Volume 56 | Winter 2017

The NERA Researcher

NERA NORTHEASTERN EDUCATIONAL RESEARCH ASSOCIATION

Inside this issue

The President's Message

Craig Wells

Dear NERA Colleagues,

As I was considering what to include in my first President's Message, I began to reflect on the various events that occurred in 2016. We suffered great losses including several beloved authors, pop icons, and activists such as Eliezer Wiesel, Umberto Eco, Harper Lee, Prince, Carrie Fisher, David Bowie, and Muhammad Ali. Our community also lost Roscoe Brown, NERA's President in 1969, who was a war hero and one of the Tuskegee Airmen in World War II. All of these people had a profound impact on how we view world events and participate in activities to support social justice. Interestingly, we as educators are not unlike these prominent public figures in that we also influence society through our work educating children and young adults to think about problems we face in society and methods for developing possible solutions. In addition, our research can shape our world views through its impact on public policy. These effects were prominent throughout the 2016 NERA conference where I saw stimulating research being presented and discussed among graduate students, academics, and professionals who care deeply about education.

One of the most challenging aspects of 2016 was the United States presidential election, which was the most vitriolic campaign in recent history. During the campaign, it appeared that winning through any means necessary was more valued than making accurate claims that were consistent with empirical evidence (although this can be said about most elections, it seemed to be worse in 2016). In fact, one of the things that shocked and frustrated me about the election was how often candidates or news agencies made assertions without a shred of empirical evidence or, perhaps worse, that were in contradiction to empirical evidence. As an educator and a scientist, I find the lack of respect for empirical evidence and how to use data appropriately appalling. Once I got over my shock, I started to think about how we can solve this *post-truth political* problem. How can we teach and emphasize the importance of the appropriate use of empirical evidence? An obvious solution (but not the only solution) is through formal education. As educators in PK-12 and higher education, we have the ability (and perhaps responsibility) to emphasize to our students the importance of and how to use empirical evidence to evaluate claims and ideas (or hypotheses) as well as the strengths and limitations of empirical evidence. We are familiar with basic research concepts such as reliability, validity, and generalizability—concepts that should be taught to all students throughout all of the grades to support critical thinking skills. By the time students graduate high school, they should understand the interplay between theories, hypotheses, facts, and observations. With these skills, we will have citizens who can evaluate the validity of claims and the quality of data sources for themselves to aid in decision-making. It is important to note that use of empirical evidence and the scientific method is not relegated to just the "science" domains such as chemistry, physics, or biology, but to any field that makes an empirical claim such as psychology, history, and other areas of social science.

Despite the sad and frustrating events that occurred in 2016, there were many successful events that we should be proud of and celebrate. For example, thanks to Charlie DePascale and the conference co-chairs, Joshua Marland, Scott Monroe, and Molly Faulkner-Bond, the NERA 2016 annual meeting was an enormous success. The conference theme was appropriate and timely given the importance and impact of

Continued on page 3

The NERA Researcher

The NERA Researcher is the official newsletter for members of the Northeastern Educational Research Association.

Message from the Editors

Happy New Year NERA Members!

We hope everyone has enjoyed a restful holiday season! As documented in several places in this issue, the 47th Annual NERA Conference was a great success. Please join us in thanking our past president, Charlie DePascale, and our 2016 conference co-chairs Molly Faulkner-Bond, Joshua Marland, and Scott Monroe for all of their hard work. We look forward to seeing what our next year has in store under the new leadership of President Craig Wells and 2017 conference co-chairs Daniel Jurich, Whitney Smiley, and Jason Kopp.

This issue contains highlights from our 2016 conference, as well as special calls for papers in The Journal of Educational Leadership and Policy Studies and chapter proposals for a special volume entitled Disruptive Views of Gender and Sexuality in K-12 and Teacher Education.

As always, a special thank you to Barbara J. Helms for her continued assistance in editing *The NERA Researcher*!

Haifa Matos-Elefonte and Katherine Reynolds
The Editors

Executive Committee

President:
Craig Wells
University of Massachusetts Amherst
cswells@educ.umass.edu

Past President:
Charlie DePascale
National Center for the Improvement of Educational Assessment
cdepascale@nciea.org

President-Elect:
Javarro Russell
Educational Testing Service
jrussell@ets.org

Treasurer:
Steven Holtzman
Educational Testing Service
treasurer@nera-education.org

Secretary:
Mary Yakimowski
Sacred Heart University
yakimowski@aol.com

Board of Directors

Carol Barry (2014-2017)
The College Board
cabarry@collegeboard.org

Felice Billups (2014-2017)
Johnson & Wales University
fbillups@jwu.edu

Andrew Jones (2015-2016)
American Board of Surgery
ajones@absurgery.org

Salika Lawrence (2015-2018)
Medgar Evers College, City University of New York
llawrence@mec.cuny.edu

Ross Markle (2015-2018)
Educational Testing Service
rmarkle@ets.org

Liz Pyburn (2016-2017)
James Madison University
pyburnem@jmu.edu

Tia Sukin (2016-2019)
Pacific Metrics
tsukin@pacificmetrics.com

Appointed Positions

2017 Conference Co-Chairs:
Daniel Jurich
National Board of Medical Examiners
jurichdp@gmail.com

Whitney Smiley
SAS
whitney.smiley@sas.com

Jason Kopp
American Board of Surgery
koppjp@gmail.com

***The NERA Researcher* Co-Editors:**
Haifa Matos-Elefonte
The College Board
hmatoselefonte@collegeboard.org

Katherine Reynolds
Boston College
reynolds.katherine@bc.edu

NERA Webmaster:
Bo Bashkov
American Board of Internal Medicine
bo.bashkov@gmail.com

NERA Conference Proceedings Co-Editor for Digital Commons:
Javarro Russell
Educational Testing Service
jrussell@ets.org

Mentoring Program Co-Chairs:
Jonathan Rubright
National Board of Medical Examiners
jrubright@nbme.org

Juliette Lyons-Thomas
Regents Research Fund
juliette.lyons-thomas@nysed.gov

STANDING COMMITTEES

Communications Chair:
Monica Erbacher Smith
University of Arizona
MonicaKErbacher@gmail.com

Membership Chair:
Tabitha McKinley
New Jersey Department of Education
tabitha.mckinley@doe.state.nj.us

Nominations Chair:
Charlie DePascale
National Center for the Improvement of Educational Assessment
cdepascale@nciea.org

Graduate Student Issues Chair:
Liz Pyburn (2016-2017)
James Madison University
pyburnem@jmu.edu

Teacher-as-Researcher Chair:
Darlene Russell
William Patterson University
russelld@wpunj.edu

AWARDS COMMITTEES

Thomas Donlon Mentoring Award Chair:
Lisa A. Keller
University of Massachusetts Amherst
lkeller@educ.umass.edu

Leo D. Doherty Memorial Award Chair:
Elizabeth Stone
Educational Testing Service
estone@ets.org

Lorne H. Woollatt Distinguished Paper Chair:
Abigail Lau
Emmanuel College
Lau.abigal@gmail.com

AD HOC COMMITTEES

Infrastructure Chair:
Tia Sukin
Pacific Metrics
tsukin@pacificmetrics.com

Conference Ambassadors Chair:
Rochelle Michel
Educational Testing Service
rmichel@ets.org

Site Selection Chair:
Javarro Russell
Educational Testing Service
jrussell@ets.org

disseminating information on peoples' behaviors (e.g., voting in the 2016 presidential election). Regardless of how much high quality research is being conducted in education, if we do not find a way to communicate the research findings to the public in an understandable manner, the positive effect will be trivial or lost completely. The two keynote speakers, Lee Badgett and Jonathan Supovitz, gave informative and thought-provoking talks on how to communicate research effectively to the public. The invited panels also provided useful information on communication from interesting perspectives. The professional development workshop topics were diverse and covered many important methodological skills educational researchers use to conduct research. The paper and poster presentations were exciting, exhibiting high-quality research that can have a positive impact on education. Lastly, Charlie DePascale's Presidential Address was a humorous, enlightening, and thought-provoking view of validity (effectively illustrated through references to Taylor Swift).

Regardless of the challenges we face in the near future, I remain optimistic that we can have a positive influence on society through education. Stephen Pinker, a Harvard professor in psychology, has argued that the world is becoming more peaceful. I would like to think that education is playing a role in making the world a better place and we as educators are partly responsible for the positive changes. Although education is not a panacea for society's ills and problems, it can have an enormously positive impact. In fact, without a well-educated public, it is not possible to have a well-functioning democracy. To do our part, the conference co-chairs, Daniel Jurich, Jason Kopp, and Whitney Smiley, and I are working diligently to plan the 2017 NERA conference, which will be held on October 18-20 at the Trumbull Marriott. The conference theme will be *Using Technology to Advance Education: Challenges and Opportunities*. We invite you to join us to share all the ways that you are doing your part to enhance the educational community. We are very excited to serve the NERA community and continue the excellent tradition and quality NERA members have come to expect.

Member News

Kurt Geisenger was unable to attend the 2016 NERA conference because Penn State University, his doctoral institution, provided him with the Alumni Fellow Award. He was definitely missed at the conference, but his absence was due to a great cause. Please join us in congratulating Kurt on such an honor!

The following articles were recently published by our members:

Matthew Duvall:

Duvall, M. (2016). Evaluating learning technology content with discourse analysis. *Educational Media International*, *53*(4).

Ashley Carpenter:

Jocson, K. M., & Carpenter, A. (2016). Translocal assemblage and the practice of alternative media toward racial justice: A pedagogical perspective. *Critical Studies in Education,* 1-18.

Olcay Yayuz:

Yavuz, O. (2016). Exploring the Impacts of School Reforms on Underrepresented Urban Students' College Persistence. *Educational Research and Evaluation*, 354-373.

Please consider submitting your professional accomplishments for recognition in *The NERA Researcher*!

Submissions may be sent to thenerarесearcher@nera-education.org.

NERA NORTHEASTERN EDUCATIONAL RESEARCH ASSOCIATION

47th Annual Conference Highlights

We would like to thank everyone who contributed to the success of the 2016 NERA conference. Participants attended from all over the country to share their research, chair and discuss sessions, participate in roundtables, collaborate on new projects, and volunteer to help the conference run smoothly. This year's program included 2 pre-conference workshops, 4 in-conference workshops, 13 symposia, 4 invited sessions, 27 individual paper sessions, a robust poster session, 2 roundtable sessions, and 2 information blitz sessions spread across two and a half days.

We were fortunate to have two thought-provoking and engaging speakers during this year's conference. Jonathan Supovitz spoke about the ways in which people perceive and internalize the results of educational research and M.V. Lee Badgett spoke about how to put the results of our research into the hands of decision makers. Their talks were thought-provoking and well received. We also had two plenary sessions where panelists continued the discussion about using educational research to effect change at both the national and local levels.

We would like to thank all of the pre and in-conference workshop facilitators for volunteering their expertise in service of NERA and our members. Thank you to Kristen Huff, Christina Schneider, Jeanne Horst, Heather Harris, Elizabeth Pyburn, Deb Bandalos, Damian Betebenner, Felice Billups, Robert Gable, and Mary Grassetti.

We would also like to thank Thai Ong and the GSIC for organizing two very thought-filled sessions for graduate students during the conference: One on grant writing and another on how to write a curriculum vitae. Thank you, Thai, for your leadership!

In addition to acknowledging the contributors to the program, we would like to thank the many volunteers who shared their valuable time and insight in making this year's conference a success. From reviewing proposals, to discussing and chairing sessions, to those that helped register new and existing members and participants at the registration desk, we appreciated your help.

We would like to offer a special thanks to Madison Holzman and the NERA Communications Committee for their tremendous support leading up to and during the conference. We would also like to thank all of the institutional and personal conference sponsors, as the annual meeting would not have been possible without their contributions.

Finally, we would like to thank each and every one of you for helping to make this year's meeting an engaging and enriching experience. At each conference the members all help to make the conference special for educational researchers from all settings and at all stages in their career.

We look forward to the exciting program that the 2017 conference co-chairs, Daniel Jurich, Jason Kopp, and Whitney Smiley are planning with Craig Wells, the 2017 NERA president.

Molly Faulkner-Bond, Joshua Marland and Scott Monroe
2016 NERA Conference Co-Chairs

NERA NORTHEASTERN EDUCATIONAL RESEARCH ASSOCIATION

2017 NERA Conference Announcement

October 18-20, 2017
Trumbull, Connecticut

Using Technology to Advance Education: Challenges and Opportunities

The rapid expansion and accessibility of innovative technology has altered the way we conduct standard activities and interact with the world around us. For example, some of you may be reading this on your smart phone or tablet, a feat scarcely imaginable two decades ago. Education, as we have seen, also has not been immune to the influence of technology. Perhaps the most noticeable example is how the internet has transformed the accessibility and dissemination of knowledge. In the classroom, tablets have begun to replace textbooks; smartboards have challenged the traditional whiteboard; and social media has been utilized to expand communication beyond classroom walls. Educational assessments have also evolved with those technological developments. Computer and web-based testing have largely supplanted paper and pencil exams. The wide ranging impacts of technology have provided educational researchers and professionals a plethora of new opportunities and challenges. However, the speeds at which these developments have evolved make it difficult for the research to keep up.

Although technology typically appears to provide numerous benefits, the true impact and potential unintended effects are often less understood. We wish to utilize the wealth of experience of our NERA members to further explore issues related to technology in education through the 2017 NERA conference theme: "Using Technology to Advance Education: Challenges and Opportunities".

As the conference co-chairs, we are working diligently to organize an engaging and informative conference that covers a diverse array of topics in educational research, primarily focused on technology. As such, we are excited to announce Virginia 'Ginny' Edwards as our keynote speaker. Ms. Edwards served as the editor of Education Week (http://www.edweek.org) from 1989 to 2016 and led the establishment of Education Week's digital presence. The timeframe and capacity in which she served has provided her a unique opportunity to observe the transformative impact of the technological boom on educational research and policy.

As our theme will focus heavily on innovation, we would be remiss if we kept the conference format static. Therefore, we are considering changing the established procedure of having two formal keynotes. In lieu of a second singular keynote speaker, we plan to hold a debate between leaders in the field regarding a controversial topic affecting the varied interests of NERA members.

In planning for the 2017 conference, we have been evaluating results from the 2016 post-conference survey to guide our decisions. Based on the feedback, we will be looking to incorporate several new session types that will help to diversify the topics covered at NERA. These include additional panel sessions, potential interviews, and the previously mentioned debate session. Respondents also noted a strong desire for continued workshop sessions at the conference. In a change from previous years, we will be soliciting proposals for conference workshops, so that all members interested in presenting have an opportunity for consideration. In line with the conference theme, we also will be improving the conference app based on survey feedback, and looking for other ways we can incorporate technology and innovation in the NERA conference.

Daniel Jurich
National Board of Medical Examiners

Whitney Smiley
SAS

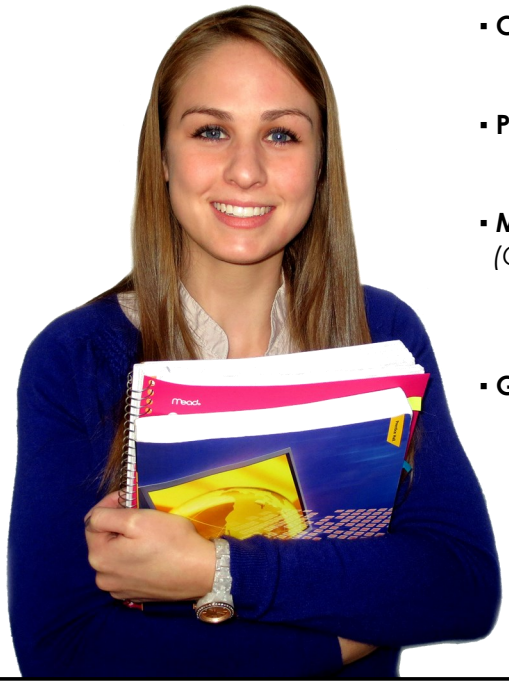
Jason Kopp
American Board of Surgery

Network with NERA members using our LinkedIn group page!

"Like" NERA on Facebook!
https://www.facebook.com/NERAconference

Follow us on Twitter!
@NERAconference

THE CENTER FOR ASSESSMENT & RESEARCH STUDIES



- **Center for Assessment & Research Studies**
  http://www.jmu.edu/assessment
- **Ph.D. Program in Assessment & Measurement**
  http://www.psyc.jmu.edu/assessment/
- **M.A. in Psychological Sciences**
  (Quantitative Concentration)
  http://www.psyc.jmu.edu/psycsciences/quantitativepsyc.html
- **Graduate Certificate in Higher Education Assessment**
  http://www.jmu.edu/outreach/assessment.shtml

MSC 6806
Harrisonburg, VA 22807
assessment@jmu.edu
Phone: 540.568.6706
Fax: 540.568.7878

JAMES MADISON UNIVERSITY

JOHNSON & WALES UNIVERSITY

IMPACT EDUCATION AT JWU

Ed.D. in Educational Leadership

AFTER NEARLY THREE DECADES of teaching Fisher students, JWU alumnus Thomas McGovern '01 became their college president. Was he prepared to lead a bold strategic plan at the historic Boston college? Yes, he says — because of his Educational Leadership Doctorate.

"The education I received at Johnson & Wales University has truly enhanced my ability to define how I can make an impact on one's education."

EDUCATIONAL LEADERSHIP PROGRAM

Center for Research & Evaluation
John Hazen White College of Arts & Sciences

Johnson & Wales University is an active member of the Carnegie Project on the Education Doctorate (CPED), a national effort by 88 institutions to reform and strengthen the education doctorate.

For more information, call 401-598-1838 or visit www.academics.jwu.edu/graduate-studies.

NERA NORTHEASTERN EDUCATIONAL RESEARCH ASSOCIATION

Living in a Post-Validity World: Cleaning Up Our Messick

2016 NERA Presidential Address

Charles A. DePascale

There are periods or events in our historical and personal timelines that are turning points; defining moments that in a significant way change all related events that follow. As we approach the presidential election and consider the events of the last month, one such turning point that comes to mind is the 1960 Kennedy-Nixon debate. The first televised presidential debate, it demonstrated the power of television as a visual medium to impact an election— and the rest, as they say, is history.

Sometimes, forces converge in remarkable ways and multiple defining moments occur within a single year. 1989 was such a year. In November 1989, we watched sections of the Berlin Wall being torn down by jubilant crowds, an iconic moment that signaled the end of the Soviet Union and the Cold War. Just one week later, in my personal timeline, I defended my doctoral dissertation and accepted a full-time job with a testing company. Of course, the following month, Taylor Swift was born in Reading, Pennsylvania. (One of those magical moments whose import and impact are only appreciated in the future.) For our field, however, the most important event of 1989 may have been the publication of the third edition of *Educational Measurement* with its chapter, "Validity", by Samuel Messick.

Validity. Ninety pages that divide people in our field into two categories: people who have never actually completed the entire chapter and people who have pored over it incessantly in search of the deep and hidden meaning. Ninety pages that serve as a defining moment and demarcation in our validity timeline. There are pre–1989 and post–1989 concepts of validity.

Like many defining moments, Validity did not simply appear out of thin air in 1989. It was the culmination of a decade of intense thinking, debate, and shaping of the concept of validity; a decade that itself was the culmination of several decades of discussion of the meaning, importance, and usefulness of the concept of validity. It could be described as an attempt to produce a Grand Unified Theory of Validity. Validity was an attempt to produce a theory which combined measurement, social, practical, and political concerns into a single definition of validity; our very own theory of everything.

We are not alone in our search for a unified theory—a theory that elegantly combines things that we partially understand and accounts for things that we do not really understand at all. Most of us, particularly in our field, are hard-wired to want things to fit together nicely. We yearn for causal connections. We want to be able to understand and explain everything. In particular, we want to be able to explain things that we do not understand and are, in fact, unexplainable. We want to see the face of God and live.

And if measurement (or assessment) is our religion no one can question that we have established validity as our god. Ebel (1961) states "[v]alidity has long been one of the major deities in the pantheon of the psychometrician. Validity is the Alpha and the Omega; literally and figuratively. In the 2014 Standards for Educational and Psychological Testing, the *Standards*, the first word of the first chapter, titled Validity, is Validity; and the final standard on page 213 (Standard 13.9) addresses the validity of an overall interpretation. As with any god, however, our ability as mere mortals to understand validity is limited. Even with Messick descending from the mountaintop, and a Shepard to guide us through his work, we have wandered the desert in search of our Promised Land for 27 years. As zealots and heretics clash through their writings and presentations, what are the consequences of their actions? Are we even sure what validity comprises? Can we communicate the meaning of validity to others?

The *Standards* begin with the following sentence describing the meaning of validity.

> "Validity refers to the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests."

The 21 words in that sentence appear to convey a relatively simple and straightforward thought. Why then have scribes devoted their lives to documenting the sacred thoughts and writings on validity? Why then can a civil discussion of validity quickly descend to the depths of sports talk radio, or worse, political discourse?

Well-intentioned educationists have built careers around interpreting the collective wisdom on validity for our children and policy makers. Yet, we are left with quotes such as these:

Validity is "the most challenging class of the semester" and the hardest for "students to understand." "Despite my best attempts to describe the holy trinity, the unified framework, or argument-based approaches to validity, few students emerge from the class with confidence that they could evaluate validity when developing, using, or even selecting tests." (Gorin)

> "Paramount among these imperfections is that the unitary conceptualization of validity as construct validity is extremely difficult to describe to lay audiences." (Sireci, 2007)

It has been 27 years since the publication of Messick's Validity. That means that most of the people in this room have spent our entire professional careers in a post-Validity world. What has been the impact? To paraphrase that great American philosopher, Ronald Reagan,

- Are we better off now than we were 27 years ago?
- Are we any closer now to understanding and being able to explain validity?

Or more to the point:

- Are we building better tests than we were in 1989?
- Are we making and promoting better interpretations of tests scores than we were in 1989?
- Are we making and promoting better use of tests and testing programs than we were in 1989?

To the extent that we think that 'No' is the appropriate answer to any of those last three questions, to what extent is that outcome an unintended consequence of Validity, and in what ways can a better understanding of validity aid in shifting the answer from 'No' to 'Yes' as we move forward.

In the next section of this address, I will outline what I perceive to be seven challenges to validation that currently exist in the field of K-12 assessment and education. All of the challenges are related in some way to Messick's Validity and our efforts to apply validity theory to an ever-changing and expanding universe of uses of testing. All of the challenges impact our ability to develop tests and implement testing programs; and also impact the ability of educators, policy makers, and the public to use assessment effectively to improve education.

Challenges:

- Our concept of validity is built around constructs and we have none.
- Our treatment of validity creates a false sense of certainty that is inconsistent with our limitations.
- It is no longer clear where, or whether, measurement fits within validity theory.
- A unified theory of validity masks real distinctions between parts of the field that are separate and should be considered separately.
- Validity is described as a never-ending process, but our tests have a shelf life.
- "Validity" becomes an end rather than the means to validity.
- We may have replaced one holy trinity with another.

Our concept of validity is built around constructs and we have none.

By default or design, building our theory of validity around construct validity turns a spotlight on the constructs that we intend to measure or the construct interpretations that we intend to make on the basis of test scores. The *Standards* define construct as "the concept or characteristic that a test is designed to measure." As examples of constructs "currently used in assessment" the *Standards* offer "mathematics achievement, general cognitive ability, racial identity attitudes, depression, and self-esteem." In addition to demonstrating that the scope of the *Standards* is much broader than K-12 educational testing, the examples suggest a concept of construct that is much narrower and well-defined than virtually all current uses of educational assessment. The assessments and testing programs that I deal with on a daily basis are intended to support inferences about school quality, teacher effectiveness, college readiness, career readiness, the politically expedient catch-all college-and-career readiness, and growth. None of those are reflected in the neat examples of constructs included in the *Standards*.

Our struggles with constructs and the influence of those struggles on the way we have framed validity and validation are not new. In 1989, the centerpiece of our field, the SAT, was embroiled in controversy over issues related to interpretation of scores and appropriate uses that are the core of validity. But what construct does the SAT measure? With the SAT, the concepts of test score, interpretation, and use are hopelessly entangled. It would difficult to imagine a worse example of an assessment against which to test this unified theory of validity than the SAT. Well, perhaps one could make an argument for norm-referenced standardized tests as a worse example, but in 1989 those tests were in the midst of their own crisis of identity having been called to task by Dr. John J. Cannell for what became known as the Lake Wobegon Effect.

For the purposes of this address, we will skip over the 1990s —a decade which began with an "Assessment Spring" but ended with the industry turning the sharp blade of Validity against itself. In short, a decade in which we learned the answer to the question, Does Psychometrics Eat Its Young?

The 1990s led us to No Child Left Behind and the current era of assessment and accountability. Since the adoption of No Child Left Behind in 2002, a primary use of K-12 assessment has been school accountability. Although a test may be designed to measure an individual student's mathematics achievement, as Linn (2009) describes, "a key use of assessment results for NCLB is the identification of schools as either making or not making adequate yearly progress." Linn also explains that implicit in the use of assessments to determine whether a school has made adequate yearly progress (AYP) is the assumption that "the differences in observed school-to-school differences in student achievement are due to differences in school quality" and the resulting inference "that a school that makes adequate yearly progress (AYP) is better or more effective than a school that fails to make AYP."

What, then, is our construct—mathematics achievement, the ability to make adequate yearly progress, school quality/effectiveness? Two of those are clearly not the concept or characteristic that mathematics test was designed to measure. So, we may be able to answer 'Yes' to the question are we building better mathematics tests than we were in 1989. It may also be possible to forge a chain that connects us directly from student achievement in mathematics to adequate yearly progress in mathematics. It becomes more difficult to extend that chain to inferences about school quality or effectiveness? And we did not stop at school quality. We added inferences about principal and teacher effectiveness to the chain.

The use of the same mathematics test scores, or metrics derived from those test scores, as indicators of student achievement, teacher effectiveness, principal effectiveness, and school quality is an example of both the complexity and cloudiness of our constructs. Now the chain that we need to forge must often link a single mathematics test score to inferences about students, teachers, principals, and schools. Inevitably, that chain will snap as it is pulled in so many directions. When that chain does snap, we are left with the questions, "What is the construct that we intended to measure?" and "What is our validity argument?"

In reality, are the validity arguments that we are building for basing inferences about teacher effectiveness, principal effectiveness, and school quality on student achievement in mathematics any stronger than the argument, "It is so because we said it is so?

Ebel (1961) shares the "story of three baseball umpires who were discussing their modes of operation and defending their integrity as umpires."

> "I call 'em as I see 'em" said the first. The second replied, "I call 'em as they are." The third said, What I call 'em makes 'em what they are."

Have we become the third umpire?

One would expect that even the most cursory validity study would conclude that defining teacher or school effectiveness in terms of student achievement in mathematics is at best a case of construct under-representation. But is a policy maker or legislator able to wash away validity concerns simply by making the claim fit the evidence?

School quality includes more than mathematics achievement, but without mathematics achievement there cannot be school quality.

Or can policy makers avoid the construct question altogether by returning to our old friend norm-referenced interpretations? When the Obama administration began to issue NCLB waivers in 2012, the accountability focus shifted from AYP and 100% proficiency in all schools to identifying the bottom 5% and 15% of all schools as identified through the accountability system as the schools most in need of assistance. ESSA, the 2015 successor to NCLB and most recent reauthorization of the 1965 Elementary and Secondary Education Act, reinforces this focus on the lowest performing schools. Of course, the shift to norm-referenced interpretations does not totally eliminate the need to consider constructs; and it also raises the question, "Would I design the same test to identify the bottom 5% of schools as I would to measure the effectiveness of all schools?"

Sadly, our success at defining a construct is not much better when we limit our inferences to an individual student's mathematics achievement or further limit our inferences to that student's achievement of the grade 4 Common Core State Standards (CCSS) in mathematics. Studies across textbooks, school curricula, and various assessments "aligned to the CCSS" will contain some areas of overlap, but no consensus of the concept or characteristic of student achievement of the grade 4 CCSS. The result, as Madaus, Russell, and Higgins (2010) describe is that in most cases it is the assessment that defines rather than measures the construct "grade 4 mathematics achievement"; and that is an issue that raises additional challenges.

Our treatment of validity creates a false sense of certainty inconsistent with our limitations.

The validity of an interpretation cannot be established by a research monograph or detailed manual. The aim for the report is to advance sensible discussion. Why should we wish for more? On matters before the public, the evidence usually is clouded. The institutions of the polity are geared to weigh up reasonable, partly persuasive, disputed arguments; and they can be tolerant when we acknowledge uncertainties. The more we learn, and the franker we are with ourselves and our clientele, the more valid the use of tests will become. (Cronbach, p. 107)

The description of validity in the *Standards* may, in fact, be consistent with the cautionary conclusion to Cronbach's (1980) *Validity on Parole: How Can We Go Straight?* presented above. The words in the *Standards* and our use of them, however, do not convey the same sense of uncertainty and humility. It is difficult to draw a direct connection between Cronbach's advice and the testing policies of NCLB and ESSA. It is much easier to draw a connection between those policies and the *Standards* and other writings that emphasize the need for assessments, assessment programs, or assessment uses that are valid, reliable, and, now, fair.

We like to present assessment and validity with technical quality terminology that connotes precision and truth. We prefer terms like measurement and reliability to terms such as estimation and probability. As assessment practitioners, we generally do not like to acknowledge the existence of alternative outcomes, explanations, or models even to the same extent as local meteorologists. This is not to suggest that we go to the other extreme and only provide a litany of alternatives as equally likely or plausible outcomes. Rather, we should strive to find the level of uncertainty that makes the best use of our professional expertise and allows policy makers and educators to effectively and appropriately integrate evidence from assessments into

their decision making.

It is no longer clear where, or whether, measurement fits within validity theory.

Validity is an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment. As will be delineated shortly, the term test score is used generically here in its broadest sense to mean any observed consistency, not just on tests as ordinarily conceived but on any means of observing or documenting consistent behaviors or attributes. (Validity, p. 13)

In the beginning, if we viewed tests as measurement instruments and testing as a measurement process, it was easy to portray validity as a measurement issue, or at least a technical issue. From the opening paragraph of Validity presented above, however, the role of measurement in validity and the relationship between measurement and validity becomes less clear. As we now conceive of tests and testing programs as accountability systems, the connection is even murkier.

We can all accept that we moved on from defining validity and validation as an attempt to answer the question "Does the test measure what it purports to measure?" But, where did moving on take us?

Shepard (1993) offered "Does the test do what it claims to do?" as "a more appropriate question to guide validity investigations." One can easily see the connection between Shepard's question and Braun's (2008) argument regarding the validity of accountability systems:

- An accountability system is imposed or implemented with the intent that it will accomplish its desired goals (e.g., increased student achievement).
- The mechanism by which this will happen is called the "theory-of-action".
- In validating an accountability system, the theory-of-action plays the same role as does the construct in test validation.
- Thus, consequential validity is the ultimate criterion by which we should judge an accountability system.

From Braun's argument, it is then a very short leap to the argument that Andrew Ho presented to this conference in 2014. Ho argued that in one sense we need not concern ourselves with the inner workings or technical quality of the accountability system at all. From an experimental design perspective,

We can view the accountability system (or testing program) as a treatment that has been implemented with the intended goal of increasing student achievement.

We design an experiment to implement the desired treatments or perhaps alternate treatments for a period of time.

At the end of the experiment, we compare student learning in the treatment groups to a control group.

Based on the outcome of the experiment, we can make an evaluative judgment about the impact of the accountability system on student achievement. And if we apply Braun's argument about the validity of accountability systems, we can evaluate the validity of the system based on the results of the experiment.

It is difficult to argue that Ho's experimental approach does not answer the Shepard's question "Does the test do what it claims to do?" But it is difficult to find a connection to measurement and it is difficult to find the sources of validity evidence delineated in the *Standards*.

In reality, however, perhaps the distinctions between Ho's experimental approach and the *Standards*, or an argument-based approach to validity are not as great as they seem. We have taken a very high-level view of Ho's experiment. When one starts to actually design the experiment and account for all of the relevant variables, perhaps the experimental design process will produce something that looks very much like the collection of evidence that would result from an argument-based approach to validation. And perhaps a close focus on the technical quality of the measures or indicators within the system will result in better outcomes; that is, will result in stronger claims of validity. Nonetheless, it could be interesting to play out the consequences of considering validity from a non-measurement perspective.

A unified theory of validity masks real distinctions between parts of the field that are separate and should be considered separately.

The beauty of the unified theory in Validity is that it makes it clear that given the complexity of our constructs, one type of validity evidence is likely insufficient to make a strong validity argument. It is likely that evidence related to the construct, criterion, and content will all be necessary to some degree to establish a strong validity argument. To the extent that the construct requires interpretation of the test scores for an intended use, there will also be a need for consequential evidence.

On the other hand, the danger with a unified theory is that by shifting the focus from the various types of evidence to an overarching concept of validity it may make it easier for some people to think that unified means simple. No, it is not possible to reach that conclusion if you have read Validity or tried to craft a validity argument or theory-of-action, but who is responsible for those tasks. Who is now responsible for compiling validity evidence?

Shepard (1993) warns of the "danger that test makers will defer to test users to evaluate intended applications" of their tests and that "[o]ften, users are not qualified, or lack the neces-

sary resources, to conduct validity investigations." She explains that "this separation of responsibility would allow test makers to study only the "scientific" meaning of the test interpretation… leaving it to the user to evaluate the test for its intended purpose" (p. 445). Having had the opportunity to participate in countless meetings with test makers, test users, policy makers, etc. over the last 27 years, I can attest that Shepard was right to be concerned.

The issue, however, may be more complex than Shepard described in 1993. With the rise of customized state assessment, the line between test maker and test user has become much less clear. To what extent is the state Department of Education the test maker and to what extent is the Department the test user? In the complex contractual relationships that exist, the testing companies, traditionally considered the test maker, have been contracted to perform specific tasks within the test making process. Is it their responsibility to go beyond those tasks or to point out the need for someone to go beyond those tasks? That is, will they feel a responsibility to study even what Shepard describes as the scientific meaning of the test interpretation?

There are also questions of conflict of interest in assembling and evaluating validity evidence. George Madaus has been a long-standing advocate for an independent board to evaluate validity. The U.S. Department of Education has established a peer review process to evaluate validity evidence pertaining to state assessment programs. However, there may still be conflicts or implicit bias in what evidence is collected and/or how it is collected.

The unified theory in Validity revealed the complex and multi-faceted nature of validity that always existed. Training test users to recognize and deal with that complexity and finding the resources to allow them to do so, however, remains a daunting task.

Validity is described as a never-ending process, but our tests have a shelf life.

Inevitably, then, validity is an evolving property and validation is a continuing process. Because evidence is always incomplete, validation is essentially a matter of making the most reasonable case to guide both the current use of the test and current research to advance understanding of what the test scores mean. (Validity, p. 13)

From the beginning, one of the anticipated consequences of a unified theory of validity was that it was too big and too complex, effectively absolving individuals from the responsibility of a rigorous process to collect evidence of validity. As Shepard (1993) warned

Finally, the complexity of Messick's model and chapter creates the same difficulty as nearly every other treatise on construct validity before his. Each emphasizes that construct validation is a never-ending process, because there are so many hypotheses to be tested across so many settings and populations and because the generalizability of findings decays over time. While the never-concluding nature of construct validation is a truism, the sense that the task is insurmountable allow practitioners to think that a little bit of evidence of whatever type will suffice. (p. 429)

Exacerbating the dilemma that validation is a never-ending process is our current view of tests as disposable or ephemeral. Historically, we viewed a test as a product that was carefully constructed over the course of several years and intended to be used for several years as a single form or a set of interchangeable parallel forms. As states shifted to custom test development, it was not uncommon for 25%, 50%, of even 100% of the items on the operational test form to be released after each test administration. A new "test" was developed and administered each year. As we move from paper-and-pencil fixed form tests to computer adaptive testing, a test may be a unique event administered only one time to an individual student. How do we reconcile this view of a test with a never-ending process of validation?

The logical answer, of course, is to reorient the validation discussion to the testing program rather than to a particular test form. Unfortunately, the life cycle of many state assessment programs has become so short that such a focus gains us little.

"Validity" becomes an end rather than the means to validity.

One of the dangers inherent in an unwillingness to accept uncertainty is that we corrupt the concept of validity. In 2014, Michael Kane addressed this conference on the importance of asking the right questions. One of the things that he cautioned against was our tendency to reduce complex situations to questions that we can answer. In validity terms, limiting ourselves to questions that can be answered results in construct underrepresentation. When evaluating a test, we would be quick to identify and point out a lack of alignment between the test items and either the content or cognitive processes that the test is designed to measure. At the same time, when compiling evidence of validity, it is so easy to skim through a list of 100-200 proposed validity analyses and select the studies that can be done quickly and easily (preferably with available data) and avoid those that would require additional resources.

As suggested in my previous discussion of constructs, a more serious threat to validity would be reframing our validity questions until those questions match our available evidence. One could think of that as the equivalent of enhancing reliability at the cost of validity by eliminating test items simply on the basis on low inter-item correlations. In the end, we may have a high coefficient alpha or generalizability coefficient, but in reality, we have neither reliability nor validity.

Alternatively, one could consider the pressure to achieve validity in the same manner that we view high-stakes testing and the pressure to achieve a certain score. When the test score itself becomes the focus rather than the knowledge and skills

that the score is supposed to reflect, there is a perverse incentive for test administrators or test takers to engage in test preparation activities or test taking practices that are likely to reduce validity. The same can be true for those charged with designing and implementing a validity study.

Note that there is a critical difference between the behaviors described above and reframing or limiting claims about test scores to only those interpretations and uses of test scores which can be supported by evidence. The former reflects an intention to distort validity, while the latter reflects an intention to promote validity.

We may have replaced one holy trinity with another

A primary purpose and accomplishment of Messick's Validity was to solidify the view of validity as a unitary concept. Validity is regarded as "the most fundamental consideration" in developing tests and evaluating the interpretation of test scores for proposed uses. A unified theory of validity, therefore, strengthens educational testing.

To a large extent, however, we never really succeeded in establishing validity as a unitary or preeminent concept with the public; and to a certain extent we may have failed to do so within the field. Whether in textbooks, research papers, or laws, it is rare to see the term *valid* without its sidekick *reliable* close by its side. For example, in the Every Student Succeeds Act, the terms valid and reliable appear together in some form at least nineteen times. This invites the question of whether one can have validity without reliability; and does little to suggest the tradeoffs that are inevitable when considering both validity and reliability. Also, as we have seen on multiple occasions over the last two decades, pairing validity and reliability may have the unintended consequence of placing an inordinate emphasis on reliability at the expense of validity.

The publication of the 2014 *Standards* further complicates the issue by adding "Fairness" to the mix by isolating Validity, Reliability, and Fairness as the foundations of educational and psychological testing. Therefore, the question now becomes, "Can one have validity without reliability or without fairness?" Like Reliability, the *Standards* position Fairness as "a fundamental validity issue" to be considered throughout the testing process.

In practical terms, have we simply recreated a Trinitarian model, replacing the holy trinity of "content, criterion, and construct validity" with our new trinity of "Validity, Reliability, and Fairness"? Will the special treatment afforded to Reliability and Fairness enhance or detract from the process of establishing validity?

Where do we go from here?

If those are the challenges that we face, where do we go from here as psychometricians, researchers, academicians, teacher educators and K-12 educators to ensure that we can confidently answer 'Yes' to the three questions I posed at the beginning of the address:

- Are we building better tests than we were in 1989?
- Are we making and promoting better interpretations of tests scores than we were in 1989?
- Are we making and promoting better use of tests and testing programs than we were in 1989?

We need to explicitly broaden the conversation about validity from tests and test scores.

As a starting point, we must keep our focus on the simple idea expressed at the beginning of the *Standards*:

> Validity refers to the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests.

From there, I believe that there are five areas in which we can improve our practice:

- We need to consider the user.
- We need to build an appreciation for and expectation of uncertainty.
- We need a shared understanding of the basic requirements to call something a construct.
- We need to figure out the role of the test in validity.
- We need to explicitly broaden the conversation about validity from tests and test scores.

We need to consider the user.

Cronbach (1980) stated "our task is not to judge *for* nonprofessionals but to clarify, so that in their judgments they can use their power perceptively." When communicating information about tests or testing programs we need to consider the users:

- What do they need to know to be able to decide whether to select a test?
- What do they need to know to be able to use test information appropriately?
- What do certain users need to know to be able to build strong validity arguments?

We need to develop solid examples or models of validity arguments for tests and testing programs designed for a particular purpose. These models should not be exhaustive validity plans that account for every possible source of evidence and potential use covered in the *Standards*. They need to be tailored

to specific needs of the users and demonstrate acceptable, reasonable validity arguments, not ideal gold standard collections of evidence that are beyond the reach of the typical test user.

We need to build an appreciation for and expectation of uncertainty.

Although most of our professional and academic writing is replete with caveats and cautious conclusions, we tend to convey a sense of precision and certainty when we couch our discussions of validity and reliability in measurement jargon. We need to better communicate that validity is a matter of degree and that validation is an ongoing process based on the application of best practices and the collection of evidence. We need to be clear that often evidence will be incomplete, while avoiding the impression that it is impossible to make an informed decision. We must accept that virtually everything that we do with tests is built around providing information to inform professional judgment.

Cronbach (1980) correctly points out that "the courts are properly impatient when asked to take seminar-room abstractions as a basis for settling concrete cases." The same is true of policy makers and the general public. All of those parties, however, are able to make informed decisions based on evidence if they know that is what they are being called upon to do.

We need a shared understanding of the basic requirements to call something a construct.

We need to be honest in communicating about constructs and our ability to "measure" them. I believe that there are two important factors to consider with regard to what we call constructs in K–12 education and how we describe the tests and test scores designed to "measure" those constructs.

First, I believe that the constructs that we are assessing must exist and be observable (at least indirectly) in the real world; that is, outside of the test. Borsboom, Mellenbergh, and van Heerden (2004) offered what they described as an "exceedingly simple" argument defining the "very basic concept" of validity:

If something does not exist, then one cannot measure it.

If it exists but does not causally produce variations in the outcomes of the measurement procedure, then one is either measuring nothing at all or something different altogether.

Thus, a test is valid for measuring an attribute if and only if (a) the attribute exists and (b) variations in the attribute causally produce variations in the outcomes of the measurement procedure. (p. 1061)

Whether one agrees with Boorsboom et al. that the argument above is sufficient to define validity, I argue that for virtually all of our intended uses of K–12 assessment, we must be able to demonstrate that our tests are measuring something that exists, is observable, and can be described even without the test. That is, there must be other means in addition to the test for identifying students who are proficient in mathematics, teachers who are effective instructors, or high-quality schools. Only if we accept that requirement will it be possible to evaluate whether differences in the construct "causally produce variations in the outcomes of the measurement procedure." If we do not require the existence of such external evidence then we are at too great a risk of the test score becoming the construct, which renders our claims and interpretations nothing more than tautologies, making validation impossible.

Second, we have to recognize, appreciate, and communicate the complexity of the constructs that we are attempting to measure and our limitations in measuring them. In her 2012 presidential address to this conference, Lynn Shelley shared "The Rat Story" and provided a description of the complex interactions of personality, social, and emotional characteristics that impact children's nonacademic and academic success; as well as the importance of understanding basic child development and brain development in creating learning environments and interpreting student performance. Mislevy (2016), in an article in the current edition of the *Journal of Educational Measurement* describes just how complex our constructs will become if we attempt to fully account for all of the forces influencing a student's performance on a complex task measuring higher-order skills. In our claims and in our validation processes we have to find a balance between the inherent complexity that is reality and the uselessness of having to create validity arguments that are so conditional and so unique that they apply only to an individual student under certain conditions at a given point in time.

We need to figure out the role of the test in validity.

Yes, we have moved beyond the notion that validity can be determined simply by answering the question, "Does the test measure what it purports to measure?" But under what conditions is it important that the test does actually measure what it purports to measure and that we understand what it measures?

Should we simply assume that the test measuring what it purports to measure is a necessary, but not sufficient component of the validity argument? Are there situations or conditions when that assumption is false?

The first question in my list of three questions asked whether we are building better tests than we were in 1989. On one hand, progress in the development of content standards, performance level descriptors, the use of information from IRT models throughout the development process, and the application of processes such as Evidence-Centered Design would support a clear answer of 'Yes' to that question. On the other hand, the shift in focus to producing student-level scores on most state assessments has reduced the number of questions administered to assess school performance within a content area from several hundred to 40–50.

We have established that an excellent test of student achievement in mathematics might be a poor indicator of school

quality. However, we need to be clear about our expectations for that test as a measure of mathematics achievement; we need to understand how the technical quality of the test impacts our ultimate judgment of school quality; and we need to be clear about who is responsible for providing the evidence to support the validity argument related to its use as a measure of mathematics achievement.

We need to explicitly broaden the conversation about validity from tests and test scores

Historically, we have spoken about tests and test scores when discussing validity or validation. Our reality, however, is that most of the validity issues that users are addressing today do not involve a simple test and test score. They involve results from a testing program which administers multiple forms of a test over multiple years or within a single year; and they involve results from accountability systems that comprise multiple indicators which often are metrics derived from one or more test scores. In many cases, the same test score is the basis for multiple indicators. Our language surrounding validity and validation must reflect that reality.

Looking ahead to the very near future, our concept about validity will have to expand to include interpretations, inferences, and decisions based on data mining and learning analytics. As past president April Zenisky demonstrated for us in her 2015 presidential address the amount of data that we have access to and the uses to which that data can be applied are growing exponentially. We cannot wait for the next revision of the *Standards* to determine how to apply validity theory to those sources of information.

Conclusion

As I mentioned in the opening of this address, in my personal timeline 1989 was the year that I defended my doctoral dissertation at the University of Minnesota and entered the assessment profession. Although I have spent most of my professional career working with assessments, I think that it is important to note that my education and training at the University of Minnesota established a broad foundation in educational research and program evaluation as well as in theoretical and applied educational measurement. My area of specialization within educational psychology was measurement and evaluation—with a heavy emphasis on program evaluation. My advisor, John Stecklein, was a leader in the field of institutional research. One of my mentors, Stan Deno, worked in the special education department and helped place me in a school-based research assistantship. I believe that background combined with my experiences working with policymakers is what enables me to take a holistic view of validity while still appreciating the importance of the details of the individual components that validity comprises.

Ebel (1961) had this to say about validity: "It is universally praised, but the good works done in its name are remarkably few. Test validation, in fact, is widely regarded as the least satisfactory aspect of test development." Cronbach titled his address to the 1979 ETS conference on new directions in assessment "Validity on Parole: How Can We Go Straight?" Newton and Shaw (2015) questioned whether validation would be better served by retiring the word validity altogether. I am optimistic, however, that the future of validity is bright. In the last few years the focus of the Assessment industrial Complex in which we operate has finally shifted its attention to the classroom—to teachers and students. As we move forward, I am confident that if we keep the focus on helping professionals to gather sufficient evidence to support their instructional and policy decisions that validity will be fine.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological review*, *111*(4), 1061.

Braun, H. (2008, September). Vicissitudes of the validators. Paper presented at the Reidy Interactive Lecture Series, Portsmouth, NH

Cronbach, L. J. (1980). Validity on parole: How can we go straight. *New directions for testing and measurement*, *5*(1), 99-108.

Ebel, R. L. (1961). Must all tests be valid? *American Psychologist*, *16*(10), 640.

Ho, A. (2014, October). *Made to be broken: The paradox of student growth prediction.* Invited address at the annual conference of the Northeastern Educational Research Association, Trumbull, CT.

Kane, M. (2014, October). *Asking the right questions.* Invited address at the annual conference of the Northeastern Educational Research Association, Trumbull, CT.

Linn, R. L. (2009). The concept of validity in the context of NCLB. *The concept of validity: Revisions, new directions and applications*, 195-212.

Madaus, G. F., Russell, M. K., & Higgins, J. (2009). *The paradoxes of high stakes testing: How they affect students, their parents, teachers, principals, schools, and society*. IAP.

Messick, S. (1987). Validity. *ETS Research Report Series*, *1987*(2), i-208. And in R. L. Linn (ed.). *Educational Measurement, 3rd edition,* 1988.

Mislevy, R. J. (2016). How Developments in Psychology and Technology Challenge Validity Argumentation. *Journal of Educational Measurement*, *53*(3), 265-292.

Newton, P. E., & Shaw, S. D. (2015). Disagreement over the best way to use the word 'validity' and options for reaching consensus. *Assessment in Education: Principles, Policy & Practice*, 1-19.

Shelley, L. M. (2012, October). *Neuroplasticity, stress, and resilience.* Presidential address at the annual conference of the Northeastern Educational Research Association. *The NERA Researcher,* Winter, pp. 5-8.

Shepard, L. A. (1993). Evaluating test validity. *Review of research in education*, *19*, 405-450.

Sireci, S. G. (2007). On validity theory and test validation. *Educational Researcher*, *36*(8), 477-481.

Zenisky, A. (2015, October). *Choose your own [data] adventure: perils and pitfalls, and lots of promise.* Presidential address at the annual conference of the Northeastern Educational Research Association, *The NERA Researcher,* Winter, p. 11-14.

NERA
NORTHEASTERN EDUCATIONAL RESEARCH ASSOCIATION

ACS VENTURES

HIGHLANDER INSTITUTE
Leveling the Field for All Learners

NERA NORTHEASTERN EDUCATIONAL RESEARCH ASSOCIATION

NERA 2016 Conference Proceedings — Call for Paper Submissions

Any paper presented at the NERA 2016 conference is eligible for inclusion in the NERA Conference Proceedings, including full papers presented in the poster session. The Proceedings are hosted and archived on DigitalCommons@Uconn (http://digitalcommons.uconn.edu/nera/) and linked to the NERA website (http://www.nera-education.org/). Each paper has a dedicated, permanent URL.

Javarro Russell is receiving submissions for the 2016 Conference Proceedings.
E-mail neraconferenceproceedings@gmail.com.

DUE DATE: Papers must be submitted electronically by February 15. The 2016 Conference Proceedings will be available on-line by April 1, 2017.

Submission Instructions

To submit your article for inclusion in the NERA Conference Proceedings, please follow these instructions:

Sign the Article Submission Agreement:
You must sign the Article Submission Agreement form indicating that you hold copyright to your paper. Article Submission Agreement forms are available on the NERA website. A scanned copy of the original signed form must be submitted along with your paper to neraconferenceproceedings@gmail.com.

A paper cannot be posted without the signed article submission agreement.

Submit your paper (electronically only) to: neraconferenceproceedings@gmail.com

1. In the subject field of the email:
   Name of first author (last, first)
   Conference Session Number of the paper to be published (i.e. 2.7 from the conference program)

2. In the body of the email please include:
   Title of paper exactly as you wish it to appear in the Proceedings (capitalize all words except prepositions, articles, etc.)
   First Author name, institution, and email address
   Second authors, institutions, and email addresses
   Keywords— identify up to five keywords (list in lowercase, separated by commas)
   Abstract of the paper (100 word limit). Please include this in the body of the email (not as an attachment). You may also include the abstract as the first section of your paper submission.

3. Attach the paper to the email:
   Attach your paper submission as a WORD document (in .doc format; not docx).
   We recommend that you use 12 pt. Times Roman font, double-spaced, APA style. Be sure that you have turned off track-changes and eliminated any comments or other review markings on the document.
   The paper will be uploaded exactly as you submit it. The editor will not do any formatting or editing of the paper (please go online to view examples).

Continued on page 18

NERA NORTHEASTERN EDUCATIONAL RESEARCH ASSOCIATION

NERA 2016 Conference Proceedings

Article Submission Agreement

E-mail scanned copy of signed agreement to neraconferenceproceedings@gmail.com

Paper Title: _____

_____

First Author Information:

Name: _____ Affiliation: _____

Address: _____

_____

E-mail Address: _____

I hold the copyright to this document and grant permission for posting it in the DigitalCommons@UConn institutional repository, or have the approval of the copyright holder to do so, and agree that it may be made available to the public in any format in perpetuity. I warrant that the posting of the work does not infringe upon any copyright, nor violate any proprietary rights, nor contain any libelous matter, nor invade the privacy of any third person or third party, nor otherwise violate Repository policies. I understand that accepted papers may be posted immediately as submitted.

_____
First Author Signature (date)

Signatures of second authors (if applicable):

_____
Second Author Signature (date)

_____
Second Author Signature (date)

_____
Second Author Signature (date)

_____
Second Author Signature (date)

Update from the Co-Chairs of the NERA Mentorship Program

Juliette Lyons-Thomas and Jonathan Rubright

The Mentoring Program's focus over the past several years has been on the broad development of new and early career NERA members, including work on the research, academic and professional development fronts. This year, the Program had a record-setting 25 mentor/mentee pairs, maintaining the growth of the program since its redesign in 2013. The Mentoring Program is focused on improving not just the quantity, but the quality of mentoring that occurs during, and outside of, the annual conference.

The Co-Chairs would like to thank the mentors for their willingness to share their time, energy, and experiences with fellow NERA members. We know that this work is time consuming, and know that you benefitted from sharing your experience with others.

Perhaps more importantly, we thank the mentees for their willingness to reach out and ask for mentorship. "Reaching out" is not something any of us are taught to do: asking for mentoring requires the courage to ask for help and the humility to admit there are things you do not know. We applaud mentees for their openness to ask for help, and are grateful to be part of a professional community that is safe and supportive enough for individuals to ask for, and receive, mentoring from fellow members.

Update from the Infrastructure Committee

Tia Sukin

We'd love to hear from you. While we make improvements every year—to the NERA website, conference registration, and membership services—we know we can get better. Please feel free to contact tsukin@pacificmetrics.com with your ideas.

In 2016, the Infrastructure Committee worked to improve the performance of conference management tools in various ways that benefited the conference committee co-chairs. We do hope that some of our changes resulted in fewer program schedule conflicts for our valued volunteers that both present their research and discuss the research of their peers. We also improved the membership and conference registration tools by allowing for simultaneous registration and membership renewal.

Automation of the NERA conference program has been problematic to date and requires more manual labor than desirable. In 2017, the Infrastructure Committee is determined to develop a process that reduces this labor-intensive task.

NERA NORTHEASTERN EDUCATIONAL RESEARCH ASSOCIATION

Update from NERA Conference Ambassadors

Rochelle Michel

Since 2014, the NERA Conference Ambassadors committee has been working behind the scenes to support the annual NERA conference. The committee is one of NERA's newer ad-hoc committees. The Conference Ambassadors are responsible for welcoming new and returning members to the annual NERA conference by encouraging networking between colleagues across the NERA membership, specifically those in attendance at the conference. As part of our role, we also share highlights from our own experiences as presenters, session chairs, discussants, elected officers, and meeting attendees. If you are considering a way in which you can serve NERA and you have attended the annual conference for at least two years, consider being a Conference Ambassador. We support both the Membership Committee and the Conference Co-Chairs, as they are integral in making each annual conference a success. Members interested in joining the NERA Conference Ambassadors committee or contributing ideas about how the Conference Ambassadors can further support the open and collegial environment found at the annual NERA conference can contact me, Rochelle Michel, at rmichel@ets.org.

We need you!

Rochelle Michel
Chair, Conference Ambassadors Committee

Call for *Researcher* Content Co-Editor!

What is this position?

The Content Co-Editor position is a three-year commitment that entails communicating with the NERA leadership and members, gathering all materials, and reviewing/editing the content in each issue.

The person in this role will take over Haifa Matos-Elefonte's responsibilities and will work with Katherine Reynolds, the Production & Design Co-Editor.

What skills are required?

The most essential skills that this position requires are:

- Great attention to detail
- Good organization skills
- Excellent time-management skills

If you are interested in this position or you know someone who might be, please get in touch with Haifa at hmatoselefonte@collegeboard.org or Katherine at Reynolds.katherine@bc.edu. Note that the new Content Editor will take over the newsletter responsibilities with the Spring 2017 Issue of *The NERA Researcher*.

NERA NORTHEASTERN EDUCATIONAL RESEARCH ASSOCIATION

The Journal of Educational Leadership and Policy Studies (JELPS) ISSN#: 2473-2826

Manuscript Deadline: July 1, 2017
www.southernct.edu/jelps/

The Journal of Educational Leadership and Policy Studies (JELPS) ISSN#: 2473-2826, sponsored by Southern Connecticut State University is a peer-reviewed electronic journal dedicated to establishing a global network that will serve as a platform for researchers, policymakers, educators and school leaders who are concerned with moving educational issues forward. More importantly, the journal will provide readers with an enhanced awareness of strategies and policies for improving educational outcomes and methods for improving school success for all students.

JELPS seeks manuscripts that address best practices and school policies in the four leadership domains: Instructional Leadership, Talent Management, Organizational Systems, School Culture and School Climate. The journal welcomes articles based on practice and research with an explicit educational leadership context, or components that examine the function of school leadership from a variety of well-balanced perspectives. All theoretical and methodological approaches are welcome (e.g., qualitative vs. quantitative; empirical vs. conceptual; discipline-based vs. interdisciplinary).

Authors will receive initial review decision notifications within 4 to 6 weeks.

Please review the Call for Articles at https://go.southernct.edu/jelps/#call-for-articles.

JELPS is also accepting papers for the Special Issue on The Gordon Paradigm of Inquiry and Practice. Manuscript Special Issue Deadline: July 1, 2017

Manuscripts should be sent to Dr. Olcay Yavuz, Editor-in Chief, Journal of Educational Leadership and Policy Studies for initial review at yavuzo1@southernct.edu.

For questions and inquiries about the submission, contact JELPS.

NERA NORTHEASTERN EDUCATIONAL RESEARCH ASSOCIATION

NERA NORTHEASTERN EDUCATIONAL RESEARCH ASSOCIATION

Call for Chapter Proposals: Disruptive Views of Gender and Sexuality in K-12 and Teacher Education

Volume Editors:
Adrian D. Martin, New Jersey City University
Kathryn J. Strom, California State University, East Bay

Past research on gender and sexuality in education and teacher education has primarily focused on identifying ways of fostering inclusive and affirmative school communities for non-cis and/or queer students who identify outside the heteronormative matrix. Much of this work has attended to theorizing pedagogies and curricula conducive towards these aims. Yet despite social progress in the legality of same-sex marriage, and considerable attention in mainstream media on issues relevant to the transgender community, non-cis and queer individuals (especially those of color) continue to experience violence, face housing discrimination, employment discrimination, and the denial of service in public businesses. In light of the growing conservative movement to not only roll back legal advances for non-heterosexual individuals, but to also promote a culture of homophobia, transphobia, and heterosexuality as normative, natural, and desirable, we argue that scholars must attend to the myriad ways in which members of the school community can counter such efforts, and how the multiple facets of the educative experience can be conceptualized beyond a heteronormative paradigm. Rethinking what it means to be queer and/or non-cis moves us beyond safe and inclusive school environments towards discourse, research, and political action to transform school communities and society at large. In so doing, we decenter heteronormativity and in turn facilitate the human capacity to identify and enact gender expression as desired rather than as socially prescribed.

Troubling heteronormative logic and static gender categories requires non-hierarchical, multiplistic frameworks that break with the reductionist thinking patterns that have reinforced gender binaries and fixed sexual identity categories. Thus, this volume seeks to explore the themes referenced above through post-structural philosophical frameworks, queer theory, and other emergent 21st century non-linear theories of human phenomena and existence. These perspectives are conducive to not only examine dominant discursive and material conditions that striate epistemological interpretations of gender and sexuality, but to also rethink what it means to be gendered, the construction of sexual categories, and the role of schools and education as perpetuating (or countering) such meaning-making processes. We envision chapters that will draw upon the work of Deleuze and Guattari, Jacques Derrida, Julia Kristeva, Michel Foucault, Luce Irigaray, Karen Barad, Bruno Latour and others to critically investigate gender and sexuality in education and teacher education in the 21st century. Our aim is that these chapters will reflect the productiveness of these bodies of thought on research on the body itself, the body in public institutions of schooling, connections between one body and another, and entanglements between bodies/matter and discursive elements. Anticipated publication date is 2018. We are interested in both conceptual and empirical manuscripts.

Continued on page 25

Possible chapter topics might include the following:

- Enacting non-cis identities in schools
- Engaging students (P-16, graduate or doctoral) on issues relevant to trans/homophobia and
- heteronormativity
- Teacher education as reproducing/countering heteronormativity
- Examining the intersectionality of identities (trans/queer individuals of color, diverse
- cultural/linguistic backgrounds, non-documented people, lower/working class individuals)
- Narratives of trans/queer students, teachers, school leaders or caregivers
- Emergent conceptualizations on gender and sexuality and the connections between these and
- schooling
- Trans/queer individuals in early childhood education, higher education, graduate/doctoral studies
- Implications of the Orlando massacre for teacher educators, scholars, and members of the school community

We invite 1000-1500 word proposals that address the following:

- Purpose/objectives and context
- Theoretical approach
- Summary of study methods
- Findings/Understandings
- Methodological/theoretical and practical significance

Timeline:

- Initial proposals are due to the editors by February 17th, 2016.
- Authors of selected proposals will be invited to develop a chapter by March 17th, 2017 and will
- be asked to peer review one full chapter manuscript.
- Chapters (5000-6000 words, APA) are due from authors by July 1st, 2017.
- Anticipated publication is 2018.

Manuscripts should be submitted to amartin6@njcu.edu.

NERA
NORTHEASTERN EDUCATIONAL RESEARCH ASSOCIATION

FORDHAM UNIVERSITY
GRADUATE SCHOOL OF EDUCATION

Southern Connecticut State University
SCSU
School of Education

BUROS
CENTER FOR TESTING

A Special Note from Phil Archer

Dear Friends,

Some of you will remember that I used to do the recapitulation, a fine talk after lunch on the last day. I stopped doing it about 14 years ago. I am writing this article in remembrance of Dick Clark, the man who did the same talk for the 25 years before me. Dick was one of my professors at SUNY-Albany. He could usually theme his talk around something having to do with the upcoming election. When I was asked to replace Dick, I realized that was not something I could do. So I decided to make up quotes and pretend NERA members said them.

Dick Clark passed away last year and a memorial service was held for him in the Albany area. Mary Horan and I went to his memorial service. I said a few words toward the end of the service, wearing a T-shirt that said, "Statistics means never having to say you're certain."

Dick was a past NERA president and an active member. Another NERA president that passed away last year was Ruben Reusch. Ruben was my advisor at SUNY-Albany. Ruben was also past president of NEERO and active in both organizations.

At the last NERA meeting, we honored Roscoe Brown, past president who was a fighter pilot with the Tuskegee Airmen during World War Two.

I think I would like Scott Brown to do my obituary because Scott may make me look good.

Phil Archer

The Graduate Lounge

Elisabeth Pyburn
James Madison University

Our Mission: The mission of the Graduate Student Issues Committee (GSIC) is to support the involvement and professional development of NERA graduate student members and to reach out to new graduate students in an effort to increase the diversity of institutions represented at NERA.

GISC News

Thanks to all the NERA members who helped make our graduate student sessions at this year's NERA conference a success! We sponsored two sessions this past year—one on grant writing and one on crafting a curriculum vita. We also hosted the annual graduate student social, which was well attended this year.

Currently, GSIC members are voting on new members and collecting submissions for the NERA Best Paper by a Graduate Student Award. We look forward to another great year.

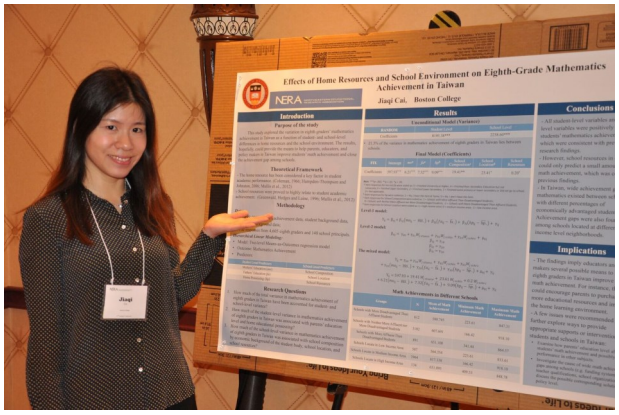
GSIC Call for New Members:

Serving on the GSIC is a great way to get involved with NERA and build relationships with other graduate students! Responsibilities include collaborating with students from various institutions to plan GSIC sponsored in-conference sessions, and the GSIC student social. New members are selected each year after the NERA Conference.

For more information on how to apply and get involved, please contact neragraduatestudents@gmail.com.

Like us on Facebook! **https://www.facebook.com/neragsic**


CPRE CONSORTIUM FOR POLICY RESEARCH IN EDUCATION | cpre.org
The Social Context of Research Use
Jonathan Supovitz
Graduate School of Education, University of Pennsylvania
Consortium for Policy Research in Education
Northeastern Educational Research Association
Trumbull CT
October 26, 2016
Marriott
TRUMBULL
MERRITT PARKWAY
PennGSE


Effects of Home Resources and School Environment on Eighth-Grade Mathematics Achievement in Taiwan
NERA
Jingl Cai, Boston College
Introduction
Results
Conclusions
Implications