



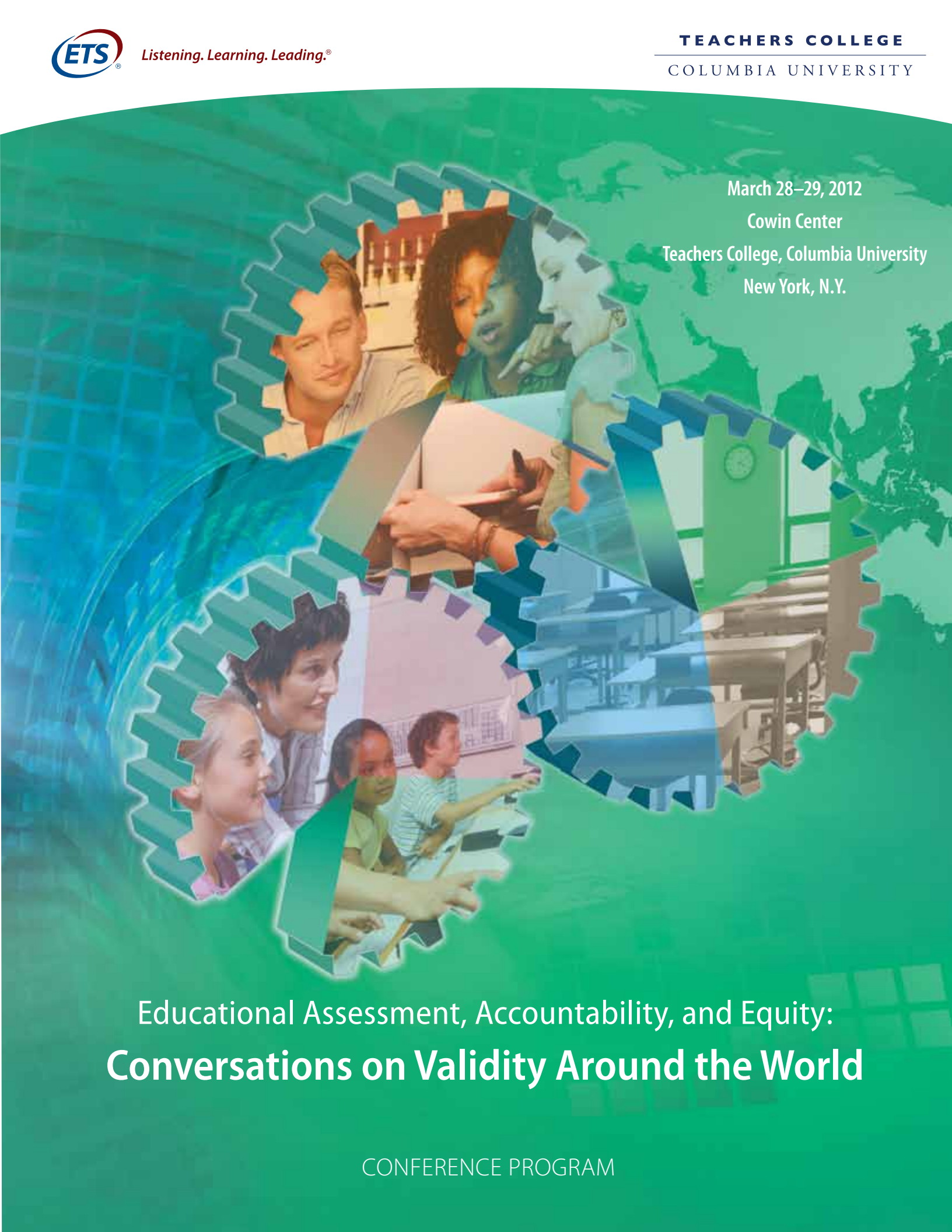
Listening. Learning. Leading.®

TEACHERS COLLEGE  
COLUMBIA UNIVERSITY

March 28–29, 2012

Cowin Center

Teachers College, Columbia University  
New York, N.Y.

The background of the cover is a vibrant green with a subtle grid pattern. Overlaid on this are several circular cutouts with gear-like edges. These cutouts contain images of people in educational settings: a man and a woman looking at a laptop, a woman pointing at a document, a woman and three children in a classroom, and a classroom with desks and a clock tower in the background.

Educational Assessment, Accountability, and Equity:  
**Conversations on Validity Around the World**

CONFERENCE PROGRAM

## TABLE OF CONTENTS

About ETS . . . . .	2
About Teachers College . . . . .	3
Welcome from ETS President and Chief Executive Officer Kurt M. Landgraf . . . . .	4
Welcome from Teachers College President Susan Fuhrman . . . . .	5
Agenda . . . . .	6–7
Presentation Abstracts. . . . .	8
Speakers . . . . .	9–10
Recent Reports from the ETS Policy Information Center . . . . .	11
Upcoming ETS Events . . . . .	12

## ABOUT ETS

Educational Testing Service is a nonprofit organization. Our mission is to advance quality and equity in education for all people worldwide.

We help teachers teach, students learn, and parents measure the educational and intellectual progress of their children. We do this by:

- ◆ Listening to educators, parents, and critics
- ◆ Learning what students and their institutions need
- ◆ Leading in the development of new products and services

**Our Mission:** To advance quality and equity in education by providing fair and valid assessments, research, and related services. Our products and services measure knowledge and skills, promote learning and educational performance, and support education and professional development for all people worldwide.

**Our Vision:** To be recognized as the global leader in providing fair and valid assessments, research, and related products and services to help individuals, parents, teachers, educational institutions, businesses, governments, countries, states, and school districts, as well as measurement specialists and researchers.

**Our Values:** Social responsibility, equity, opportunity, and quality. We practice those values by listening to educators, parents, and critics. We learn what students and the institutions they attend need. We lead in the development of products and services to help teachers teach, students learn, and parents measure the intellectual progress of their children.

[www.ets.org](http://www.ets.org)

## ABOUT TEACHERS COLLEGE

Teachers College, Columbia University, is the oldest and largest graduate school of education in the United States and also consistently ranked among the nation's best. Its name notwithstanding, Teachers College (TC) is committed to a vision of education writ large, encompassing our four core areas of expertise: health, education, leadership, and psychology.

TC sees its leadership role in two complementary arenas: One is as a major player in policymaking to ensure that schools are reformed and restructured to welcome all students regardless of their socioeconomic circumstances. The other is in preparing educators who not only serve students directly but coordinate the educational, psychological, behavioral, technological, and health initiatives to remove barriers to learning at all ages. For more than 100 years, TC has continued to:

- ◆ Engage in research on the central issues facing education
- ◆ Prepare the next generation of education leaders
- ◆ Educate the current generation of leaders in practice and policy to meet the challenges they face
- ◆ Shape the public debate and public policy in education
- ◆ Improve practice in educational institutions

Created in 1887 to prepare a new breed of school teacher to educate immigrant children, TC has been home to many of the defining figures in American education, including John Dewey, James Earl Russell, Edward Lee Thorndike, Isabel Maitland Stewart, Morton Deutsch, Edmund Gordon, and Maxine Greene.

TC was founded by the philanthropist Grace Hoadley Dodge and philosopher Nicholas Murray Butler to provide a new kind of schooling for the teachers of disadvantaged children in New York City, one that combined a humanitarian concern to help others with a scientific approach to human development and schooling. Historically, TC has a commitment to diversity and social justice.

Today, TC's three highly complementary and interrelated areas of study in education, psychology, and health work fulfill our vision of TC as a preeminent international human resource development institution, committed to systematic teaching and learning in all the major educative institutions. Students are candidates for Masters of Arts (M.A.), Master of Education (Ed.M.), Master of Science (M.S.), Doctor of Education (Ed.D.), and Doctor of Philosophy (Ph.D.) degrees.

A combination of distinguished faculty, students dedicated to developing as both scholars and practitioners, renowned research facilities, and all of the cultural and social resources of New York City make TC an exciting place to be and learn. Quick facts on size of the student body, faculty, and other information is available at [www.tc.edu](http://www.tc.edu).

## WELCOME FROM ETS PRESIDENT AND CHIEF EXECUTIVE OFFICER KURT M. LANDGRAF

Dear Colleagues:

Welcome to the international conference “Educational Assessment, Accountability, and Equity: Conversations on Validity Around the World.” Co-convened by Educational Testing Service (ETS) and Teachers College, Columbia University, the conference is an opportunity for educators, policymakers, and assessment specialists to examine issues related to the valid and meaningful use of information generated by educational assessments.

We hope to use this conference to encourage discussions about validity among technical and nontechnical professionals who are in a position to influence the appropriate use of educational assessments in applied educational settings.

Over the next two days, you will hear presentations from distinguished researchers on such topics as systems that are designed to measure student achievement and to evaluate teachers, schools, and institutions; where and how validity breaks down; and how it can be improved in both policy and practice.

ETS and TC have been at the forefront of the development and application of innovative methods and technologies used in educational assessments, in the United States, and around the world. Given the prevalence and impact of educational assessment — on individuals, institutions, public policy, and national economies — it is incumbent on assessment professionals to be attentive and responsive to the issues that assessments raise, and to the needs of educators and policymakers regarding assessment.

Thank you for your participation in the conference. I hope you enjoy the discussions and exchange of information, and that you find it valuable in your own work.

Regards,



Kurt M. Landgraf  
President and Chief Executive Officer  
ETS

## WELCOME FROM TEACHERS COLLEGE PRESIDENT SUSAN FUHRMAN

Dear Colleagues:

Welcome to TC and this important conference, which grew out of a partnership between TC’s Assessment and Evaluation Research Initiative (AERI) and ETS.

At the beginning of the 20th century, a brilliant young psychologist named Edward Lee Thorndike, who had recently joined the faculty of TC, began developing objective measurements of intelligence for use with human subjects. Thorndike subsequently created methods for measuring individual abilities and achievement in many domains, which he divided into three broad areas: abstract intelligence, mechanical intelligence, and social intelligence. His work helped to launch the modern field of educational testing and assessment.

Around the same period, James Earl Russell, who later became TC’s first dean, taught the world’s first course on comparative international education. A succession of TC thinkers, including Richard Wolfe, R.L. Thorndike, Elizabeth Hagen, and Harold Noah, have since helped to develop methods of measurement and evaluation for education systems in different countries. Today, we continue our research and development on international assessment programs, including the Trends in International Mathematics and Science Study and the Programme for International Student Assessment.

More than ever, these three areas — educational testing, evaluation, and international educational development are of paramount concern for us at TC.

In the United States and elsewhere, discussions surrounding educational testing, accountability, and equity often grow tense. On the one hand, educational assessment is seen as a means for finding out what students understand and don’t, and how the information can be used to facilitate the development of every student to realize his or her full intellectual potential. On the other hand, testing and assessment lie at the center of accountability policies that directly or indirectly affect students, parents, teachers, schools, districts, and education systems in general. While both these aims are laudable, they — and their champions — sometimes find themselves at loggerheads. And these are only two of many “high stakes” uses of educational assessments.

What are the consequences that various stakeholders face when tests and assessments, or the information they give us, is misused or inappropriately used? How can we make better use of educational assessment information around the world? The organizers of this conference have invited the perspectives of various stakeholders — assessment developers, researchers, policymakers, school-based practitioners, and other test users — to confront these significant questions on validity. We are very pleased to have you here to join in the search for answers.

With my best wishes,



Susan Fuhrman  
President  
Teachers College

## AGENDA

March 28–29, 2012  
Cowin Center  
Teachers College, Columbia University  
New York, N.Y.

### Wednesday, March 28, 2012

11 a.m.–2:30 p.m.	On-site Registration
1–2:15 p.m.	Guided Walking Historical Tour of Columbia University and Teachers College (Optional by registration)
2:30–3 p.m.	Opening Ceremony <b>Introductory speakers:</b> John Allegrante (TC, Columbia University) Susan Fuhrman (TC, Columbia University) Ida Lawrence (ETS) <b>Overview and orientation:</b> Madhabi Chatterji (TC, Columbia University)
3–5:15 p.m.	Validity, Fairness, and Testing <b>Keynote speaker:</b> Michael T. Kane (ETS) <b>Panelists:</b> Edmund W. Gordon (TC, Columbia University) Sébastien Georges (Centre International d'Études Pédagogiques) Kadriye Ercikan (University of British Columbia) Alina A. von Davier (ETS) <b>Moderator:</b> Madhabi Chatterji (TC, Columbia University)
5:30–7:30 p.m.	Reception and Dinner Buffet (Taste of New York)

### Thursday, March 29, 2012

7:30–9 a.m.	Breakfast Buffet
9–11:15 a.m.	Models of Teacher Evaluation and School Accountability Around the World <b>Keynote speaker:</b> Adrie Visscher (University of Twente) <b>Panelists:</b> Aaron Pallas (TC, Columbia University) Drew Gitomer (Rutgers University) Jakob Wandall (Independent Consultant, Denmark) Haniza Yon (MIMOS Berhad) <b>Moderator:</b> Hank Levin (TC, Columbia University)
11:15 a.m.–1 p.m.	Lunch
1–3:15 p.m.	Validity Issues in International Large-Scale Assessments <b>Keynote speaker:</b> Michael J. Feuer (George Washington University) <b>Panelists:</b> Hans Wagemaker (International Association for the Evaluation of Educational Achievement) Eduardo Backhoff (Universidad Autónoma de Baja California) Robert Laurie (New Brunswick Department of Education) Val W. Plisko (retired – National Center for Education Statistics) <b>Moderator:</b> Eugenio Gonzalez (ETS)
3:15–3:45 p.m.	Light Refreshments
3:45–6 p.m.	Bringing the Validity Conversations Home: When Education Measures Go Public* <b>Keynote speaker:</b> Eva L. Baker (University of California, Los Angeles) <b>Panelists:</b> Nick Lemann (School of Journalism, Columbia University) David Steiner (City University of New York and former New York state Education Commissioner) Kevin Welner (University of Colorado at Boulder) Leo Casey (United Federation of Teachers, New York City) <b>Moderator:</b> Jeffrey Henig (TC, Columbia University)  Closing Ceremony
6–6:15 p.m.	Provost's Opening Remarks for the AERI Institute Thomas James (TC, Columbia University)

\*Session co-sponsored by the National Science Foundation, the National Education Policy Center at the University of Colorado at Boulder, and the Department of Educational Policy and Social Analysis at Teachers College, Columbia University.

## PRESENTATION ABSTRACTS

### **The Chimera of Validity**

*Eva L. Baker*

This presentation will treat the long-valued topic of validity, with a brief review of its history through recent guidelines, to redefine and assist its use in educational assessment and testing. As a context, the 1999 “Standards for Testing and Assessment” will be used. Three examples will be given of ways in which validity claims have not been technically supported and potential causes will be considered. Given the potential changes in assessment design and delivery, such as technological options, the development of badges, and other qualifications to recognize deep and extended student work, the future of validity issues can be sketched.

### **Validity Issues in International Large-Scale Assessments: Truth and Consequences**

*Michael J. Feuer*

This paper has three goals: to consider the policy context of international comparisons of educational achievement in the United States and elsewhere; to explore possible lessons for educational comparisons from cross-national studies in economics and demography; and to propose an expanded framework for validity definition in cross-national assessments, borrowing from Samuel J. Messick’s formulation of “consequential validity.”

### **Validity, Fairness, and Testing**

*Michael T. Kane*

Validation requires two kinds of arguments — an *interpretive argument* and a *validity argument*. The *interpretive argument* specifies the proposed interpretation and use of test scores in terms of a sequence of inferences and assumptions leading from observed test performances to claims about attributes of the test taker, and typically, to decisions about the test taker. The *validity argument* validates the proposed interpretation and use by evaluating the plausibility of the inferences and assumptions in the *interpretive argument*. That is, in validating a proposed interpretation or use, we first lay out the claims being made and then we evaluate these claims. An evaluation of a test use necessarily requires an evaluation of the consequences of the use. If the consequences are, on the whole, positive, the use is justified. If the consequences are judged to be more negative than positive, the use is unjustified. Evaluating the consequences of a decision rule tends to be difficult and potentially contentious, but if we want to evaluate test uses, it is necessary. Fairness plays a major role in the evaluation of both the interpretations and uses of test scores. In the context of interpretations, fairness can be defined in terms of the consistency of score meanings and implications (e.g., predictions of various criteria) across groups. For test uses, fairness can be defined in terms of the appropriateness of the decisions based on test scores, and, therefore, depends mainly on the consequences of these decisions.

### **Models of Teacher Evaluation and School Accountability Around the World: A Dutch Perspective**

*Adrie Visscher*

A description will be given of how the Dutch schools inspectorate holds schools and teachers accountable for their performance, and how Dutch schools are performing. Next, the potential of various ways of using evaluation data for performance improvement will be discussed. It is stated that benefitting most from evaluation data requires a school organization that operates in a goal-oriented way and in which teaching is deprivatized, performance feedback is used for improvement, and teachers are supported in improving the knowledge and skills that matter for student achievement. An important prerequisite for such a mode of operation is a fruitful balance between external evaluation and improvement.

## SPEAKERS

### ***Eva L. Baker***

Eva L. Baker is a Distinguished Professor in the divisions of Psychological Studies in Education and Social Research Methodology at the UCLA Graduate School of Education and Information Studies at the University of California, Los Angeles (UCLA). Baker has directed the UCLA Center for the Study of Evaluation since 1975. She also is the Director of the National Center for Research on Evaluation, Standards, and Student Testing, a competitively awarded national institution funded by the U.S. Department of Education.

Baker is a member of the National Academy of Education and a recipient of the 2007 ETS Henry Chauncey Award for Distinguished Service to Assessment and Educational Science. She was a congressionally appointed member of the National Council on Education Standards and Testing and chair of the Board on Testing and Assessment, National Research Council, National Academies (2000–2004). Baker is a former president of the American Educational Research Association (2006–2007), a former president of the Educational Psychology Division of the American Psychological Association, and a former editor of “Educational Evaluation and Policy Analysis.” She was co-chair of the committee to revise the *Standards for Educational and Psychological Testing* (1999).

Baker’s research is focused on the integration of instruction and measurement, including design and empirical validation of principles for developing instructional systems and new measures of complex human performance. She also is involved in the design of technologically sophisticated testing and evaluation systems of assessment in large-scale environments for both military and civilian education.

### ***Michael J. Feuer***

Michael J. Feuer is Dean of George Washington University’s Graduate School of Education and Human Development. Previously, he served as the Executive Director of the Division of Behavioral and Social Sciences and Education in the National Research Council of the National Academies, where he directed the Board on Testing and Assessment and the Center for Education. He has served as a senior analyst at the Congressional Office of Technology Assessment, where he specialized in educational technology and measurement. He also has held faculty positions at Drexel and Georgetown universities and has been a senior adviser to many foundations and educational institutions in Europe, Israel and the United States.

Feuer has published widely in education, economics, and public policy journals, and has written essays, reviews, and poems for newspapers and magazines. His most recent book, “Moderating the Debate: Rationality and the Promise of American Education,” was published by Harvard Education Press in 2006. Feuer was the Milton Angoff Lecturer at ETS in February 2011. He also is an elected member of the National Academy of Education, and fellow of the American Association for the Advancement of Science and the American Educational Research Association.

Feuer holds a Ph.D. in public policy from the University of Pennsylvania and an M.A. from the Wharton School. He graduated *cum laude* from Queens College (City University of New York) with a major in English literature and journalism, studied public administration at the Hebrew University of Jerusalem, and studied political science at the Sorbonne.

## SPEAKERS *(continued)*

### **Michael T. Kane**

Michael T. Kane is the Samuel J. Messick Chair in Test Validity at Educational Testing Service. From September 2011 to August 2009, he served as Director of Research at the National Conference of Bar Examiners. From 1991 to 2001, he was a professor in the School of Education at the University of Wisconsin, where he taught measurement theory in the Department of Kinesiology. From 1982 to 1991, Kane served as a vice president for research and development and a senior research scientist at American College Testing in Iowa City. He also served as Director of Test Development at the National League for Nursing from 1976 to 1982.

Kane's main research interests are validity theory and practice, generalizability theory, licensure and certification testing, and standard setting.

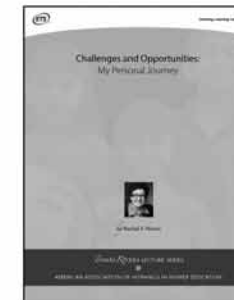
He holds a Ph.D. in education and an M.S. in statistics from Stanford University. He also has a B.S. from Manhattan College and M.A. from the State University of New York in physics.

### **Adrie Visscher**

Adrie Visscher is an associate professor of Behavioural Sciences at the University of Twente in the Netherlands, where he has been teaching and researching school effectiveness since 1983. He is especially interested in the organizational and management aspects of educational institutions and in how school performance feedback can support improving school performance. Currently, he is involved in projects on data-driven teaching, in which school staff are trained for two years with respect to how they can analyze their performance data, as well as translate the findings into goals to be achieved and instructional strategies to be carried out. The effects of the longitudinal training course on school quality are studied in 150 schools.

## RECENT REPORTS FROM THE ETS POLICY INFORMATION CENTER

Download copies from [www.ets.org/research/pic](http://www.ets.org/research/pic)



### **Challenges and Opportunities: My Personal Journey**

In the 2011 Tomás Rivera Lecture, Rachel F. Moran, Dean of the UCLA School of Law, combines her personal experience with an analysis of the pressing issues affecting educational opportunity for Hispanic Americans.



### **ETS Policy Notes — Optimizing Talent: Closing Educational and Social Mobility Gaps Worldwide**

This issue provides highlights from the first of three international conferences on education and social mobility co-sponsored by ETS and the Salzburg Global Seminar.



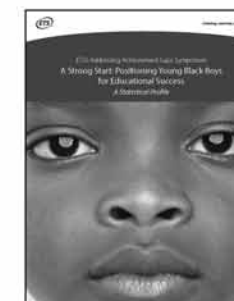
### **The Mission of the High School: A New Consensus of the Purposes of Public Education?**

This report provides a perspective on a variety of issues that need to be addressed as the mission of the high school is transformed to prepare all students for college and careers.



### **ETS Policy Notes — The Family: America's Smallest School**

The challenge of helping families prepare their children for school success was the subject of "The Family: America's Smallest School," the 14th in ETS's series of "Addressing Achievement Gaps" symposia.



### **A Strong Start: Positioning Young Black Boys for Educational Success**

This brief statistical profile provides a snapshot of the rugged terrain many young Black males traverse during the course of their lives. Beginning at birth, the data describe the individual, family, school and community factors that impact a child's development, as well as suggest important transition points where opportunities to develop human capital are gained or lost.



### **Access/Acceso: Rising to the Challenge of Improving Higher Education Opportunities for Latinos**

In the 2010 Tomás Rivera Lecture, California chancellors Charles B. Reed and Jack Scott address higher education issues ranging from access and recruitment to retention and graduation of underrepresented students.

## UPCOMING ETS EVENTS

### **Future Directions for Kindergarten Readiness and Assessment: Policies, Measures, and Issues to Consider**

May 17–18, 2012

Educational Testing Service  
Chauncey Conference Center  
Princeton, N.J.

For details, contact Jeanette Negroni at [jnegroni@ets.org](mailto:jnegroni@ets.org).



Copyright © 2012 by Educational Testing Service. All rights reserved. ETS, the ETS logo and LISTENING. LEARNING. LEADING. are registered trademarks of Educational Testing Service (ETS). 18874



*Listening. Learning. Leading.®*

[www.ets.org](http://www.ets.org)