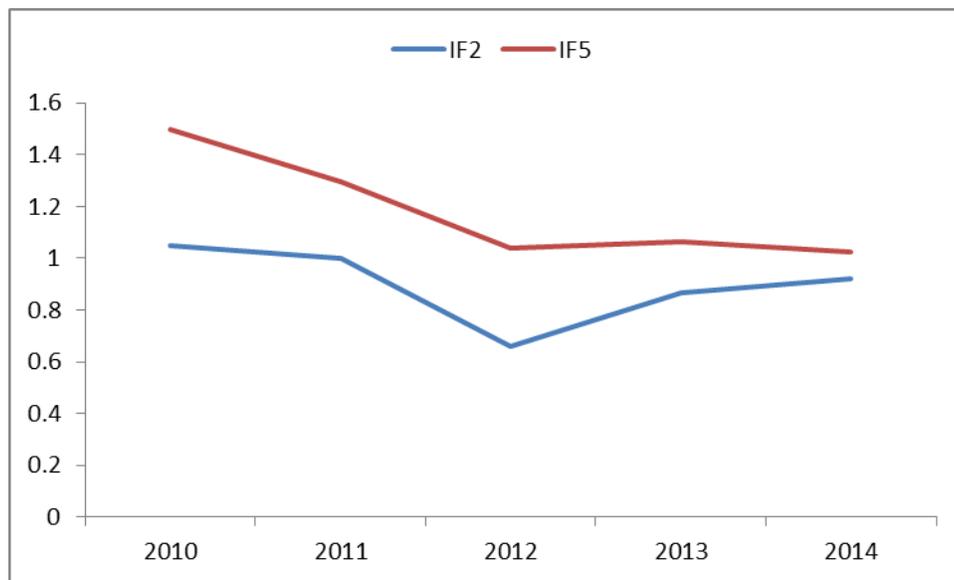
Table 3. 2013 IF2 Calculation

Cites in 2013 to articles published in	2011	27	Number of citable articles published in:	2011	23
	2012	12		2012	22
	Sum	39		Sum	45
IF2 calculation	Citations to recent articles		39	= 0.867	
	Number of citable articles		45		

Table 4. 2014 IF2 Calculation

Cites in 2014 to articles published in	2012	12	Number of citable articles published in:	2012	22
	2013	35		2013	29
	Sum	47		Sum	51
IF2 calculation	Citations to recent articles		47	= 0.922	
	Number of citable articles		51		

From the data in Tables 2-4, it is apparent that IF2 increases when more articles are referenced very quickly (within 2 years of publication). The IF5 gives a snapshot over a longer time span, making it useful in gauging the broader range of citation activity. The IF2 and the IF5 will vary depending on the nature of the academic discipline, journal maturity, and the life span of a journal's articles. Journals that publish articles with a very limited life span will have a higher IF2 and lower IF5 and vice-versa. Figure 1 presents the IF2 and IF5 for JEM over the past 5 years.

**Figure 1. JEM IF2 and IF5 metrics: 2010-2014.**

Other Metrics

The immediacy index indicates how quickly articles in a journal are cited. This index is computed by dividing the number of citations to articles published in a given year by the number of articles published in that year. Journals that are published frequently (e.g., *Educational Researcher*, which is issued nine times annually by the American Educational Research Association) will have an advantage over a journal that is published quarterly (e.g., *JEM*) because the index is a per-article average of articles published in the given year. In addition, articles published in the first issue of an annual volume will have a better chance of being cited than articles published in the last issue of the volume because those in the first issue will have appeared for a longer amount of time than those in the last issue.

The cited half-life is the median age of articles cited in a journal during a calendar year. All *JEM* years in Table 1 have a cited half-life of 10 years. Using 2014 as an example, this means that half of the cited articles were published before January 1, 2005, and half were published afterwards. Newer journals typically have a shorter cited half-life than older journals. It is not surprising that *JEM*, published for almost 52 years with its first volume appearing in 1964, has a cited half-life of 10 years. Finally, journals that publish articles on a continuous basis with rapid review and dissemination (e.g., *AERA Open*, published by the American Educational Research Association) tend to have a shorter cited half-life than journals that follow what may be characterized as more-traditional journal publication models (e.g., quarterly, annually).

Eigenfactor Metrics: EF and AI

The EF represents the number of weighted citations for the year, excluding journal self-citations, to articles published in the previous 5-year period. Citations are weighted according to the quality of the citing journal with citations from higher quality journals weighted more than citations from lesser journals (Bergstrom, 2007; Bergstrom, West, & Wiseman, 2008; West, Bergstrom, & Bergstrom, 2010). The EF is elegant conceptually:

The scientific literature forms a network of scholarly articles, connected by citations. Each connection in this network, that is, each citation reflects the assessment of an individual scholar regarding which papers are interesting and relevant to his or her work. Thus contained within the vast network of scholarly citations is the collective wisdom of hundreds of thousands of authors.

Our approach is to rank journals much as Google ranks web pages. While Google uses the network of hyperlinks on the web, we use citations in the academic literature as tallied by JCR. By this approach, we aim to identify the most influential journals, where a journal is considered to be influential if it is cited often by other influential journals. While this might seem hopelessly circular, it is not: we can iteratively calculate the importance of each journal in the citation network by a simple mathematical algorithm.

This iterative ranking scheme, which we call Eigenfactor, accounts for the fact that a single citation from a high-quality journal may be more valuable than multiple citations from peripheral publications. We measure the influence of the citing journal *divided by the total number of citations appearing* in that journal. (Bergstrom, 2007, p. 314)

The AI is the average influence of a journal's articles over the first 5 years after publication. The AI is calculated by dividing a journal's EF score by the number of articles in the journal. AI scores are normalized so that the mean article in each index group in the Thomson Reuters JCR database has an article influence of 1.00. Thus, a journal with an AI score of 10 means that the average article in that journal has ten times the influence of the mean article in the JCR database for the index group in which the journal is included. Recall there are three index groups in this database: (a) Science Citation Index, (b) SSCI, and (c) Arts and Humanities Citation Index. Figure 2 gives the EF and AI scores for *JEM*, which is included in the SSCI index group.

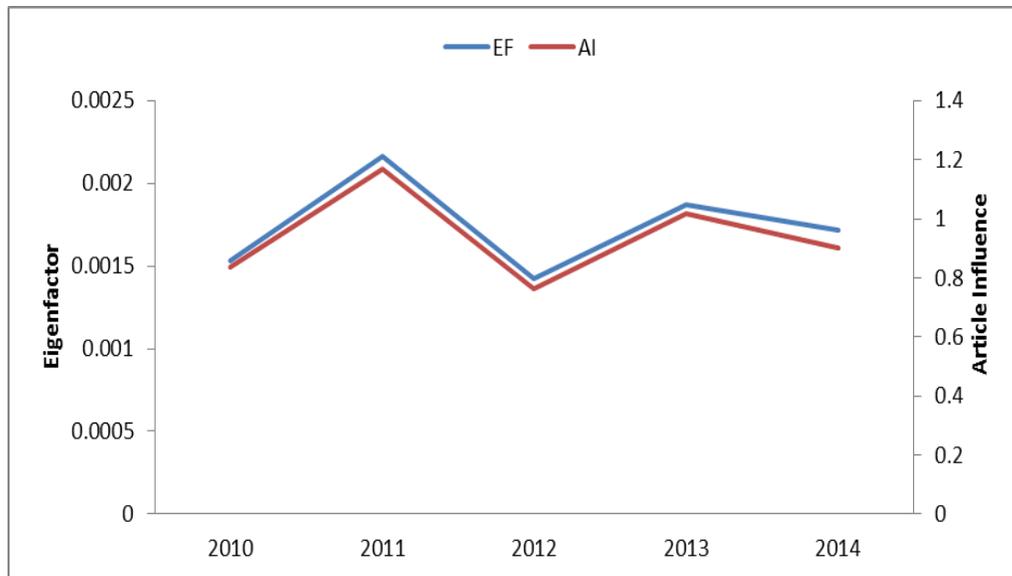


Figure 2. *JEM* EF and AI scores: 2010–2014.

EF metrics are available at no cost at the [Eigenfactor Project web site](#) for every journal in the JCR from 1997–2013. The web site also provides the percentile ranking of a journal’s EF and AI.

I queried the web site by the JCR category, psychology, mathematical, and sorted by AI to obtain the percentile rankings and EF metrics shown in Figure 3. The psychology, mathematical category in the Eigenfactor Project database includes only journals in the SSCI group. In addition, the information is for 2013, the most recent year in the Eigenfactor Project database.

Notice there are journals that publish articles associated with psychometrics and educational measurement that do not appear in Figure 3. Two obvious omissions include *Psychometrika*, the official journal of the Psychometric Society and *Educational and Psychological Measurement*, published by Sage. Further research revealed that in 2008, both journals moved from inclusion in the SSCI group to the Science Citation Index group in the Eigenfactor Project database. While the JCR allow journals to be indexed in more than one of the three Web of Science index groups, the Eigenfactor Project database allows a journal to be included in only one index group.

This article describes some of the bibliometrics Wiley provides for *JEM*; however, there are several other approaches that attempt to measure journal impact. Some other journal citation databases not addressed in this article offer different metrics (for example, but not limited to Scopus®, an Elsevier subscription service, and Google Scholar). It is important to remember that the familiar journal impact factor is but a 2-year measure, easily influenced by the number of citations and the time it takes an article to get published. Measures that include a broader range of citation activity, such as the IF5, and those that weight citations by the quality of the citing journal (the EF and its associated AI), offer alternatives that deserve consideration.



Journal Ranking for *PSYCHOLOGY, MATHEMATICAL*

Order	Journal	Percentile	EF ↓	AI ↓	EFn ↓
1	BEHAV RES METHODS 1554-351X	EF: 92 AI: 95	0.022	2.3	2.4
2	PSYCHON B REV 1069-9384	EF: 91 AI: 91	0.020	1.6	2.3
3	J EDUC BEHAV STAT 1076-9986	EF: 55 AI: 83	0.003	1.1	0.3
4	J EDUC MEAS 0022-0655	EF: 47 AI: 80	0.002	1.0	0.2
5	APPL PSYCH MEAS 0146-6216	EF: 56 AI: 75	0.003	0.9	0.3
6	APPL MEAS EDUC 0895-7347	EF: 27 AI: 51	<0.001	0.5	<0.1

Figure 3. 2013 journal percentile rankings sorted by AI for JCR category, psychology, mathematical. Journal name (sponsor/society/publisher): *Behavior Research Methods* (Psychonomic Society), *Psychonomic Bulletin & Review* (Psychonomic Society), *Journal of Educational and Behavioral Statistics* (American Educational Research Association and American Statistical Association), *Journal of Educational Measurement* (National Council on Measurement in Education), *Applied Psychological Measurement* (Sage), and *Applied Measurement in Education* (Taylor & Francis).

References

Bergstrom, C. T. (2007). Eigenfactor: Measuring the value and prestige of scholarly journals. *C&RL News*, 68(5), 314–316.

Bergstrom, C. T., West, J. D., & Wiseman, M. A. (2008). The Eigenfactor metrics. *Journal of Neuroscience*, 28(45), 11433–11434.

Briggs, D. C. (2015). Editorial. *Educational Measurement: Issues and Practice*, 34(2), 1.

Garfield, E. (1955). Citation indexes for science: A new dimension through association of ideas. *Science*, 122(3159), 108–111.

Thomson Reuters. (2015). Web of Science: The complete citation collection. Retrieved from <http://wokinfo.com/citationconnection/>

West, J. D., Bergstrom, T. C., & Bergstrom, C. T. (2010). Big Macs and Eigenfactor scores: Don't let correlation coefficients fool you. *Journal of the American Society for Information Science and Technology*, 61(9), 1800–1807.

CAREER CONTRIBUTION COMMITTEE UPDATE

Wayne Camara, ACT, Committee Chair

The NCME Award Committee for Career Contributions to Educational Measurement completed its review of ten nominees this fall and forwarded the information on the final selection to the Board. A majority of this year's nominees were carried over from 2015, but a few additional submissions were received for 2016. Barbara Plake will be chairing the committee next year and returning members include Suzanne Lane, Michael Kane, and Wayne Camara, and Stella Kim (graduate student member).

This is an annual award that honors a person whose contributions over a career have had a widespread positive impact on the field of educational measurement. These influential contributions might include one or more of the following: theoretical or technical developments; ideas or conceptualizations of information to the public about educational measurement that has widely influence public understanding; and/or applications of theory through procedures, instrument, or program development that have influenced broadly the nature of measurement and practice. The recipient of the award will receive a monetary award of \$3,000. Additionally, there will be a plenary award ceremony with an invited presentation by the award recipient at the annual meeting. The call for 2017 will be published in early spring with an August deadline for submission of materials.

FUND DEVELOPMENT COMMITTEE UPDATE

Linda Hargrove, Wayne Camara, Linda Cook, Deborah Harris, Suzanne Lane, Seohong Pak, and Cathy Wendler



Linda Hargrove (chair)



Wayne Camara



Linda Cook



Deborah Harris



Suzanne Lane



Seohong Pak



Cathy Wendler

NCME's 2015 Annual Meeting marked the beginning of Year 2 for NCME's Mission Fund. The fund provides a donor-supported avenue for supporting activities that advance NCME's mission in the science and practice of measurement in education. More than 120 members contributed almost \$39,000 in Year 1. Thanks to their support, NCME was able to sponsor a pre-session workshop for more than 40 state assessment professionals at CCSSO's June 2015 National Conference on Student Assessment. The workshop addressed the implications of the *Standards* in everyday practice and how they may be used to prepare for the peer review process. ([Click here](#) to view slides and audio from the workshop.)

As the end of the calendar year approaches, the Fund Development Committee encourages NCME members to build upon the momentum started in Year 1 by donating to NCME's Mission Fund in Year 2. While NCME members have contributed in many ways over the years to advancing NCME's mission, both monetarily and nonmonetarily, three members kicked off Year 2 contributions to the Mission Fund through the donation of forthcoming book royalties. These contributors are Suzanne Lane, Mark Raymond, and Thomas Haladyna, editors of the second edition of the *Handbook of Test Development*. Second edition editors were inspired by the generosity of the original editors of the *Handbook*, Steven Downing and Thomas Haladyna, who donated the first edition's royalties to NCME several years ago.

Please consider joining past and current donors to the Mission Fund to continue building funds in Year 2. If you have previously donated, please consider donating again. The sooner funds are raised, the sooner further activities that reinforce NCME's mission, vision, and goals can be supported. As a dedicated development fund, the Mission Fund has the potential to support activities that go beyond those that can be supported through NCME's annual operating budget. As outlined in the [NCME Mission Fund brochure](#), types of activities include:

- Workshops or small conferences designed to improve measurement and testing theory and application or expansion of measurement knowledge into other fields, such as teacher education
- Funding for the dissemination of measurement knowledge and procedures both domestically and internationally, such as projects for promoting proper test use
- Outreach activities such as grants to encourage exchange in the science and practice of measurement
- Support for the professional development of graduate students and early career scholars

As the end of the calendar year approaches, NCME members who desire to make one or more tax-deductible cash donations to the Mission Fund can donate:

1. Online by going to the [NCME Sign In Page](#)

- Sign in with your member number and password **before** choosing the Donate link in the upper right hand corner;
- Next, choose the Donate link; and
- Enter the amount for the Mission Fund before moving to complete credit card payment information.

OR:

2. By using the downloadable [NCME Mission Fund brochure](#).

- Print, complete, and detach the Contribution Form in the brochure and
- Mail or fax to NCME as instructed in the brochure.

As an added convenience, members also may donate to the Mission Fund when paying for membership renewal online.

NCME FITNESS WALK/RUN CORNER

*Jill van den Heuvel, Alpine Testing Solutions
Katherine Furgol Castellano, Educational
Testing Service
(with Brian French, Washington State
University, Pullman, advising)*



We hope you've had a great fall full of leaf peeping, apple picking, pumpkin carving, and Thanksgiving eating! As the chill nips the air, walking and running outside may start to feel like a battle against the elements. Enjoy the challenge!

We have been busy planning a great NCME Fitness Walk/Run event in Washington, DC in April 2016. Note that the NCME conference has been moved up a day from previous years, spanning Saturday to Monday instead of Friday to Sunday. The Fitness Run will still be the day after the NCME Breakfast as it has been for several years, but it will now be on a Monday: April 11, 2016 (5:45 – 7 am). Remember to sign up for the run/walk when registering for the conference, and to encourage your friends and colleagues to register too. If you convince enough of them, you may have a chance to beat defending champions University of Iowa (university category) or Pacific Metrics (testing company category) in the Team Participation Competition!



This year we are bringing back the “design the shirt” contest! We are looking for the NCME membership to create and then select the main design for the shirt (we will add sponsor logos to the winning design). Designs should use a maximum of two colors and not include any copyrighted images. Submissions can be for a single front/back print or a separate front and back print. Look out for more details in coming weeks from NCME via email. Submissions will be due to [Jill](#) via email no later than February 1, 2016, voting will occur mid-February and the winning design will be revealed on the shirts at the conference! We are looking forward to your creative submissions!

Also, we're looking for creative additions, changes, and incentives for the 2016 NCME Fitness Run/Walk so look out for these and email [Katherine](#) if you have any fun ideas.

Keep moving and happy holidays!

NCME SUNRISE YOGA

Kristen Huff, ACT, and Brian Patterson, Pearson

Please join us for the second annual NCME Sunrise Yoga. We will start promptly at 6:30 am for 1 hour on Saturday, April 9 at the Renaissance in the meeting room level foyer (near registration). There will be a \$10 reservation fee that covers your mat rental. **NO EXPERIENCE NECESSARY.** Just bring your body and your mind, and our friends from  flow yoga center will do the rest. Floor space is limited, so sign up soon. Namaste.





Inaugural NCME Sunrise Yoga Session, Chicago, April 2015. Kristen Huff is pictured holding the yoga sign.

NCME ELECTIONS

NCME elections are now in progress. Polls will remain open until January 15, 2016. You will need the ballot link that was e-mailed to you to get to the candidate bios and voting options on the elections site. Additional ballot links were sent in a reminder e-mail recently. Please contact the NCME office if you have any trouble getting to the elections site.

Ballot

President-Elect

Randy Bennett, Educational Testing Service
Brian Gong, Center for Assessment

Board Member From a Testing Company

Susan Davis-Becker, Alpine Testing Solutions
Denny Way, Pearson

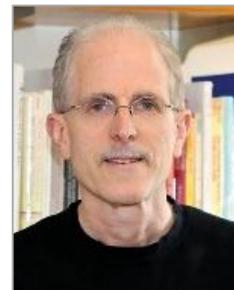
Board Member at Large

Derek Briggs, University of Colorado at Boulder
Marianne Perie, Center for Educational Testing and Evaluation at the University of Kansas

Candidate Bios

Randy Bennett

I began working at Educational Testing Service (ETS) in 1978 as a predoctoral intern. After receiving my degree from Teachers College, Columbia University, I took a position at ETS. It's 36 years later, and I now work in the Research & Development division as the Norman O. Frederiksen Chair in Assessment Innovation. For most of that 36-year period, my research has concentrated on integrating advances in the cognitive and learning sciences, technology, and measurement to create new approaches to assessment.



Two examples of that research stand out for me. From 1999 through 2005, I directed the NAEP Technology-Based Assessment project, which explored the use of computerized testing in NAEP and laid much of the groundwork for the program's current transition to technology delivery. We conducted three major studies for the National Center for Education Statistics (NCES) in mathematics, writing, and problem solving with technology. I believe these studies were the first to administer computer-based performance assessments to nationally representative samples of elementary and secondary school students. They were also the first to collect and use log-file data in such samples to try to measure the cognitive processes employed in problem solving.

As a second example, since 2007, I've been directing the *Cognitively Based Assessment of, for, and as Learning* (CBAL) initiative. Through CBAL, we've created domain theory in the English language arts (ELA), mathematics, and science, including an extensive set of publicly available ELA learning progressions. From that theory, we've tried to generate assessments that model good teaching and learning practice—assessments that, in essence, are learning experiences in themselves. And we've tried to evaluate whether the theory embodied in those progressions and assessments is empirically supported. A variety of organizations have used CBAL results in the form of theory, assessments, or professional development. Examples include the Smarter Balanced Assessment Consortium, GlassLab, the World Bank, Impaq International, and the Singapore Examinations and Assessment Board.

With respect to research dissemination, my publications can be found through the [ETS ReSearcher](#) or [Google Scholar](#). I'm particularly honored to have been coauthor of two chapters in NCME-sponsored publications: "Technology and Testing" in *Educational Measurement* (4th ed.) and "Validity and Automated Scoring" in the first book in the new *Applications of Educational Measurement and Assessment* series. I've also contributed to NCME's research dissemination efforts as an editorial board member for both the *Journal of Educational Measurement* and *Educational Measurement: Issues and Practice*.

I am currently an officer of the International Association for Educational Assessment (IAEA), which is composed of governmental and private testing organizations. Through IAEA, I've had the very good fortune to develop relationships with educational measurement organizations around the world.

If I had one hope for the future of NCME, it would be for the organization to become a more vocal advocate within the measurement community for intentionally designing tests to have greater positive impact on teaching and learning and with the public for the value that well-constructed and appropriately used tests can provide.

Brian Gong

I received a Ph.D. in education from Stanford University and have worked in applied measurement settings for my entire career, seeking to both improve assessments and assessment practices, and to use assessments to improve public schooling. I worked at Educational Testing Service as a research scientist for several years, developing classroom assessments in middle school science that would help “assessments be sensitive to instruction, and instruction sensitive to assessment.” Work in innovative assessments led to my taking a position in the Kentucky Department of Education (KDE) during the well-known Kentucky systemic reform years. As associate commissioner of education in KDE, I worked on validation efforts and improvement of large-scale assessments in programs such as writing portfolios, performance tasks, early large-scale state assessments with extensive constructed response, and accountability systems. With a colleague, I cofounded the Center for Assessment, a nonprofit consulting organization providing technical assistance primarily to state departments of education. Over the past 17 years through the Center for Assessment, I have been involved in helping state departments of education and others address every aspect of operational large-scale assessment and accountability. After serving as the executive director of the Center for Assessment for 10 years, I am currently a senior associate at the Center for Assessment.



I have served on many national committees, including on the committee responsible for the 2014 revision of the APA/AERA/NCME *Standards for Educational and Psychological Testing*, on the technical advisory group for the U.S. Department of Education’s (USED) Peer Review for English Language Proficiency assessments, and on the technical advisory group for the USED guidance on growth models. I have regularly presented at NCME, the Council of Chief State School Officers (CCSSO), and AERA’s annual meetings.

NCME is the premier U.S. professional organization for professionals involved in assessment, evaluation, testing, and other aspects of educational measurement. NCME provides the mechanism for fostering sharing of contemporary scholarship with other members of the community (primarily through the NCME annual meetings, *JEM*, and the recently launched book series), professional development (through courses at the NCME annual meeting as well as the *Instructional Topics in Educational Measurement Series [ITEMS]* and other courses), networking and communication (through the newsletter), a professional home for members (through strong institutional identify, prestigious professional awards, etc.), and official professional representation (on things such as the joint *Standards for Educational and Psychological Testing*).

In terms of serving its internal community, NCME is doing an outstanding job. This is fortunate because the need for NCME’s contributions to the larger public community have never been greater. The need for scholarship and research continues at an increased pace, fueled by changes in technology and desires for educational measurement in new constructs and contexts. Application of measurement in new testing situations calls for operational and methodological advances. Policy demands more actionable information from testing and evaluation, while being more credible. At the same time, there has never been more opposition to testing and evaluation, or misunderstanding and distortion of measurement.

NCME can make a great contribution by providing leadership with partners in sharing accurate and relevant scholarship, technical advice, and outreach for communication to those policy makers who shape and practitioners who are most affected by the purposes and uses of testing. This is a fundamental obligation of attending to consequential impact. I would be humbly enthusiastic to work with the NCME Board and members to strengthen and extend NCME’s efforts in this vital area.

Susan Davis-Becker

In receiving my Ph.D. in assessment and measurement from James Madison University, the program provided me with an amazing start in the field through opportunities to gain consultation experience with higher education assessment. From there, I joined the team at the Oscar and Luella Buros Center for Testing and began collaborating with a variety of educational and credentialing testing programs. Since 2007, I have been a part of Alpine Testing Solutions and currently serve as a senior psychometrician and director of professional credentialing. In this role, I consult with a variety of national and international certification and licensure programs on program design, psychometric issues, validity research, and related policy considerations. One of my favorite aspects of this role is the frequent challenge to create solutions for clients that align with a program's purpose that also strive to meet our profession's expectations.



My research opportunities have come predominantly from applied experiences with a range of testing programs in education, certification, and licensure. I enjoy presenting and publishing work focused on the challenges/issues that arise in practice along with practical solutions and ideas to help address these challenges. I have presented regularly at conference such as NCME, AERA, Institute for Credentialing Excellence, Association of Test Publishers, and International Test Commission; and published my research efforts individually and collaboratively in numerous white papers, journal articles, and book chapters. My recent research has included topics such as test design, test evaluation, standard setting, and score reporting.

I have enjoyed being an active member of NCME for more than 10 years now. Volunteering within the organization on various committees and recently serving as the *NCME Newsletter* editor has been a great opportunity to learn more about different aspects of the industry while interacting with a number of great colleagues. Most recently, I am excited to be working with the NCME Editorial Board and a colleague on leading the production of a volume on testing practices in the credentialing industry.

I am honored to be considered by colleagues as a representative to the NCME Board of Directors. In representing the membership, I bring an applied perspective that is based on experience balancing theory, policy, and practice. I am particularly interested in helping the organization grow in visibility across the domains in which testing practices are applied.

Walter (Denny) Way

I am pleased to offer this biographical statement in support of candidacy for the NCME Board. I have been in the measurement field for more than 25 years. I received a Ph.D. in educational measurement and applied statistics from the University of Iowa in 1988. My current job is vice president of assessment solutions and design at Pearson. I have worked for Pearson for 11 years, at various times leading groups in psychometrics, research, test development, and content technology. In my current role, I am responsible for the design and implementation of assessment solutions for Pearson's business units worldwide, as well as Pearson's assessment activities in support of the Partnership for Readiness for College and Careers (PARCC).



Prior to joining Pearson in 2004, I spent 16 years with Educational Testing Service (ETS) in a variety of technical and management roles. While at ETS, I directed research and operational psychometrics for testing programs in higher education admissions, licensure and certification, and K-12 assessment. I was involved in the development of several of ETS' computerized adaptive testing programs in the 1990s. I continue to maintain an interest in computer-based testing and serve as a senior advisor to the teams at Pearson that are responsible for computer-based and computerized adaptive testing. Additionally, my research interests include equating and scaling, test design, and applications of technology to large-scale assessment. Most recently, I have contributed to chapters in the new *NCME Applications of Educational Measurement and Assessment* book series on topics that include the design and development of end-of-course tests; the comparability of scores based on different testing conditions, modes and

devices; and estimating, interpreting, and maintaining the meaning of credentialing test scores. As these chapter topics illustrate, I have broad experiences with a variety of measurement applications, as well as the research that has supports them.

I have been active in NCME throughout my career. I have been a constant presenter and/or discussant at the annual NCME conference since my graduate school days. I served as an advisory editor for the *Journal of Educational Measurement* from 2005 to 2008 and was a member of the joint committee revising the *Standards for Educational and Psychological Testing*, published in 2014. I served on the *NCME Career Award Committee* from 2011 to 2013, chairing the committee in 2012. I am the current chair of the *NCME Standards and Test Use Committee*.

As someone with extensive experience tackling real life measurement problems, I am keen on promoting applied research with strong practical applications and would like to see more recognition of applied work in various NCME publications and initiatives. I believe my views in this area and advocacy of practical measurement applications will benefit the NCME Board during a time when there are unprecedented changes occurring in our field.

Derek Briggs

I received my Ph.D. in education from the University of California, Berkeley in 2002 with a specialization in psychometrics and applied statistics. Prior to that I graduated with a major in economics in 1993 from Carleton College, and in between these two events I worked as an assistant analyst at the Congressional Budget Office in Washington, DC. I had expected to focus on the economics of education in graduate school, but after one course on measurement in my first semester I was hooked. I was very lucky to have three amazing mentors during my time at Berkeley, each of whom had a profound impact on my development as a scholar. I started as an assistant professor at the University of Colorado, Boulder in the fall of 2003. I was promoted to associate professor with tenure in 2009, and promoted to full professor in 2013. I am currently the chair of the Research and Evaluation Methodology program and director of the Center for Assessment, Design, Research and Evaluation. These days my own students are becoming professors and practicing psychometricians, something that makes me very proud. Time flies.



I have been a member of NCME since 2000 and have served the organization in many capacities over the years. I am just completing a 3-year term as the editor of *Educational Measurement: Issues and Practice (EMIP)*. This has been a labor of love for me. It has been a privilege and an honor to serve NCME in this manner. I've had the opportunity to work with so many submitting authors and reviewers to produce published manuscripts that I hope have had a positive influence on the field. In addition to my role as *EMIP* editor, I have served as chair of the Outstanding Doctoral Dissertation Award Committee, as a longstanding reviewer of NCME conference proposals and *Journal of Educational Measurement* manuscripts, and as a frequent presenter and discussant at the NCME annual conference.

One of my favorite lines from the bio on my faculty web page is as follows "Dr. Briggs's long-term research agenda focuses upon building sound methodological approaches for the measurement and evaluation of growth in student learning. His daily agenda is to challenge conventional wisdom and methodological chicanery as they manifest themselves in educational research, policy and practice." I like to think that I'm the sort of person that can indeed successfully challenge conventional wisdom, but with a smile on my face while appreciating, with humility, just how much there is for me (and all of us) to keep learning about our field. I want to challenge my friends and colleagues at NCME to do more, but to always think critically about what they are doing and why. This is the role I would plan to take as an at-large member of the NCME Board.

Marianne Perie

I am currently the director of the Center for Educational Testing and Evaluation at the University of Kansas. I received my B.A. in psychology from Cornell University and my Ph.D. in educational research, measurement, and evaluation from the University of Virginia. I began my postdoctoral career working at the American Institute for Research (AIR) as AIR started an assessment division, working on such oldies but goodies as the Voluntary National Test (VNT) and the California High School Exit Examination. While there, I conducted cognitive labs to determine the suitability of international assessments such as TIMSS and PISA for U.S. students. After 8 years, I moved to Educational Testing Service (ETS) to focus specifically on standard setting across the K-12, licensure, and international programs. I also spent 2 years working on NAEP reporting and validity studies. I next moved to the National Center for Improvement of Educational Assessment and focused on the integration of test design, psychometrics, and education policy. In 2013, I moved to CETE in order to interact more with graduate students. There, I've worked on an alternate assessment consortium, two state testing programs, and a career and technical education certification program. I've also cotaught a course in standard setting (at UMD) and in current issues in educational assessment (at KU).



My interests lie primarily in the areas of standard setting and validity evaluation. I've worked with state departments of education and their vendors to improve the technical documentation of their assessments around a validity argument. I'm also interested in equitable assessment of students with disabilities and English language learners. In addition, I've worked to differentiate formative, interim, and summative assessments and believe strongly that an assessment that tries to serve too many purposes serves none well.

I've been an active member of NCME since 1991. I've served on multiple committees and have chaired the Outreach Committee, the Brenda Loyd Dissertation Committee, and the Publications Committee. For three years, I also chaired the planning committee of the Council of Chief State School Officer's (CCSSO's) National Conference on Student Assessment and helped bring joint sessions to both their and NCME's annual conferences.

I believe NCME has a role to play in informing policymakers about the technical requirements of legislation they craft. Too often, our voices have been given only polite reception and not serious consideration. As the Elementary and Secondary Education Act is being reauthorized, states are developing new accountability systems that require the use of assessment data in teacher, school, and district evaluations. Measurement professionals need to be part of the conversation when designing such systems.



The NCME Mission Fund

Make a Difference!

Advance NCME's Mission,
Vision, & Goals

- *The Mission Fund allows NCME to carry out a variety of mission-driven activities such as workshops and small conferences, outreach, dissemination, and support of early career scholars and students.*
- *The Mission Fund supported its first activity, a workshop at the June 2015 CCSSO National Conference on Student Assessment to help more than 40 state personnel understand implications of the newly released AERA/APA/NCME Standards. [Click here](#) to view the presentation.*
- *Please help NCME reach its \$50,000 goal for Year 2. Make your tax-deductible donation today at: [NCME Sign In Page](#) (member login required). Click "Donate" at top right, enter amount, and proceed to checkout.*

Newsletter Advisory Board

LANIE BRADSHAW, University of Georgia
HEATHER BUZICK, Editor, Educational Testing Service
WILLIAM DARDICK, George Washington University
SUSAN DAVIS-BECKER, Alpine Testing Solutions
JOHN DENBLEYKER, Houghton Mifflin Harcourt
ELLEN FORTE, edCount
SARA HENNINGS, American Dental Association
ANDREW HO, Harvard Graduate School of Education

GERALD MELICAN, Consultant
MARIA ELENA OLIVERI, Educational Testing Service
THANOS PATELIS, Center for Assessment
S.E. PHILLIPS, Consultant
JONATHAN ROLLINS, University of North Carolina, Greensboro (Grad Student Rep)
EDYNN SATO, Pearson
CHRISTINA SCHNEIDER, Center for Assessment
BRANDI WEISS, George Washington University

Send articles or information for this newsletter to:

Heather Buzick
Educational Testing Service
660 Rosedale Road
Princeton, NJ 08541

e-mail: hbuzick@ets.org

The newsletter is published quarterly (March, June, September, and December). Download issues at <http://www.ncme.org/ncme/News>

The *Newsletter* is not copyrighted; readers are invited to copy any articles that have not been previously copyrighted. Credit should be given in accordance with accepted publishing standards.