FIVE ETS CHAIRS

Frederic M. Lord Chair in Measurement and Statistics
Samuel J. Messick Chair in Test Validity
Edmund W. Gordon Chair for Policy Evaluation and Research
Norman O. Frederiksen Chair in Assessment Innovation
Ralph W. Tyler Chair in Large-Scale Assessment
table of contents

Our Research Chairs Connect Us to the Future........................................3–5

Lord Chair: Robert J. Mislevy .................................................................6–7

Messick Chair: Michael T. Kane .........................................................8–9

Frederiksen Chair: Randy E. Bennett ..................................................10–11

Gordon Chair: Michael T. Nettles ......................................................12–13

Tyler Chair: Irwin S. Kirsch .................................................................14–15

About ETS Research & Development ................................................16

This brochure contains excerpts from five extensive interviews with current holders of ETS’s Research Chairs. It can be downloaded at www.ets.org/research/chairs.
Ida Lawrence, Senior Vice President of ETS’s Research & Development (R&D) division, considers the five Research Chair positions — and the scholars who hold them — among ETS’s most precious resources.

“We have world-class leaders here, but they have been hidden in plain sight,” she says.

The creation of the Research Chair positions, named after pioneers in educational assessment, seems to reflect a grand strategy, but they were actually added pragmatically, one after the other, over a decade.

“We created the chairs to attract scholars and luminaries, and keep those that we have,” Lawrence says.

ETS’s Board of Trustees view the Research Chair positions as a way to honor the company’s scientific pioneers and recognize today’s scholars. The first, the Frederic M. Lord Chair in Measurement and Statistics, was created to encourage Paul W. Holland to return to ETS. He left ETS in 1993 for the University of California, Berkeley, and returned in 2000 to assume the chair. Named after Fred Lord, a pioneer in psychometrics, the first chair honors a man once called the “Father of Modern Testing.”

“I accepted the Lord Chair because it allowed me to do research that I could not pursue otherwise. Among other things, with the help of two great ETS colleagues, I wrote the book on kernel equating that I had thought about for many years. As Fred Lord’s work had done before me, this has led to many other ideas that have rationalized the field of test equating,” Holland says.

Theoretical challenges, practical approaches

ETS’s Research Chairs differ from those of the academic world, as they reside within a nonprofit organization that creates, administers, and scores more than 50 million tests every year.

Our business is closely tied with our social mission — to advance quality and equity in education by providing fair and valid assessments, research, and related services — and this connection gives our research a practical orientation. Among the five chairs, even the most theoretical challenges are often approached in a down-to-earth way. They are keenly aware that they ultimately serve test takers around the world.
“I spend a lot of time thinking about what we actually do and what kind of assumptions we make,” says Michael T. Kane, who has held the Samuel J. Messick Chair in Test Validity since 2009. Messick was a leading authority in validity theory and worked at ETS for more than three decades.

“One of my early papers was about the validity of licensure exams, laying out a model for how to think about that,” Kane says. “I have generally found that thinking of such problems helps me stay focused and grounded.”

Kane also appreciates the freedom that the chair position provides, allowing him to look at what ETS does from alternative perspectives and ask, as he puts it, “the ultimate questions.”

“The theoretical can be very practical. If you have a theory that tells you how to solve certain problems, that is the most practical thing you could possibly have. It’s not so much the practical versus the theoretical, but that you want conceptual systems and methodologies that are internally consistent and coherent, and that also address problems that need to be addressed,” he continues. “You need theory to do that, but the theory has to be grounded and responsive.”

**Working to improve the practice of assessment**

Robert J. Mislevy, who assumed the Lord Chair in 2010, also takes a practical approach to research.

“My first task is connecting cognitive psychology and learning psychology with psychometric and measurement principles. My job is to try to get the most bang for ETS’s buck in making improvements in the field,” he says.

“Fred Lord was very much connected to the practical problems of assessment, but he also read theoretical statistics. Paul Holland worked on the theory of casual reasoning and applied those ideas to the kind of inferences you can do from the National Assessment of Educational Progress, also known as NAEP,” Mislevy continues. “He applied sometimes esoteric theories about causation to very practical problems of assessment design and policy-making. I am trying to do the same thing. I want to, in the spirit of Fred Lord, have a foot in the practical and necessary problems and whatever research fields going on, so that we can leverage it to improve the practice of assessment.”

Randy E. Bennett holds the Norman O. Frederiksen Chair in Assessment Innovation, named after a pioneer in constructed response, performance assessment, and simulation. Frederiksen was particularly interested in applying findings from cognitive psychology to testing, as well as in creating assessments that influenced individuals and institutions positively.

The Cognitively Based Assessment of, for, and as Learning (CBAL™) initiative, which Bennett leads, carries on Frederiksen’s legacy.
“We use advances in learning sciences, technology, and measurement to create new testing approaches that both measure well and model good classroom practice for teachers and students,” Bennett says.

**Taking on complex policy issues in education**

The fourth chair was created in 2004 in honor of Edmund W. Gordon, who was referred to as “the premier Black psychologist of his generation” in *The New York Times*. He is a pioneer in educational assessment and professor emeritus at Yale and Columbia universities. In 2004, ETS named Michael T. Nettles, Senior Vice President of the Policy Evaluation & Research Center (PERC), as the first holder of the Edmund W. Gordon Chair for Policy Evaluation and Research. Nettles is an expert on fairness in the educational system, and worked as a research scientist at ETS from 1984 to 1989. In 2003, he returned to lead ETS’s policy research program.

“The goal of the Gordon Chair is to advance the research and address complex policy issues in education, including the pursuit of equitable education,” says Nettles, who is also involved in assessment development targeting disadvantaged groups. “Gordon has been a prolific thought leader in this field. He has established many theoretical bases for the work and conducted much of the early research.”

The fifth chair was created in April 2012 in honor of Ralph W. Tyler and is held by Irwin S. Kirsch, an expert on large-scale international assessments who, since 2000, has led the Organisation for Economic Co-operation and Development’s (OECD’s) Reading Expert Group at the Programme for International Student Assessment (PISA). He currently serves as the International Project Director for the 2015 round of PISA, which is administered to 15-year-olds in more than 80 countries. He also leads the Programme for the International Assessment of Adult Competencies (PIAAC), the first international adult assessment to be delivered on computers.

“Large-scale survey assessments provide various stakeholders with information about groups rather than individuals,” Kirsch says. “They are growing more important to ETS and are increasingly popular among policymakers around the world.”

A former Chair of the Department of Education as well as Dean at the University of Chicago, Tyler was an eminent scientist who advised seven U.S. presidents on educational issues. He may be best known for his role in laying the groundwork for the development of NAEP, which began in 1969.

Together, the five Research Chairs offer ETS a platform to look both deep and far.

“The chairs highlight the fact that ETS is a significant organization in terms of innovation and scientific status,” Lawrence says. “You could say that the scientists holding the chairs are keeping us in touch with the future.”
Robert J. Mislevy assumed the Lord Chair in 2010 when he returned to ETS, after serving a decade at the University of Maryland in College Park, Md. Looking ahead, he expects fundamental changes in assessments, technology, and cognitive science.
Mislevy joined Fred Lord’s Model-based Measurement group when he first came to ETS as a young researcher. During the 1980s and 1990s, he worked on the NAEP program and got involved in cognitive science and psychometrics while working for the Office of Naval Research. He says he enjoyed the idea of connecting various disciplines and linking them to developments in the real world. One result of this approach was Evidence Centered Design (ECD) of assessments, which he developed with a team of ETS researchers. The promise of ECD was that it could replace the professional “hunch” with clear reasoning based on sound evidence, a technique that would prove useful in new types of assessment projects, whether it was used to assess service technicians for F-15 fighter jets or video recordings of aspiring teachers working in classrooms.

Mislevy’s interest in new assessment environments includes new computer-based platforms and the challenges they present.

“We have interactive environments where people learn and are assessed, but our specific tools from psychometrics are not always up to the job,” he says. “However, if you understand the principles of psychology, of building assessments, and of how to use statistical methods to manage that evidence, you can then create psychometric methods that are appropriate to forms of assessments that nobody has seen yet.”

ECD can be used to create such open-ended assessments, but the test delivery also presents challenges. Mislevy is working with Universal Design for Learning (UDL) as a way to get developers to start thinking about fairness and accessibility earlier in the process.

“With UDL, you are trying to figure out different ways for people to access and interact with what the test is supposed to be about,” he says.

Mislevy’s approach to assessment design shares aspects of software development.

“We drew on ideas that we found in a number of fields and software design was certainly one of them, including object oriented programming. Another major source was research on evidentiary reasoning,” he says.

“My job is to apply statistics, technology, and cognitive psychology to practical problems in assessment, which also is what Fred Lord did,” Mislevy continues. “I am trying to figure out the underlying evidentiary principles in assessment, how to make them explicit with language, representation, processes, and then how to put that to work in new and cutting-edge forms of assessment. That is in a nutshell what I am working on, and these are things that motivate us who hold these chairs. We ask what we can do to push the frontier forward, and use the underlying principles of assessment in the context of ETS to do things that haven’t been done before.”

He adds that today is an exciting time to work in the assessment field.

“For 80 years we saw many important improvements in the field, but now we are facing a whole new world of technology and science — a paradigm shift,” Mislevy says. “What were the underlying assessment principles that worked for the past 80 years? Can we leverage them to do principled assessment in the new world, and to do it in a valid and effective way?”
Early in his academic career, Michael T. Kane, the Messick Chair, studied physics. He earned a master’s degree in the subject at the State University of New York (SUNY) at Stony Brook and was well on his way toward earning a Ph.D. But he changed his mind, and studied mathematical methods in education at Stanford University instead. He ultimately earned a master’s degree in statistics and a Ph.D. in education from Stanford and returned to SUNY at Stony Brook. However, his career there was cut short when New York state shut down Stony Brook’s education department in a restructuring effort.
Kane’s next stop was the National League of Nursing, which developed licensing tests for the American Nursing Association.

“The league did a study of the test’s validity, which is why I took the job,” he says. The validity model and especially how it deals with consequences are issues he continues to work on today.

“The generic challenge we encounter in educational measurement and in behavioral measurement in general has to do with trade-offs,” Kane says. “We can’t have everything, so we have to figure out how we can get the most of what we want with the least of what we don’t want. We know for instance that we can make tests more reliable by making them longer, and that we can make very reliable tests by using lots of multiple-choice questions. But sometimes, we are interested in more complex performances.”

“For example, when you take the Regents’ exam in history, you have performance tasks where you have to write essays, requiring you to have developed skills at producing a coherent analysis of some fact or situation,” he continues. “The test may show you a cartoon about Boss Tweed and legislation to control corruption, and then ask questions about how they relate. You can’t really do that in a multiple-choice question, but multiple-choice tests have good reliability and cover the whole spectrum, so that students who hadn’t spent a lot of time on that particular aspect of American history wouldn’t be severely penalized.”

Kane explains that it may be good to design a test so that it has some performance tasks, even if that means giving up something in terms of reliability and introduces some difficulty in equating it. On the other hand, relying only on one or a few performance tasks risks underestimating someone’s skills in an area.

He also focuses on the overall importance of creating fair and valid assessments within the scope of ETS’s charter.

“ETS is not just a business. In many ways, it functions like a business, but it is a nonprofit organization that provides assessments in order to create educational opportunity for people all over the world,” Kane says. “We need to be concerned about producing tests that are of high quality because the consequences of not doing so are unacceptable. There is too much at stake.”

ETS works diligently to ensure valid exams, particularly as new challenges are presented.

“We are always looking at improving the efficiency with which we allocate resources, especially when we’re dealing with a wide variety of testing programs. We are also on the cutting edge of developing new kinds of assessments and that is probably where validity issues need to be addressed the most. If you are talking about predicting first-year GPAs using a college admission test, those issues have been hashed over for 40 or 50 years. But there are a lot of new issues we need to consider when we, for example, work on new assessment formats in math or innovative language assessments,” Kane says.

“We have a lot of elegant statistical models, value-added models in particular, that hold some promise. But they have a lot of potential problems of their own, and we should pay attention to that as a research organization,” he continues. “It’s a challenge, but an exciting one for ETS.”
Randy E. Bennett — who holds the Frederiksen Chair — seeks to overcome our traditional tests’ perceived negative impact on teaching and learning by integrating advances in the learning sciences, measurement, and new technology.

Bennett, who was born in Brooklyn, had his mind set on becoming a writer like his father, Jay. But he knew that it would be a difficult career, so he had a backup plan of becoming an English teacher when he entered SUNY at Stony Brook.
Once out of college, he tried the writer’s life. “It didn’t take me long to realize that it was hard to sit down each day and work on my novel, so I started to look around for teaching jobs,” Bennett says.

His search began in 1975, when New York City reeled under a fiscal crisis and The Daily News ran its famous front-page headline: “Ford to City: Drop Dead.” It was difficult for newly minted teachers to land a job. Bennett ended up working in poor neighborhoods, where the schools were often plagued by violence and many teachers and administrators had given up on their jobs.

If there was one silver lining, it was that Bennett gained a broad experience of teaching English, remedial reading, and remedial math. It also spurred him to apply for graduate school, and he was accepted into a master’s program in reading and learning disabilities at Teachers College, Columbia University. In 1977, he entered the school’s doctoral program, focusing on educational measurement and special education. While searching for an internship to support himself during his work on his dissertation, he found one at ETS and soon realized that he wanted to become a researcher at the organization.

There were no job openings for research scientists when his internship was over, but Bennett found a part-time job that allowed him to meet colleagues across different areas of ETS. One of them, Gerry Bogatz, was a program director in the Division of Elementary and Secondary School Programs. She hired him for a full-time position developing ideas for new products for students with disabilities. This was in the early 1980s, also known as the beginning of the “micro-computer” revolution.

“The IBM® Personal Computer had just come out, and I bought one,” Bennett says. “My salary at the time was about $25,000, and that machine cost $5,000 with the printer. It was 20 percent of my gross income. That was a big investment, but I felt that becoming familiar with the computer was going to be important.”

He didn’t know it then, but his investment in a PC would open a door for him in the Research division when one of the staff members received a large grant from IBM. The computer giant asked ETS to run a pilot program to train teachers to use the IBM PC in schools.

“The Research division was looking for people to help implement this project,” Bennett explains. “And I had an IBM PC, which gave me experience few others had at the time.”

He continued working on assessment projects for students with disabilities, but gradually grew more interested in using technology to assess students in general. In 1985, Bennett began working with Henry Braun, who was a director of a department that would later become part of today’s Statistical Analysis, Data Analysis, and Psychometric Research area in R&D. They wanted to know how to score responses to open-ended questions more effectively — a problem that was of no immediate practical value. There were no educational computer-based tests (CBTs) commercially available before ETS launched its first CBT, the Computerized Placement Tests, in 1986. And that test, like virtually all other CBTs that followed, had only multiple-choice items.

Bennett is closely associated with computers and the use of technology in assessment, but his focus was always on bringing together technology, cognition, and assessment.

“The integration of cognitive and learning science, measurement, and technology, which characterized our automated scoring studies, has been the theme of almost everything I’ve done since,” he says.

This integration is also reflected in the CBAL initiative, which Bennett launched in 2007. “CBAL began with a deep analysis of the knowledge, processes, strategies, and habits of mind that were important for success in the domains of reading, writing, and mathematics,” he says. “Based on this analysis, the CBAL team has created prototype assessments that not only measure those competencies, but model the learning and instruction of them.”
As a young man, Michael T. Nettles first dreamt of becoming a professional baseball pitcher, then a sports journalist. Today, he is Senior Vice President of PERC, holder of the Gordon Chair, and head of ETS’s Early Childhood Research Assessment Center.
When Nettles entered the University of Tennessee in 1973, he studied political science and economics as well as the financial underpinnings of education and politics. Segregation was still very much a factor at the time, and college admissions and funding were tightly intertwined with racial discrimination. As a graduate student, Nettles experienced these dynamics first hand when he landed an internship and later a job at the Tennessee Higher Education Commission (THEC). Tennessee and 18 other Southern states were sued for operating segregated educational systems (Geier v. Tennessee, 1968), a practice that was first struck down by the U.S. Supreme Court (Brown v. Board of Education, 1954).

The state was court-ordered to dismantle its segregated system, but it tried a work-around by setting up a separate school belonging to the University of Tennessee in downtown Nashville in order to compete with and undercut Tennessee State University, which had a predominantly African-American student population.

“This was a blatant perpetuation of the dual system,” Nettles says.

He continued on to graduate school, earning master’s degrees in political science and education, and a Ph.D. in education. One thing he learned from his work on the Geier case was the importance of being able to present evidence supported by facts.

As Nettles plunged into the world of data, seeking tangible proof to build a case for equity, his fact gathering and analysis led to an ever-growing list of studies and publications centered on educational equality and quality of education. His passionate interest in data eventually led Nettles to ETS, as the organization took interest in his research. Joan Baratz, then the Director of PERC, asked him to present at ETS after seeing one of his presentations in Washington, D.C. Nettles visited ETS in May of 1984. By July, he was working as a research scientist for the Policy Research division.

His first stay at ETS was short. Lamar Alexander, President of the statewide University of Tennessee system, asked Nettles to join as Vice President in 1989. Nettles became a professor at the University of Michigan three years later, and soon joined the National Assessment Governing Board, which runs NAEP.

In 2003, ETS’s President & CEO Kurt M. Landgraf, invited Nettles to return to ETS and appointed him as Director of the Policy Research division. In 2004, Nettles assumed the Gordon Chair.

Nettles currently focuses on issues of equity in education through research, while working with ETS’s Assessment Development area on assessments that can help reduce disadvantages among demographic groups. This work is part of his long-term effort to create knowledge and awareness, and to address achievement gaps experienced by children from disadvantaged social and ethnic populations. These are the students, according to Nettles, who receive limited exposure to high-quality instruction and assessment.

“I am focusing on performance gaps across racial groups and social classes, and trying to craft solutions that would help overcome differences in opportunity and performance,” he says. “I am looking at a variety of strategies to help advance the disadvantaged.”
While growing up in Baltimore, Irwin S. Kirsch — the Tyler Chair — honed his navigational and teamwork skills through sailing and lacrosse. Today, he uses similar skills working for international organizations like the OECD, where he plays a leading role in the development of large-scale international assessments like the PISA and PIAAC.
Kirsch's path to ETS and the field of international large-scale assessments began at the University of Maryland, where he studied psychology. He subsequently worked as a research assistant at the Kennedy Krieger Institute, which was connected with the Johns Hopkins School of Medicine. One of his responsibilities was to work on a research project with John Guthrie, sometimes sharing lab resources with another young research assistant named Mary Seifert.

When Guthrie was offered a job as Director of Research at the International Reading Association in Delaware, he asked Mary, who by then had become Mrs. Kirsch, to continue as his assistant. This presented a dilemma for the young couple, as Kirsch had enrolled in Johns Hopkins' master's program in communications disorders. Guthrie suggested that he might consider enrolling in the educational psychology program at the University of Delaware. After finishing his master's program, he earned a doctorate in measurement and evaluation at Delaware. Kirsch and Guthrie wrote and published several articles dealing with the measurement of reading literacy and received a grant to study literacy in the workplace.

Kirsch's work on literacy eventually led him to ETS.

“One day in June of 1984, ETS called to ask if I was interested in becoming the Project Director for a new national study, the Young Adult Literacy Study, or YALS, which was being funded through the NAEP group at ETS,” Kirsch explains. “I agreed to take on the role, and that is what brought me to ETS and to large-scale assessments.”

At ETS, Kirsch met with Victor Bunderson, who was then Vice President for Research Management, and he eventually moved to the R&D division, where he worked on a project that investigated how to link assessments with instruction. His interest in literacy continued to broaden and Kirsch became involved with several other large-scale national assessments — one for the U.S. Department of Labor and another for the U.S. Department of Education. He also worked with researchers in Canada, who wanted to know if the methodology he had helped develop around literacy could be applied in the bilingual Canadian context. This led to the development of international assessments, such as the International Adult Literacy Study and the Adult Literacy and Life Skills survey.

In 2005, Kirsch, ETS colleague Kentaro Yamamoto, and a representative from Statistics Canada approached the OECD with an idea for a new international adult assessment, which eventually became PIAAC.

“Working with these surveys provides ETS with a unique opportunity to expand our methodological and technological capabilities,” Kirsch says. “This, in turn, will better support the needs of policymakers and other stakeholders — both here in the United States and around the world.”

An example of how data from large-scale assessments can be used to inform policymakers is America's Perfect Storm: Three Forces Changing Our Nation's Future, a report published by ETS’s Policy Information Center in 2007. Kirsch wrote the report with Yamamoto, Henry Braun of Boston College, and Andrew Sum of Northeastern University.

“It is a perfect example of a policy report that relies on and benefits from survey data, both national and international,” Kirsch says.
About ETS Research & Development

ETS's R&D staff includes internationally distinguished research scientists, psychometricians, and assessment experts. Together, they work to:

- provide fair, accurate, and meaningful products and services
- drive innovation that advances assessment, learning, and teaching
- provide content, methods, and processes for manufacturing our products
- advance the field of educational measurement
- promote effective and equitable education policy

Our research focuses on core areas such as assessment innovation, cutting-edge psychometrics, validity and fairness, natural language processing, and automated scoring. For more information about R&D, visit www.ets.org/research.
FIVE ETS CHAIRS

Frederic M. Lord Chair in Measurement and Statistics
Samuel J. Messick Chair in Test Validity
Edmund W. Gordon Chair for Policy Evaluation and Research
Norman O. Frederiksen Chair in Assessment Innovation
Ralph W. Tyler Chair in Large-Scale Assessment

Listening. Learning. Leading.®

www.ets.org