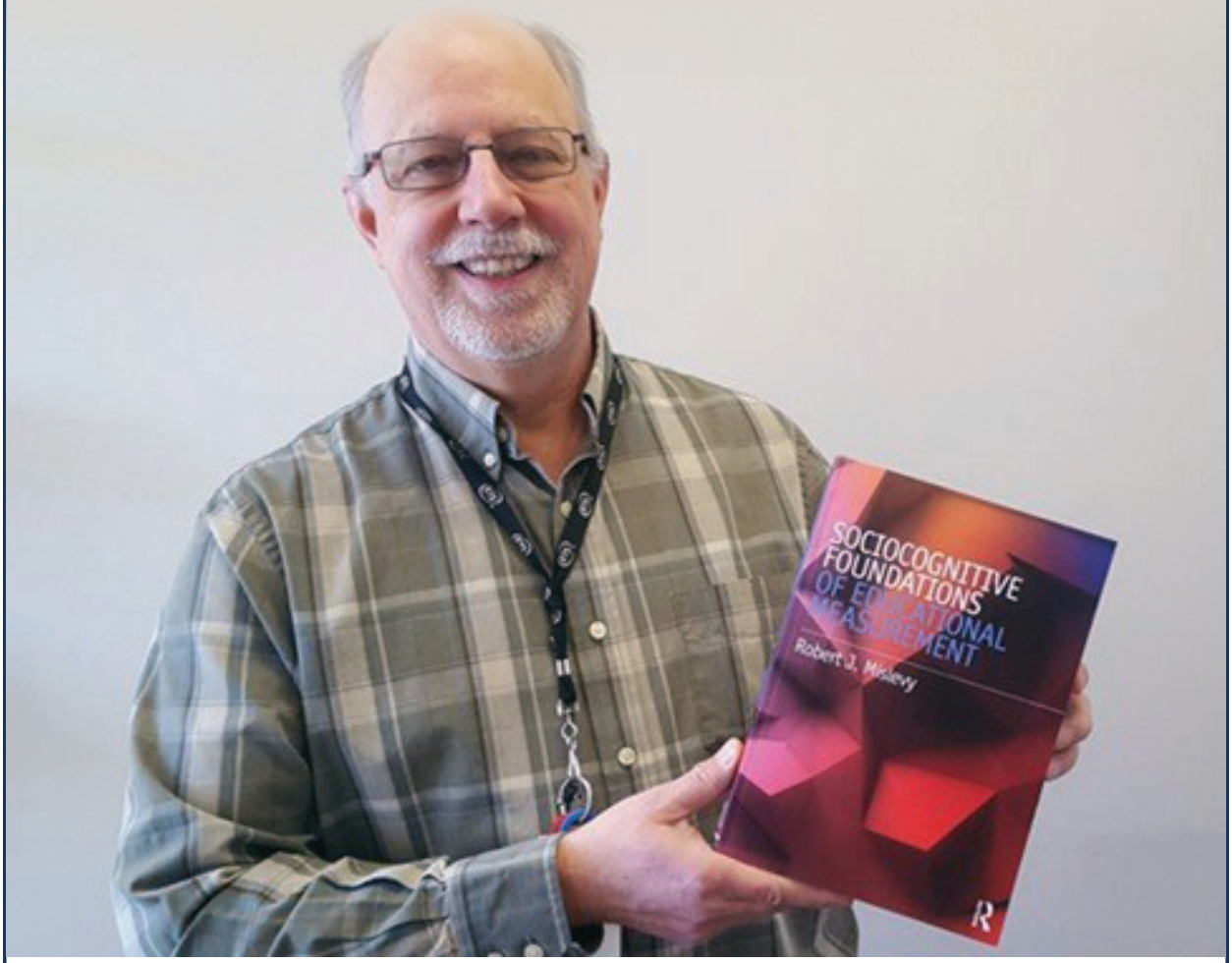
REMINISCENCES:

REFLECTIONS ON THE LIFE OF

Robert J. Mislevy



A collection of remembrances
for a life that shaped so many.

Reminiscences: Reflections on the Life of Robert J. Mislevy

Edited by Maria Elena Oliveri, Eric M. Tucker, Sheryl L. Gómez,
David Slomp, and Norbert Elliot

About these Reminiscences

In the spirit of Robert J. Mislevy's own scholarship, this collection gathers the evidence traces of a life defined by brilliance and kindness. Co-editors Maria Elena Oliveri, Eric M. Tucker, Sheryl L. Gómez, David Slomp, and Norbert Elliot aim to honor Bob as a friend, colleague, and mentor. These reminiscences reflect a joint effort between *The Journal of Writing Analytics*, The Study Group, and these co-editors. The authors of each remembrance together made this volume possible, and we are grateful for each inspiring and heartfelt contribution. Beyond honoring Bob's memory, this collaboration documents progress of a wider and varied campaign to document his scholarship and launch new research programs that build upon his foundational ideas. From psychometric innovation to Bayesian reasoning, from Evidence-Centered Design to sociocognitive foundations of educational measurement, the work continues through the community he inspired. We offer these reminiscences as both a celebration of the past and a foundation for the future of educational measurement. We invite you to read these reflections as both a memorial and the first chapter in the research programs inspired by his intellectual journey.

REMINISCENCES:
Reflections on the Life of Robert J. Mislevy

December 2025

Reminiscences: Reflections on the Life of Robert J. Mislevy

Maria Elena Oliveri, *Purdue University*

Eric M. Tucker, *The Study Group*

Sheryl L. Gómez, *The Study Group*

Norbert Elliot, *New Jersey Institute of Technology*

With heavy hearts and deep gratitude, we share this volume of *The Journal of Writing Analytics* (*JWA*), devoted to the life and work of Robert J. Mislevy. His passing on May 22, 2025, is an immeasurable loss, not only to educational measurement scholars but to the many of us who had the privilege of working alongside him, learning from him, and being shaped by his brilliance and generosity of spirit. His absence is deeply felt. As the reminiscences gathered here make clear, his influence remains vividly present in our scholarship, our institutions, and our practices with colleagues and students.

This collection, *Reminiscences: Reflections on the Life of Robert J. Mislevy*, brings together voices from across the communities Bob touched. In gathering these reminiscences, we found ourselves tracing the constellations of ideas, commitments, and human connections that Bob set into motion over the course of his career. Each contribution in this collection illuminates a different facet of his influence—sometimes through theoretical breakthroughs, sometimes through methodological innovations, and often through the inspirational, calm way Bob helped people become better thinkers, collaborators, and versions of themselves. Taken together, the pieces sketch a view of a scholar who reshaped scholarly fields and programs of research while never losing sight of the people in it. Contributors to this volume include Russell Almond, Eva L. Baker, John T. Behrens, Randy Elliot Bennett, David Dorsey, T. J. Elliott, Howard T. Everson, Michael J. Feuer, Drew H. Gitomer, Edmund W. Gordon, Jiangang Hao, Saad Khan, Ida M. Lawrence, Jessica Mislevy, Maria Elena Oliveri, André A. Rupp, Valerie J. Shute, Sandip Sinharay, Eric M. Tucker, Alina A. von Davier, Howard Wainer, David M. Williamson, Duanli Yan, and Diego Zapata-Rivera.

As these scholars demonstrate, Bob's work has profoundly shaped contemporary thinking in educational measurement, validity theory, fairness, and opportunity to learn, among other areas. Throughout a forty-plus-year career, he authored—both alone and in multidisciplinary collaborations—books, chapters, and articles on the modeling of human cognition, learning, and culture. As his obituary in *Psychometrika* documents,

> *"He was a past-president of the Psychometric Society, an elected member of the National Academy of Education, an elected fellow of the American Educational Research Association, and served on National Academies of Science committees concerning assessment, instruction, and psychology. Among other honors, he received the Award for Significant Contribution to Educational Measurement and Research Methodology from the American Educational Research Association, the E.F. Lindquist Award from the American Educational Research Association and ACT, the National Council on Measurement in Education's Award for Exceptional Achievement in Educational Measurement (five times), and awards for career and lifetime contributions from the Psychometric Society and the National Council on Measurement in Education." (Levy & Almond, 2025, p. 1)*

Perspectives and Themes

The first set of contributions in this collection explores how Bob opened new horizons in assessment design and in the optimization of complex measurement processes. T. J. Elliott, Russell Almond, Duanli Yan, Jiangang Hao, Diego Zapata-Rivera, Randy Bennett, Howard Wainer, Alina von Davier, and David M. Williamson each describe how Bob's frameworks and questions sparked innovations that continue to ripple through our technical and methodological foundations. These authors note that some of Bob's contributions informed innovations in measurement, such as exploring the role of games in educational assessment and building validity arguments for personalized assessment. Valerie Shute outlines Bob's influence, enabling stealth assessment, thus enhancing efforts to identify the cognitive, affective, and contextual factors that influence learning, building, and testing intelligent tutoring systems. Diego Zapata-Rivera highlights how Bob's work guided the analysis of accessibility features, the supports provided for English language learners, and learning and assessment in digital learning environments. Ida Lawrence and Eva Baker reflect on how Bob's approach traveled through institutions—at Educational Testing Service (ETS, where he served as Frederic M. Lord Chair in Measurement and Statistics from 2011 until his retirement in 2021), at the UCLA Center for Research on Evaluation, Standards, and Student Testing (CRESST, where he published approximately 29 technical reports), and beyond—showing how his designs for assessment reshaped research programs, partnerships, and the day-to-day work of R&D teams. T. J. Elliott underscores how Bob's insistence that claims still matter equips us to navigate an era of misinformation, while Duanli Yan and David Williamson trace how evidence-centered design (ECD) and Bayesian networks continue to shape contemporary scholarship.

A second group of authors turns to situative and sociocognitive perspectives, approaches Bob helped illuminate to re-imagine traditional measurement. Russell Almond, Maria Elena Oliveri, and Edmund W. Gordon reflect on how these perspectives reoriented assessment around meaning-making, context, and the lived experience of learners and communities. This work is illustrated in Bob's *Sociocognitive Foundations of Educational Measurement* (Mislevy, 2018) and culminated with his co-authored chapter on socioculturally responsive assessment (Mislevy et al., 2025). This work had just been published when Bob passed.

The third theme, by far the most recurrent, centers on Bob's profoundly human presence as a mentor, collaborator, and friend. Maria Elena Oliveri, Eric Tucker, Jiangang Hao, Sandip Sinharay, Valerie Shute, André Rupp, Michael Feuer, Jessica Mislevy, Eva Baker, Ida Lawrence, David Dorsey, and David Williamson show how Bob infused scientific pursuits with patience, curiosity, intellectual generosity, and humility. Drew Gitomer, in a reflection aptly titled "In Gratitude: The Gift of Bob Mislevy," underscores how Bob's questions and

collaborations quietly redirected entire careers and research programs. These pieces remind us that scholarship is animated not only by ideas but by the relationships that support them. We see those relationships in play in the 2021 special issue of this journal, *Meeting the Challenges of Workplace English Communication in the 21st Century* (Oliveri et al., 2021a; Oliveri et al., 2021b; Oliveri et al., 2021c; Oliveri et al., 2021d). Bob's frameworks underpinned sociocognitive models of communicative competence and guided the design of tasks that honor the complexity of real workplaces.

A fourth contribution, from John Behrens, offers a meditation on cognitive recalibration, how Bob nudged us to think differently, pose different questions, and reorient our intellectual compasses in ways we often recognized only in hindsight.

We are grateful to each contributor for sharing these heartfelt, keenly observant memories.

Finally, Saad Khan brings forward a theme that threads through Bob's career even as technology transformed around him: the evolving role of digital tools and computational infrastructures in shaping what assessment can do—and what it ought to do.

Across these themes, what emerges is more than a portrait; it is a transformation of minds and hearts shaped by Bob's example. This collection honors the ideas he championed, the communities he built, and the humane, quietly transformative way he shaped our world.

The volume opens with a family remembrance from his oldest daughter, Jessica, who shares perspectives about a loving and devoted family man, followed by tributes from long-time colleagues; collaborators, and mentees; leaders in assessment and innovation; and scholars in allied fields who came to know Bob as a "soulful psychometrician" and a "brilliant soul." Some contributions began as memorial remarks delivered at services in College Park and Princeton; others were written specifically for this issue. Together, they form portions of a mosaic of Bob's life—from the backyard campouts and "Bob jokes" of family memory to the technical breakthroughs and institutional transformations that reshaped our understanding of assessment in education.

We are grateful to each contributor for sharing these heartfelt, keenly observant memories. In the same way that Bob never failed to recognize those who inspired and worked alongside him, we want to thank the family members, friends, colleagues, and mentees who have contributed their reflections. This collection is a product of a community, just as Bob always believed that science (and particularly assessment) was a team activity. The presence of so many voices is itself evidence of the profound influence he had on all of us.

Across these pages, a consistent portrait of Bob's character emerges. Edmund Gordon opens one of his two contributions by addressing us as "family, friends, and colleagues of the late Professor Robert Mislevy, a gentle person and giant scholar," then reflects, in a line that many of us have carried with us, "On my scale of human decency, no one stands higher than he." Others echo the same sentiment in different words. One colleague writes that Bob was "the gentle, lifelong refutation" of the myth that brilliance must have sharp edges, proving that "the most effective tool for solving intractable problems isn't bombast, but collaboration." Jessica observes, with the clarity of a daughter who watched both his career and his parenting up close, that "for most whose career achievements even come close to rivaling those of my father, it's accompanied by ego. . . . Neither was the case with Bob Mislevy. Despite his brilliance, my father was modest and humble. Despite his success, my father was present and hands-on."

Conceptual Legacy

Those qualities of gentle intellectual rigor and deep humility shaped every aspect of Bob's work. As many readers of *JWA* will recall, his scholarship transformed how we conceptualize, design, and validate assessments. Through frameworks such as ECD (Mislevy et al., 2003) and his sociocognitive approach to measurement (Mislevy, 2018), Bob helped redefine assessment as a system of evidentiary reasoning, one that must remain responsive to the cognitive, social, and ethical dimensions of human development. He emphasized that assessments are not merely mechanisms for producing scores but structured arguments about what people know and can do, grounded in context and purpose. It is no accident that ECD has become a foundational grammar for work at the intersection of writing, discourse, and analytics.

Assessments are not merely mechanisms for producing scores but structured arguments about what people know and can do, grounded in context and purpose.

At the core of this assessment evolution is a simple yet demanding conviction: any score must be a defensible story about a learner, understood as inseparable from the learner's context in which performance occurs. For those of us working in writing analytics, that insight resonates deeply. Our field likewise treats texts, discourse, and data traces as evidence in a larger narrative about writers' abilities and circumstances. It is fitting, then, that this collection, focused on a scholar who saw evidence as a story, should itself be a collection of stories, each reminiscence a piece of evidence in a broader validity argument about Bob's life and legacy. T. J. Elliott's remembrance, "Claims Will Always Matter," picks up this thread explicitly, arguing that Bob's architecture for evidentiary reasoning offers tools not only for assessment but for democratic life, where the stakes of our claims extend far beyond test scores.

One of Bob's key contributions is also his ability to look beneath the machinery of assessment and notice the epistemic scaffolding that quietly shaped what could be known, inferred, and valued. John Behrens captures this as the moment when Bob "lifted the veil"—when the hidden architecture of reasoning, evidence, and modeling suddenly came into view, revealing assessment as a coherent act of meaning-making rather than a technical ritual. Once the veil was raised, there was no returning to the earlier, more mechanical view. Bob showed that every measurement argument rests on an underlying story about learners and the world, and he gave the field the conceptual tools to make those stories explicit, examinable, and improvable.

At the heart of this revelation was Bob's insistence that cognition is never disembodied. Before "situated cognition" gained wide traction, Bob articulated how knowing and doing unfold within contexts, communities, histories, and tools, and how assessment must be built to respect that ecology. His work reframed epistemology as a living practice: a system of commitments about what counts as knowledge, how evidence connects to claims, and how humans interpret and act on information in real settings. The frameworks he developed did not merely improve assessment design; they enhanced the field's sense of its own purpose.

His legacy endures not as a monument but as an ongoing invitation to think with greater clarity, humility, and imagination.

Today, scholars and practitioners work inside the intellectual clearings Bob established, designing assessments that are more transparent, more coherent, and more attuned to the richness of human thought. His legacy thus endures not as a monument but as an ongoing invitation to think with greater clarity, humility, and imagination.

Bob also helped the field imagine actionable assessments; tasks that revealed understanding through action rather than recitation. He championed performance assessments, simulations, scenario-based tasks, and game-like environments that allowed reasoning, collaboration, and strategic decision-making to unfold dynamically. These designs treated

assessment as an unfolding encounter rather than a static checkpoint, capturing the grain of thinking as it moves through authentic challenges. Under Bob's guidance, the field discovered how carefully engineered tasks could make visible the cognitive, social, and contextual dimensions of competence that traditional formats left untouched.

As technology evolved, Bob widened the horizon even further. He opened vistas for both technology-enhanced assessment and assessment-enhanced technology—a dual transformation Behrens later described as a revelation in its own right. Interactive simulations, telemetry-rich digital tasks, automated scoring, and stealth assessment frameworks all derive their validity arguments from the groundwork Bob laid. These innovations continue to echo across the education and human-performance sectors, where his ideas guide everything from adaptive learning systems to workplace readiness evaluations.

Inspirational Collaboration and Collegiality

Mentorship was never prescriptive; it was invitational—rooted in deep listening, genuine curiosity, and guidance that felt like your own discovery.

If Bob's concepts changed how we think about assessment, his way of being set a model for how we might work together. Over and over, contributors describe his mentorship as "never prescriptive; it was invitational." Maria Elena Oliveri writes that he "approached every conversation with humility and care, listening deeply and asking questions that gently but profoundly shifted your thinking." Jiangang Hao remembers that "he did not lecture or impose. Instead, he listened carefully, took my questions seriously, and guided me toward insights that felt like my own discoveries." For Duanli Yan, who worked with Bob for more than three decades, that style of guidance was life-changing: "I would not be where I am today without him, and I am eternally grateful for his mentorship and friendship."

Mentorship for Bob was inseparable from collaboration. John Behrens tells a story that has become emblematic: At the end of an incandescently brilliant talk, after laying out a complex model, Bob would do something unusual. He would meticulously thank collaborators—from software engineers and classroom teachers to game designers and mechanics—"ensuring the work was seen not as his own, but as the product of a community." In his actions, science was a team sport. That habit of recognition and appreciation expressed a methodological stance. Bob believed that high-quality ideas emerge when people reason about evidence together, across roles and disciplines. He worked, as Edmund Gordon puts it, "at the borders and edges, where ideas can mingle and be better heard." He invited linguists, teachers, designers, technologists, psychometricians, and students into the same conversations, and he treated everyone's questions as worth his time. It is no surprise that so many of the authors in this issue are themselves now leading work at those borders, in computational psychometrics, game-based learning, caring assessment, and workplace communication.

Science was a team sport, a community effort where ideas flourished as people reasoned together across roles, disciplines, and lived experiences.

These reminiscences also remind us that Bob's scholarship and collegiality were rooted in a rich family life. Jessica's opening piece offers a portrait of her father as "incredibly kind, patient, and giving"—the dad who combed out tangled hair, drove late-night carpools, made space for the pool table to become a Lego village, and, on a crowded London subway platform, shouted, "COMING THROUGH!" to make sure his family stayed together. She calls his fifty-year marriage with Robbie "his most important collaboration in life," noting that they were "truly best friends and partners in life," providing "a strong, positive model for what marriage can be." Those stories of campouts, classic rock debates, silent films, and the red Mustang convertible help us see how the same patience, steadiness, and playfulness that defined Bob at home also shaped his presence as a colleague and mentor. As one contributor writes, "Brilliance and humility, insight and kindness, could inhabit the same person. In Bob, I saw a clear expression of that ideal." Family, broadly conceived, was at the center of Bob's life.

Taken together, the contributions in this collection testify that Bob's greatest legacy lies not only in the frameworks and methods he left us, but in the people he formed. Many authors explicitly frame their tributes as a way to say thank you for opportunities, conversations, and "moments of humanity" that shaped their careers. They describe how his questions continue to guide their research, how his example shapes their own mentoring, and how his insistence on fairness, context, and care has become part of their scholarly DNA. In Howard Everson's words, "Robert Mislevy left us all many trees to sit under. My students and I will take every opportunity to enjoy the shade."

Each reminiscence is a seed planted in someone's life, now blooming in its own distinct form.

We observe that the contributors to this volume are tending to aspects of Bob's body of work, sustaining traditions that scholars will study for years to come. This collection, then, is not a conclusion to Bob's story, but a new chapter in it, a modest corner of a growing, living archive of Bob's impact that will continue to grow as long as we, his community, carry forward his ideas and ways of working. Each reminiscence here is a seed he planted in someone's life, now blooming in a distinct form. As you read, you become part of that living legacy: reflecting on his ideas, smiling at his jokes, perhaps finding inspiration for your own next project. In this way, remembrance becomes forward-looking.

The larger tribute is to mentor with extra patience, collaborate more generously, and infuse a bit more 'rigorously applied kindness' into our everyday work.

As editors, we are profoundly grateful for the time we had with Bob Mislevy. It was a privilege and a blessing to know him, to learn from him, and to be uplifted by his example. The field of educational measurement (and the community of writing analytics) owes him an incalculable debt for giving us new languages and models to understand learning, and for showing us that behind every data point is a person. If we honor Bob only with words, we miss the heart of his example. The larger tribute is to mentor with extra patience, collaborate more generously, and keep asking the deep questions he posed, to infuse a bit more of his "rigorously applied kindness" into our everyday work.

Bob Mislevy was a scholar who not only advanced disciplines but also guided and inspired communities. We end this introduction with profound gratitude—for his ideas that challenged us and his kindness that sustained us. The pages that follow testify to a simple truth: each in our own way, we were blessed to know him, and we are better for it. His light lives on in every mind he touched and every heart he inspired.

References

Levy, R., & Almond, R. G. (2025). Obituary: Robert J. Mislevy (1950–2025). *Psychometrika*, 1-6. doi:10.1017/psy.2025.10049

Mislevy, R. J. (2018). *Sociocognitive foundations of educational measurement.* Routledge.

Mislevy, R., Almond, R. G., & Lukas, J. F. (2003). A brief introduction to evidence-centered design (Research Report RR-03-16). Educational Testing Service. https://files.eric.ed.gov/fulltext/ED483399.pdf

Mislevy, R. J., Oliveri, M. E., Slomp, D., Wolf, A. C. E., & Elliot, N. (2025). An evidentiary-reasoning lens for socioculturally responsive assessment. In R. E. Bennett, L. Darling-Hammond, & A. Badrinarayan (Eds.), *Socioculturally responsive* assessment. pp. 199–241). Routledge. https://doi.org/10.4324/9781003435105-13

Oliveri, M. E., Mislevy, R. J., & Slomp, D. (2021a). Principled development of workplace English communication part 1: A sociocognitive framework. *The Journal of Writing Analytics*, *5*, 34–70. https://doi.org/10.37514/JWA-J.2021.5.1.02

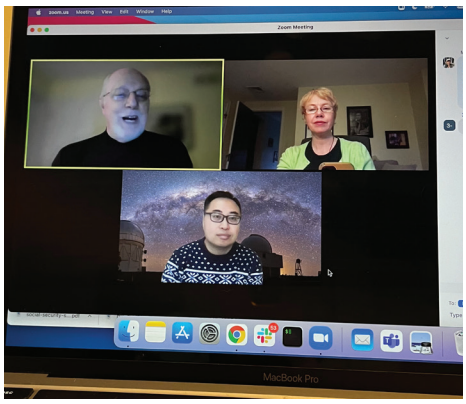
Oliveri, M. E., Slomp, D. H., Elliot, N., Rupp, A. A., Mislevy, R. J., Vezzu, M., Tackitt, A., Nastal, J., Phelps, J., & Osborn, M. (2021b). Introduction: Meeting the challenges of workplace English communication in the 21st century. *The Journal of Writing Analytics*, *5*, 1–33, https://doi.org/10.37514/JWA-J.2021.5.1.01

Oliveri, M. E., Slomp, D., Rupp, A. A., Mislevy, R. J. (2021c). Principled development of workplace English communication part 2: Expanded evidence-centered design and theory of action frameworks. *The Journal of Writing Analytics*, *5*, 71–108. https://doi.org/10.37514/JWA-J.2021.5.1.03

Oliveri, M. E., Slomp, D., Rupp, A. A., & Mislevy, R. J. (2021d). Principled development of workplace English communication part 3: An integrated design and appraisal framework. *The Journal of Writing Analytics*, *5*, 109–141. https://doi.org/10.37514/JWA-J.2021.5.1.04

A Legacy in Photographs

Anne Ruggles Gere, Norbert Elliot, Frances Ward, Rhonda Maneval, Robert J. Mislevy, Diane Kelly-Riley, Jeffrey Hogrefe. Modern Language Association, New York, January 7, 2018



Bob, Jiangang Hao, and Alina von Davier

Bob and David Williamson

Jiangang Hao and Bob

Maria Elena Oliveri and Bob

Bob, Duanli Yan, and Alina von Davier

Alina von Davier and Bob

- Bottom left: Ruth, Richard, Robert and Joseph Mislevy (1964)
- Top left: Robert and Roberta Mislevy (1981)
- Top right: Top row left to right: Remi Gottheil, Meredith Hughes, Zachary Hughes, Robert Mislevy; Bottom row left to right: Luke Gottheil, Alexander Hughes, Elliot Hughes, and Jessica Mislevy (2024)
- Bottom right: Roberta, Meredith, Robert, and Jessica Mislevy (2006)

Jessica Mislevy

Director of Digital Learning & Technology Policy program
SRI International

Delivered at the University of Maryland

Atrium, Adele Stamp Student Union
3972 Campus Drive, College Park, MD

October 19, 2025

Remarks Delivered, Celebration of Life of Robert J. Mislevy

Hello everyone. My name is Jessica Mislevy, Bob Mislevy's oldest daughter, and I'm honored to share a few remarks on behalf of our family.

Many here can speak to my father's scholarship, his impact on education and assessment, and what a joy he was to work with and learn from. So, I won't even try. Instead, I would like to reflect on his life as a loving and devoted family man.

Once, as a young child, I was playing with my dad in the backyard. He was lifting me up and spinning me around. I distinctly remember thinking, "Wow, my dad is the strongest man in the world!" Well, with more life experience, I came to realize that, no, physically, he was not the strongest man in the world—but when it comes to strength of character, he was arguably so.

He was an incredibly kind, patient, and giving person—so too as a dad. I remember him combing the knots out of our long hair before bed with No More Tangles spray. How he let me and Meredith commandeer the pool table in the rec room for months with the extensive Lego village we constructed because the green felt made for the perfect faux grass. And there was the time he and another neighborhood dad set up a tent in the backyard so the kids could have a campout; one by one we children abandoned the tent and retreated to our beds, leaving only the dads to sleep uncomfortably outside.

He drove us to and from gymnastics practices, piano lessons, and choir rehearsals. Some of my favorite memories with my dad are from those one-on-one times in the car. This is where he introduced us to classic rock. Artists like Led Zeppelin, Bob Dylan, Eric Clapton, and of course, The Beatles. Debates about who was the best rock guitar player, and whether we agreed with the rankings on the radio station's countdown of the best songs. Perhaps his influence is why, in fourth grade, when we were asked to play our favorite song for music class, I brought in The Who's "Baba O'Reily," while the rest of my class played New Kids on the Block and MC Hammer.

Much bemoaned, he also made us watch classic movies—not just ones filmed in black and white, but several from the times before "the talkies" even had sound! *Citizen Cane*, *Metropolis*, *The Cabinet of Dr. Caligari*. At least he let us watch that last one on fast forward. Ever the educator, he'd talk about what made the movie famous—like how a lighting or camera technique was used for the very first time and copied by filmmakers ever since. I will say my love of Hitchcock mysteries stuck.

He also helped us with our homework. You'd think that having a statistician as a father would be a big perk when you're stuck on a fifth-grade math assignment, but I recall instances where I debated whether to ask him for help. I'd have to leave soon for gymnastics practice and my dad's explanations were never quick—usually culminating with something like, "Now when you get to calculus, you'll see that . . . " Just what every 10-year-old wants to hear.

My dad was funny, and boy, did he love a pun. *Dilbert* and the *Far Side* comics were up there, too. One family friend dubbed the jokes he made "Bob jokes"—that is, dad jokes that were actually smart and interesting. A one-liner of his that still makes me chuckle came from a time he and another dad had taken all five of us kids out for a meal. My dad accidentally spilled his own drink all over the table and said, "Well, when you go out to eat with kids, you have to expect spills." My sister and I may not have always fully appreciated his humor, though. For back-to-school night in elementary school, kids were encouraged to leave a note for their parents on their desk; apparently, the one Meredith left specifically instructed my dad not to tell any of his "so-called" jokes.

One of the things that made my father such a great dad and grandpa was that no matter his age, he was still a kid at heart—a little goofy, at times mischievous, and often too smart for his own good. I enjoyed hearing stories about "Little Bobby"—like the time in elementary school when he finished his math assignment early and got in trouble for prancing around the classroom, distracting other students. His teacher tied him to his chair; in those days, you could do that. He said he remembered sitting there thinking he could get out of the binds if he really wanted to but knew it probably wasn't a good idea to try. And the stories about all the odd jobs he had in his teens, including delivering wedding cakes and working at the macaroni factory—playing tricks on new employees and breaking his finger on the industrial equipment in a moment of pure, "Dude, watch this" stupidity. I'm sure those of you who knew my father in high school and college have plenty of stories you could tell.

Even so, my dad always took his commitments to his family seriously. He was too level-headed to go through a midlife crisis. The closest he got was when he decided he no longer wanted to wear boring ties, and instead, he started donning ones with bold colors and abstract prints. And after decades of driving practical family cars, I was glad our mom finally talked him into buying the red Mustang convertible he'd wanted since his youth. He made such sacrifices for us, and so selflessly. Like when he took the job at the University of Maryland and commuted hours by train every week for years so Meredith and I could finish high school in New Jersey.

Though my dad didn't officially become a professor until the early aughts, he was always a natural and gifted educator. In fact, both my mom and dad had quite the reputation for explaining things. This instilled in us the value of education and inspired a love of learning. Some may recall that my father often used his hands to illustrate and emphasize his points. On one of the treks between our house in New Jersey and the University of Maryland, Meredith and I were in a separate car following my parents in their car. Here we are on a high-speed interstate, and their car just starts going slower and slower. Meredith and I are wondering what the heck is going on, but we felt it would be inappropriate to pass our parents. Finally, through the back window, we see one of my dad's familiar hand gestures, and instantly, we knew why he was driving so slowly—he was explaining something!

Little did we know that measurement, statistics, and evaluation would become a family business. Meredith and I didn't start there, nor did we intend to follow in his footsteps (which would be impossible shoes to fill anyway). I majored in criminology, Meredith in linguistics, but in time, we each organically found our way to psychometrics. Without pushing, he laid a foundation for the curiosity and critical thinking that are so fundamental to assessment. It's true that our family actually sat around the dinner table discussing construct-irrelevant variance. He even enjoyed watching the reality show *Project Runway* with us because he said it was a great example of design under constraints, revealing brittle versus deep expertise. The show's elimination catch phrase—"In fashion, one day you're in, and the next day you're out."—he jokingly rewrote: "In statistics, 95% of you will be in, and 5% of you will be out."

My dad was so proud to hood us when we earned our graduate degrees from the department in which he taught. He also liked that it got him out of being on the exam committee each semester one of us was taking comps due to the conflict of interest. He ordered multiple copies of an issue of the *Journal of Educational Measurement*, in which we both had an article published. Once he attended a presentation I was giving at NCME and asked if I wanted him to plant any questions during the discussion. I told him it would be great if he could just nod along in agreement with everything I was saying. Ended up that he leaned against the light switch and turned off the lights during my presentation. Seriously though, anytime someone in the business asks me, "Are you related to Bob Mislevy?" I'm proud to answer that I'm his daughter—sure, because of the prominence of his work, but more so because of the caliber of his character.

Though my dad had just a few precious years with his grandsons, he managed to seed that inquiry mindset in the next generation, introducing them to magic tricks and science experiments. He loved helping them learn, grow, and explore the world around them. He loved telling stories about the funny things they would say and do, and marveling how their insights revealed their understanding of the world—

the core of assessment in a way. One time, when Alex and Elliot were maybe three years old, we were all over at my parents' house, and the boys asked if they could push the buttons on my parents' answering machine. In response, the boys heard, "No!" from all the adults but one—Grandpa, who said, "Sure!" We all turned to look at him, and he goes, "What? That's why they like me so much."

Now any remarks about my dad as a family man would be woefully incomplete without mention of his most important collaboration in life—the one he had with our mother, Robbie, his beloved wife of 50 years. Meredith and I learned so much by the example they set. They provided us with such a strong, positive model for what marriage can be, a relationship to aspire to. They were truly best friends and partners in life, treating each other with love, kindness, and respect. Having married young, my parents grew up with each other and into new life stages. By championing each other's passions and aspirations, they grew stronger together instead of growing apart. They supported one another over the years, such as when my mom worked full-time so my dad could finish graduate school, or when my mom went back to school for her teaching degree. Of course, there were times they disagreed, but they never lost sight of the fact that they were on the same team. They loved to explore new places and hobbies, like playing volleyball in their 30s or traveling to photograph national parks and wildlife in their 60s. They never stopped wanting to learn from each other. They never stopped laughing together. To see them with one another, you knew you were in the presence of something exceptional.

Lest you think my father was too perfect, he had his endearing quirks. Mild color blindness sometimes made for interesting outfits. And as smart as he was, he had a terrible knack for inadvertently spoiling movies and TV shows. Perhaps the most infamous example of this was when he asked, "Now, the last time you watched *Game of Thrones*, was John Snow alive?" He still thought this question was clever, if you know the plot point it references. My dad also had his absent-minded professor moments—like when he injured his foot after inadvertently jogging in my mother's sneakers, which were sizes too small. And he occasionally had strange ideas of social etiquette—like the time at a formal work dinner when he ordered crab cakes as his appetizer *and* the crab cakes as his entrée. Or when he wanted to wear his expensive academic regalia to the Renn Fest, arguing that the ensemble was historically accurate.

There's one story from a trip our family took to London, now more than 20 years ago. My mom, sister, and I were accompanying my father as he traveled abroad for work, and in his free time, we rode the Tube around to see the sights. There, on a crowded subway platform, we nearly got separated when boarding the train—a big deal when you're in a foreign country and before the prevalence of cell phones. Suddenly, we hear my mild-mannered father yell, "COMING THROUGH!" as he pushed his way through the crowd to join us in the train car. While we've laughed about it many times over the years, this one moment just exemplifies what he would do to protect his family.

When my dad received his diagnosis, as much as he wanted more time—with his grandsons, especially—he said he was at peace if this was to be it. That he'd had a good, long life and was proud of what he accomplished professionally, of his loving marriage with our mom, and of the children he raised. He gave us as much time as he could, despite how difficult the last two years must have been for him since our mother passed. He saw us to our paths professionally, as wives and as mothers. We take some comfort in thinking of Bob and Robbie reunited.

For most whose career achievements even come close to rivaling those of my father, it's accompanied by ego. Or their families suffer the costs as they devote themselves to their work. Neither was the case for Bob Mislevy. Despite his brilliance, my father was modest and humble. Despite his success, my father was present and hands-on. Perhaps this is why his loss is felt so profoundly both personally and professionally. Each in our own way, we were blessed to know him, and we were better for it.

So let me close by saying that we love you, Dad, and can never thank you enough. We celebrate your life and strive to live ours by your example. Today and always.

Edmund W. Gordon

John M. Musser Professor of Psychology, Emeritus, Yale University
Richard March Hoe Professor, Emeritus of Psychology and Education, Teachers College, Columbia University.

Delivered at the University of Maryland

Atrium, Adele Stamp Student Union 3972 Campus Drive, College Park, MD

October 19, 2025

Respect for the Situative

To the family, friends, and colleagues of the late Professor Robert Mislevy, a gentle person and giant scholar.

I am here at his memorial today because of my love and respect for his scholarship and humanity. There are many good things that can be said for Bob, and not much that is less than good. Perhaps the worst is his willingness to compromise! Bob often reminded me that his strategy was to work the borders and edges, where ideas can mingle, to effect change. "Avoid head-on challenges," he would argue. He thought we could make more progress that way, showing respect for the position of the other even while destroying it. But Bob also thought the laws of the Universe are too relational to support absolutes. To understand and certainly to resolve most problems, we simply *have* to consider the multiple perspectives from which a position can be considered. An argument that is respectful of the situative and that accommodates the possibility of the other, even the contradictory, may be more persuasive—and it may be closer to reality.

In divinity school, I was taught that this was a weak position from which to argue. But who among us can claim, with any evidence, that Robert Mislevy was a weak scholar? Who among us will admit to being stupid enough to challenge the reasoning of Bob Mislevy?

What does the danger sign read? *At Your Own Risk!*

Bob was the consummate scholar. Always on target. Always thoughtful. Always closer to being correct than to being wrong. Still, no matter what, he was respectful and kind to others.

Professor Robert Mislevy was among the finest human beings I have known in my 104 years of walking beside some of the most able and decent human beings on earth. He moved us forward by working at the borders and edges where ideas can mingle and be better heard.

May his spirit remain with us while his body rests in peace.

Maria Elena Oliveri

Research Associate Professor
College of Engineering
Purdue University

Transforming Educational Measurement Through Theory, Evidence, and Mentorship

It is with a heavy heart and deep gratitude that I write this article recognizing the extraordinary contributions of Dr. Robert J. Mislevy. Bob's recent passing is an immeasurable loss, not only to the field of educational assessment but to all of us who had the privilege of working alongside him, learning from him, and being shaped by his quiet brilliance and generosity of spirit. His absence is deeply felt, and it is difficult to find words that fully convey how much he meant: to me personally, and to the many scholars, students, and colleagues whose lives he touched.

Bob profoundly reshaped how we conceptualize, design, and enact assessment systems. Over the course of his career, he built an intellectual legacy rooted in principled reasoning, methodological rigor, and an unwavering commitment to ensuring that assessments serve rather than constrain learning. His influence extends far beyond psychometrics and educational measurement. He helped redefine assessment as a system of evidentiary reasoning, one that must remain responsive to the cognitive, social, and ethical dimensions of human development.

My own work, like that of many in the field, has been deeply and enduringly shaped by Bob's scholarship, mentorship, and friendship. In the years we spent collaborating—across projects, papers, and long, searching conversations—I came to know a person of rare depth and clarity.

What made working with Bob so meaningful was not just the depth of his ideas but the spirit with which he shared them. He approached every conversation with humility and care, listening deeply and asking questions that gently but profoundly shifted your thinking. His mentorship was never prescriptive; it was invitational. He encouraged you to ask better questions, to write more clearly, and to think more responsibly about the impact of your work.

Bob had a remarkable gift for theoretical writing. His ability to weave together insights from psychometrics, learning theory, philosophy of science, and sociocultural perspectives gave his work a kind of intellectual architecture that was at once rigorous and expansive. But what made our collaboration so special was how his theory opened doors for me to do the kind of applied work I valued most, work that is grounded in real educational contexts and in the lives of learners and educators. As someone who came to assessment from the lived experiences of being a classroom instructor and a literacy mentor and working with adult learners in workplace settings, I often sought tools and frameworks that could honor the complexity of those settings while still supporting valid and reliable assessments. Bob's ideas gave me that foundation. They allowed me to design systems that were principled, but also adaptable to the messy, rich realities of learning and communication in the real world.

Together, we brought different but deeply complementary strengths to our joint work. He had a remarkable gift for bridging theory and practice, for posing foundational questions with humility and precision, and for creating space where multiple disciplinary voices, educators, cognitive scientists, and psychometricians could come together in the shared pursuit of more valid, more fair, and more inclusive assessments. In turn, I was able to bring those ideas to bear in practical applications that mattered to me: designing assessments that captured the richness of workplace communication, supporting multilingual learners navigating complex literacy demands, or creating frameworks that made explicit not just what we measured, but why it mattered, and for whom.

Our work was grounded in the idea that assessment should serve learning, not constrain it, and that validity is not just a technical property but an ethical responsibility. That shared commitment—and the dialogue it enabled between theory and lived experience—made our collaborations uniquely meaningful to me. With Bob, the work was never abstract. It was always connected to people's lives, to real consequences, and to the hope that our tools could contribute, even in small ways, to more just and responsive educational systems. His presence made the work richer, more honest, and more human. I am profoundly grateful for all I learned from him, and for the friendship, insight, and generosity he brought to every collaboration.

Bob changed the way I think—not just about assessment design but about what it means to do this work with integrity. He modeled a kind of intellectual leadership that was rigorous yet never rigid, always open to new ideas, and grounded in deep respect for learners and their diverse experiences. His insights continue to shape how I build systems, frame research questions, and mentor others. I carry his influence with me every day.

This article is a small tribute to Bob's extraordinary legacy. It reflects the wide reach of his ideas, the depth of his impact, and the enduring admiration of those who had the privilege to call him a collaborator, mentor, and friend. For me, this is also a way to say thank you, to honor the profound imprint he left on my life and my work.

We miss him dearly.

An Intellectual Legacy

Robert J. Mislevy's work transformed the field of educational assessment, reshaping both its theoretical foundations and its practical applications. Over a career spanning decades, Bob developed frameworks that brought together psychometric rigor with cognitive and sociocultural insights, opening up new possibilities for how we conceptualize, design, and evaluate assessments. He is perhaps most widely known for his pioneering work on evidence-centered design (ECD), a systematized approach to assessment design that aligns evidence with the claims we wish to make about learners. ECD formalized a structure for ensuring that assessments are not ad hoc collections of tasks but principled tools for reasoning about knowledge and performance.

But Bob's contributions extend far beyond ECD. He brought to the field a rare epistemic clarity, ensuring that every measurement rests on assumptions about what we value, what we observe, and how those observations support inferences. His writings on validity, grounded in Toulmin's argumentation model, presented assessment as an evidentiary process shaped by context, purpose, and interpretation. Through this lens, validity is not simply a statistical property; it is an ongoing argument, a reasoned claim that must be scrutinized through both technical and ethical perspectives.

Bob also led the way in integrating cognitive, social, and cultural dimensions in assessment design. His sociocognitive approach emphasized that learning and knowing are not isolated acts; rather, they emerge in interaction with tools, communities, and settings. This insight pushed the field toward more equitable and responsive assessments, especially for linguistically and culturally diverse learners. Through projects spanning digital simulations, language assessments, and socioculturally responsive models, Bob's work has had a wide-ranging impact on education systems and researchers across the globe.

Robert J. Mislevy's Lasting Contributions to Educational Measurement

Robert J. Mislevy's scholarship has deeply shaped modern educational measurement, with his ideas leaving an indelible mark on theory, practice, and policy. Through decades of work, he consistently advanced a vision of assessment as a process of *reasoned inference*, grounded in cognitive and sociocultural theory, and made rigorous through statistical modeling. His contributions—from ECD to Bayesian networks, from simulation-based assessment to sociocognitive perspectives on validity—have

guided the field toward more principled, transparent, and inclusive forms of assessment. Few figures in educational research have been as prolific or as influential in redefining how we conceptualize and implement assessments in various contexts.

At the heart of Mislevy's work is the understanding that assessments are not simply tools for generating scores; they are structured arguments about what people know and can do, based on observable behavior. This foundational idea runs through his most influential work on ECD. First articulated in the late 1990s and formalized in a series of highly cited papers and reports (e.g., Mislevy et al., 2003), ECD provides a systematic framework for designing assessments that begins with clear claims about learners, identifies the evidence needed to support those claims, and then constructs tasks that will elicit that evidence. This framework introduced a layered architecture (student, evidence, and task models) that brought unprecedented clarity to the assessment design process. ECD is important to educational measurement because it shifted the focus from post-hoc validation of test scores to validity by design. It challenged test developers to embed reasoning about validity into the earliest stages of assessment design, making it easier to align tasks with construct definitions, ensure fairness, and accommodate new forms of evidence (e.g., digital traces, discourse data). Today, ECD principles are foundational in the development of simulation-based assessments, language tests, and writing analytics. ECD has helped reconceive psychometrics not as a separate phase of test development but as part of an integrated evidentiary argument. This influence is not just conceptual but visible in institutional practice. Indeed, testing companies and educational technology organizations routinely use ECD as a standard for ensuring alignment and defensibility in their assessment systems.

In parallel, Mislevy was also instrumental in bringing Bayesian psychometric modeling and graphical models into mainstream educational measurement. Long before Bayesian methods became standard in applied psychometrics, he demonstrated how Bayesian networks could be used to model complex relationships between student knowledge, task features, and observed responses. His work (e.g., with Almond and Steinberg) showed how probabilistic graphical models could support adaptive testing, intelligent tutoring systems, and diagnostic assessments where traditional unidimensional models were insufficient. The books *Bayesian Networks in Educational Assessment* (2015) and *Bayesian Psychometric Modeling* (2017, with Ken Levy) provide comprehensive accounts of these methods and remain core references in the field. These models are crucial because they enable the incorporation of prior information, contextual knowledge, and multi-source data in assessment, capabilities essential for modern, dynamic learning environments. Bayesian reasoning supports real-time inferences in digital environments and allows assessments to evolve as evidence accumulates. In this way, Mislevy's work expanded the boundaries of what educational assessments can capture and interpret, laying a foundation for complex, scenario-based, and formative assessments that adapt to learner needs.

Another key contribution of Mislevy's scholarship is his development of a sociocognitive perspective on assessment and validity. In works such as *Sociocognitive Foundations of Educational Measurement* (2018), Mislevy argued that test performance is never just a reflection of internal traits or skills; it is always shaped by the social, cultural, and linguistic contexts in which tasks are situated. Drawing on activity theory, situated cognition, and sociocultural learning theory, he showed how understanding assessment requires more than psychometric rigor; it demands attention to how learners engage with tasks, interpret them, and respond based on their cultural backgrounds, prior experiences, and linguistic resources. This sociocognitive lens has had a transformative impact on validity theory. Mislevy's view complements and extends Michael Kane's argument-based approach, emphasizing that validity is an ongoing process of justification grounded in context. His work has influenced fairness guidelines, accessibility research, and assessment development for linguistically diverse learners. It has also informed language and literacy assessments by highlighting the need to make purposeful decisions about task language: for example, simplifying language that is not construct-relevant while preserving complexity in domain-specific terms. These ideas have helped shift the field from viewing fairness as a checkbox exercise to embedding equity considerations at the heart of assessment design.

Mislevy also brought these theoretical insights into practice through his work on simulation- and game-based assessment, a growing domain where traditional test paradigms fall short. He demonstrated how the principles of ECD, sociocognitive theory, and Bayesian modeling could be used to interpret rich process data from interactive environments. Whether working on network simulations like Cisco's Packet Tracer or epistemic games that mimic real-world professions, Mislevy showed how assessments embedded in learning environments could generate meaningful evidence, so long as they were anchored in thoughtful design and sound inference models. His work in this area anticipated current trends in AI-enabled learning systems, where ongoing, unobtrusive assessment is crucial.

In summary, Robert J. Mislevy's work has transformed educational measurement by making it more principled, inclusive, and responsive to the complexities of learning. His contributions have provided both conceptual frameworks and technical tools for designing assessments that are not only psychometrically sound but also pedagogically meaningful and socially just. From ECD to Bayesian modeling, from sociocognitive validity to simulation assessment, Mislevy has consistently expanded what is possible in the field, earning him a place among the most influential assessment theorists of our time. His legacy endures not only in publications and citations but in the very language and logic that today's assessment professionals use to design, analyze, and interpret educational assessments.

In this reflection, I focus on two contributions that have shaped my own thinking and practice: (1) epistemic thinking and the idea of "ground truth" in assessment design, particularly in relation to activity theory; and (2) a sociocognitive approach to measurement that centers linguistic and cultural diversity. In both areas, my collaborations with Bob were intellectually rich and personally transformative, giving rise to frameworks and tools that continue to guide my work.

Epistemic Thinking and the Problem of Ground Truth

One of Bob's most enduring lessons for me was the need to be epistemically honest about what an assessment can tell us. He often asked, "What's the nature of the construct we're trying to measure? And how do we know what we know?" These questions may seem abstract, but they are central to grounding assessment design in real-world activity. In our conversations, Bob drew from activity theory and situated cognition to remind us that knowledge is not something we can measure like height or weight; it must be inferred from performances embedded in contexts that matter.

This view had a profound influence on how I approach assessment in complex domains, particularly in simulation-based environments. In our work on performance assessments and in projects such as *Next-Generation Assessments of 21st Century Skills* (Oliveri & Mislevy, 2019a, 2019b), we used ECD to build models that aligned claims with actions taken in simulated professional tasks. But more importantly, Bob challenged us to think about *ground truth*; that is, what does success really look like in this context, and how do we model it? He emphasized that evidence must not only be aligned with constructs but also with the underlying activity system: who is acting, with what tools, toward what ends. This thinking came to life in our joint work on workplace English communication, where we needed to assess not only whether someone *knew* English but whether they could *use* it in authentic, situated ways. Here, Bob's framing helped us make critical design decisions: which data to collect (e.g., discourse, log files), how to interpret it, and what claims we could responsibly make about communicative competence.

Sociocognitive Perspectives and Linguistic-Cultural Responsiveness

Another major influence from Bob's work was his clear call that measurement must reflect the sociocognitive realities of diverse learners. Assessment is never culturally neutral, he would say. It always reflects assumptions about language, knowledge, and valued ways of participating. This principle was central in our series of papers on workplace English communication (Oliveri et al., 2021a, 2021b, 2021c, 2021d), where we extended ECD to include sociocognitive frameworks that attended to real-world complexity.

Bob's influence was especially visible in our approach to task design and language use. We spent considerable time thinking about how to distinguish between language that should be simplified, terms or phrasing not central to the construct, and technical language that must remain complex because it is *part* of the skill domain. Bob encouraged us to be precise about these distinctions, always returning to the assessment's purpose and the meaning we wanted to support. He also pushed us to make our theory of action explicit: What are the intended and unintended consequences of using this assessment in workplace or educational settings? This thinking culminated in the development of an integrated design and appraisal framework that built on both ECD and sociocultural theory. Later, in our contribution to *Socioculturally Responsive Assessment* (Mislevy et al., 2025), we further extended these ideas, offering a lens to design assessments that acknowledge the political, linguistic, and social realities of the populations they serve. That work, to me, exemplifies Bob's unique ability to move fluently between formal psychometric modeling and deeply humanistic commitments.

More Than a Mentor

What made working with Bob so meaningful was not just the depth of his ideas, but the spirit with which he shared them. He approached every conversation with humility and care, listening deeply and asking questions that gently but profoundly shifted your thinking. His mentorship was never prescriptive; it was invitational. He made you want to ask better questions, to write more clearly, and to think more responsibly about the impact of your work.

For me, Bob's legacy lives not only in the frameworks and publications we created together, but in how I now approach scholarship, mentorship, and collaboration. He reminded me that assessment is not just about scores; rather, it's about sensemaking, argumentation, and care. It's about ensuring that what we design reflects the complexity of human lives and the values we hold.

His presence made the work richer, more honest, and more human. I am profoundly grateful for all I learned from him, and for the friendship, insight, and generosity he brought to every project. This work was never just academic with Bob. It was an opportunity to grow and learn.

Conclusion

In reflecting on Robert J. Mislevy's profound contributions to educational measurement, I am deeply grateful not only for the intellectual depth and rigor of his scholarship but also for the mentorship and guidance he has offered throughout my career. His work has fundamentally reshaped how we think about assessment, moving the field beyond traditional psychometrics toward a more nuanced, integrated understanding of evidence, cognition, and context. I have learned immensely from his frameworks, particularly ECD, and have drawn inspiration from the clarity and precision with which his work connects theory to practice. His generous mentorship, thoughtful feedback, and unwavering commitment to advancing the science of assessment have left an indelible mark on my own professional journey.

As I continue my work in assessment design and educational research, I know that his legacy will remain a guiding force. The principles he championed—rigorous evidence reasoning, alignment with cognitive theory, and attention to fairness and authenticity—will continue to shape my approach to the challenges and opportunities ahead. Robert Mislevy's influence is not just seen in the field's body of knowledge, but also in the many scholars and practitioners he has inspired. I count myself fortunate to be among them, and I carry forward his legacy with deep appreciation and enduring respect.

References

Almond, R. G., Mislevy, R. J., Steinberg, L. S., Yan, D., & Williamson, D. M. (2015). *Bayesian networks in educational assessment*. Springer.

Behrens, J. T., Mislevy, R. J., DiCerbo, K. E., & Levy, R. (2012). Design and discovery in educational assessment: Evidence-centered design, psychometrics, and educational data mining. *Journal of Educational Data Mining, 4*(1), 11–48. https://doi.org/10.5281/zenodo.3554641

Levy, R., & Mislevy, R. J. (2016). *Bayesian psychometric modeling*. CRC Press.

Mislevy, R. J., Oliveri, M. E., Slomp, D., Cropped Eared Wolf, A., & Elliot, N. (2025). An evidentiary-reasoning lens for socioculturally responsive assessment. In R. E. Bennett, L. D. Darling-Hammond, & A. Badrinarayan (Eds.), *Socioculturally responsive assessment: Implications for theory, measurement, and systems-level policy* (pp. 199–241). Routledge.

Mislevy, R. J. (2018). *Sociocognitive foundations of educational measurement*. Routledge.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives, 1*(1), 3–62. https://doi.org/10.1207/S15366359MEA0101_02

Oliveri, M. E., & Mislevy, R. (Eds.). (2019a). Challenges and opportunities in the design of "next-generation assessments of 21st century skills" [Special issue]. *International Journal of Testing, 19*(2).

Oliveri, M. E., & Mislevy, R. (2019b). Introduction to "challenges and opportunities in the design of 'next-generation assessments of 21st century skills'" special issue. *International Journal of Testing*, *19*(2), 97–102. https://doi.org/10.1080/15305058.2019.1608551

Oliveri, M. E., Slomp, D., Rupp, A. A., & Mislevy, R. J. (2021a). Principled development of workplace English communication part 2: Expanded evidence-centered design and theory of action frameworks. *The Journal of Writing Analytics, 5*, 71–108. https://doi.org/10.37514/JWA-J.2021.5.1.03

Oliveri, M. E., Slomp, D., Rupp, A. A., & Mislevy, R. J. (2021b). Principled development of workplace English communication part 3: An integrated design and appraisal framework. *The Journal of Writing Analytics, 5*, 109–141. https://doi.org/10.37514/JWA-J.2021.5.1.04

Oliveri, M. E., Mislevy, R. J., & Slomp, D. (2021c). Principled development of workplace English communication part 1: A sociocognitive framework. *The Journal of Writing Analytics, 5*, 34–70. https://doi.org/10.37514/JWA-J.2021.5.1.02

Oliveri, M. E., Slomp, D. H., Elliot, N., Rupp, A. A., Mislevy, R. J., Vezzu, M., Tackitt, A., Nastal, J., Phelps, J., & Osborn, M. (2021d). Introduction: Meeting the challenges of workplace English communication in the 21st century. *The Journal of Writing Analytics, 5,* 1–33. https://doi.org/10.37514/JWA-J.2021.5.1.01

Rupp, A. A., Gushta, M., Mislevy, R. J., & Shaffer, D. W. (2010). Evidence-centered design of epistemic games: Measurement principles for complex learning environments. *The Journal of Technology, Learning and Assessment, 8*(4). https://ejournals.bc.edu/index.php/jtla/article/view/1623

Eric M. Tucker

The Study Group

Modeling What Matters: Honoring Robert J. Mislevy's Parables of Evidentiary Reasoning

Robert "Bob" Mislevy often shared thought experiments. Imagine, he'd say, that an expert English-speaking chemist learning German and a native German undergraduate both take a German-language chemistry test. He'd ask, "If a test taker struggles writing an essay, is it their skill at chemistry or German?" The same low score, he'd explain, might tell two completely different stories—one of a language barrier, the other of a knowledge gap. With that simple parable, he'd raise a point: any score must be a defensible story about a learner, understood as inseparable from their context.

The mythology of brilliance often suggests it comes with sharp edges. Bob Mislevy, one of the most influential measurement scientists of his generation, was the gentle, lifelong refutation of that myth. He possessed a razor-sharp intellect paired with a profound generosity and decency, proving that the most effective tool for solving intractable problems isn't bombast, but collaboration.

To understand Bob's legacy, this remembrance focuses on the stories he told. He frequently taught by parable, or short stories that illustrate deeper insights, holding up for admiration teams of game designers, F-15 mechanics, and software engineers to reveal an architecture for how we communicate with evidence. Central to his work was the conviction that a test score is not a fixed trait, but a defensible story about learning.

Bob's most enduring contribution was not a static product but a dynamic process—a blueprint for how professionals with varied perspectives might solve problems too complex for any single person. Aspects of his approach were formalized in his seminal framework, evidence-centered design (ECD), which serves as a formal "grammar" for experts from disparate fields to reason together about evidence. He consistently shifted credit to his colleagues, believing the quality of human interaction is inseparable from the quality of the intellectual product. Through a lifetime of practicing kindness and decency, his greatest model was his own life, a demonstration that brilliance and humanity are not conflicting virtues, but mutually reinforcing forces.

I met Bob while serving as an Educational Testing Service/MacArthur Gordon Commission Fellow, and from that first conversation, I learned that he treated complex ideas and colleagues with a signature blend of warmth and penetrating humor. I was struck by his stillness when he listened intently. In a world that rewards quick answers and loud assertions, Bob was patient, curious, and humble. He was a brilliant statistician with a genius for collaboration, an art he practiced with a kindness that was both his method and his message. Through collaboration on the Gordon Commission Study Group and the Sociocognitive Foundations of Educational Measurement Reading Group, I spent time with my friend and mentor in a range of contexts over nearly 15 years. I appreciate the opportunity to offer this remembrance.

The Pedagogy of the Parable

Bob was renowned as one of the "most soulful psychometricians"; his kindness and character were said to be among the only things that exceeded his deep intellect. Though brilliant, he was remarkably humble and approachable, always shifting credit to his collaborators. His generous collaboration on tough challenges was legendary; decades after a project ended, he would ask after the high school teacher who wrote the activities or the young engineer on a project. He loved telling human stories because he believed that much of the work's value lay with the people who did it and the learners it served.

A Defensible Story About Learning

Bob saw a model of a *defensible story about learning* in the AP Art Studio portfolio—a story textured not by multiple choice, but by charcoal, clay, and light. The challenge was translating hundreds of hours of personal studio time into a fair, common standard. The "miracle," he'd say, was the rubric that artists, educators, technologists, and psychometricians built together. It was a bridge between the private language of a student's creativity and a shared, evidentiary claim, allowing raters to make inferences without sanding off the very edges that made the art unique.

The Power of Wrong Answers

Bob loved the drama of a well-written math problem. Consider two students solving 4 and 1/10 − 2 and 8/10. One freezes, considers the problem, regroups 4 as 3 and 10/10, and subtracts. His neighbor, however, confidently changes the denominators "just to be safe," a "buggy" habit that too often yields the wrong answer. Same task, but two different windows into a student's thinking. The second student's error wasn't random; it was evidence of a mislearned rule, a scar from a past lesson. Right-or-wrong is not a fixed trait but an inference about an approach to navigating misconceptions when solving a specific problem.

He saw the same power in physics. On the Force Concept Inventory, Bob would point to a ball-toss item where the tempting wrong answer implies a lingering "push" is needed to keep the ball in motion. Each answer, he'd explain, was a window into a learner's mind, not just a tally mark. The wrong answers weren't random—they were principled, revealing the durable, intuitive beliefs that shape how a novice sees the world. In his telling, the test wasn't just a test; it was a probe designed to turn misconceptions into data.

Assessment Inside Authentic Activity

Bob believed that skills live inside their work. A doctor's English, he'd argue, is not a lawyer's English, so any assessment must live inside the world where that language does its job. He championed tests like the Occupational English Test (OET), which wasn't a grammar quiz but a measure of English braided with the messy, high-stakes genres of clinical practice—reading a patient chart, following a clinical dialogue, parsing a prescription note. He called its design a humane argument because it was built from the authentic demands of the work itself. He saw the same proof that capability lives in context by watching international graduate students teach beginners. You learn if someone can teach by watching them teach, especially when faculty from that same setting—who know what good looks like—are the raters. His conviction was clear: to know whether someone can navigate a system, you have to let them navigate it. This, in his eyes, was a true, humane story about learning in action.

Process Data and Evidence-Centered Design

One of Bob's favorite stories was about Hydrive, an intelligent tutor that trained F-15 hydraulic mechanics in weeks instead of months. The simulation created a dynamic portrait of skills by logging every action, allowing one to infer a trainee's troubleshooting strategy from the sequence of their choices. He prized how the model updated its inferences by accounting for a trainee's instructional history.

He also described the Dental Interactive Simulations Corporation (DISC), where a student's trail of judgment in a virtual clinic provided the evidence. Every decision and course-correction, especially when new information forced a revision, formed the evidentiary claim. Because the tasks mirrored actual professional decision points, the inferences were tethered to how work is really done.

Similarly, Bob celebrated the Cisco Networking Academy lab, where students use a command line in the Packet Tracer simulator to find a bug planted by an instructor. The simulator logs every step, making the pathway the inference. The evidence was the authentic work of repairing the fault, not responses marked on a bubble sheet. The design revealed *how* students found the problem, not just *if* they found it, making

the tool both a learning experience and a certification of competence. Bob always credited the village that built it, including John Behrens at Cisco and colleagues Russell Almond and Malcolm Bauer.

With SimCityEDU: Pollution Challenge!, Bob was excited not by the game's polish but by what he called the "live argument you could have with a city." A student could try a policy and watch the system—the economy, air quality—push back in real time. In these decision traces, he could read an engineering mind at work, grasping feedback loops and unintended consequences. He called it a "system-of-parts test," whose telemetry revealed a student's cause-and-effect models. He would roll the credits for the teams at GlassLab, EA, and ETS before landing his lifelong lesson: context is the construct, so assessment belongs inside authentic activity.

Any account of Bob's stories is incomplete without mentioning his most important stories: his loving marriage and lifelong partnership with Robbie, his beloved wife of 50 years. Bob shared about their love for national parks and exploring the world to photograph wildlife together. He was immensely proud to hood his daughters (Jessica and Meredith) when they earned their degrees from his department at Maryland, finding joy in how they followed their own intellectual paths. He loved being Grandpa Bob, and especially loved his grandsons Luke, Alex, and Elliot. He was a gentle soul whose kind voice and endless patience made him a blessing to anyone who met him.

Conclusion

At the end of an incandescently brilliant presentation, after laying out a complex model or a groundbreaking analysis, Bob would do something unusual. He would meticulously thank collaborators, from the software engineers to the high school teacher who wrote curriculum activities, ensuring the work was seen not as his own but as the product of a community. Science was a team sport. He made it clear how indebted he was to Air Force mechanics, dental hygienists, and video game designers—for their willingness to reason about evidence together. For Bob, this wasn't just a courtesy.

In a career dedicated to crafting *defensible stories about learning*, the most defensible story is the one Robert Mislevy told through his actions. While he gifted our field intricate statistical models, he built us up through simple parables, demonstrating that the stories we tell about each other—the generous crediting of collaborators, the patient framing of a problem—are among the most powerful models we have. His career was a quiet refutation of the myth that brilliance must have sharp edges, proving that the most generative intellectual force is the patient cultivation of collective intelligence.

To separate Bob's kindness from his brilliance is to miss the point. Generosity was the operating system for his scholarly and design work, the method by which he made the complex clear and the contentious collaborative. He built statistical bridges connecting evidence to claims, but his more enduring structures were human ones. Bob Mislevy's legacy is not a monument to be admired but a practice to be inhabited: a challenge to ask the clarifying question, share the credit, and solve the next tough problem together, with a rigorously applied kindness.

John T. Behrens

Director, Office of Digital Strategy, College of Arts & Letters
Director & Prof. of the Practice, Technology & Digital Studies Program
Concurrent Prof. of the Practice, Computer Science & Engineering
University of Notre Dame

Cognitive Re-calibration

Robert J. (Bob) Mislevy was wise, gracious, generous, funny, and—especially important to me—interested in technology. Bob changed how I thought about myself both professionally and personally, and changed the trajectory of my career. He was a blessing to anyone who met him.

I first intersected with Bob when I read his 1994 tour de force, "Evidence and inference in educational assessment" (Mislevy, 1994). Having long been dismissive of psychometrics as overly focused on large-scale testing and ignorant of the wider world of assessment concerns (e.g., classroom assessment), I was bowled over by the piercing thoughtfulness of the piece. I came away thinking, "Well, if this is psychometrics, perhaps I should give it a chance."

That chance came just a few years later when I was working in industry and was tasked with starting "the best assessment program in the technology sector" in the service of the emergent Cisco Networking Academy that later grew to serve 10,000 schools in 160 countries. I knew that while I was not up for the job, we might have a chance if we could enlist Bob Mislevy. I was happy to hear that Bob had a new framework called evidence-centered design (ECD), and we both agreed we would use Cisco's rapidly growing global learning network as a launching pad for ECD's ideas. After a few years, the initial work was published in a special issue of the *International Journal of Testing* under the editorship of Bruno Zumbo. I was extremely lucky to share ongoing collaborations related to ECD and many other ideas for more than a decade after that.

Of course, Bob was brilliant technically, but, more essentially, he was wise. He opened doors to getting things done by solving very difficult problems, mathematically or conceptually or operationally. Yet, not only did he open doors to serve children and adults around the world through assessment experiences, he plumbed the depth of fundamental questions about what it means to think, to reason, to quantify belief, to move from observation to a socially negotiated claim. His thinking went far beyond how to solve a single problem as he addressed deeper underlying questions: What are we doing here? How do we think about this or anything? How do we update our beliefs?

To me, Bob was in the business of revelation, a word that comes from the Latin "to lift the veil or unveil." Bob's work lifted the veils around statistical reason, system automation, the social underpinnings of our work, and the social consequences. Bob was doing spade work at the foundations of psychometrics, education, psychology, and related fields.

Bob was amazingly gracious. By all rights, his stature would have allowed him to be a jerk or at least a curmudgeon. He was not. I never heard him say anything negative about anyone, ever. In fact, he was so nice that we needed some linguistic calibration when my team and I at Cisco first started working with him in 2000. After a few months, we had to call an ad hoc meeting to talk about the "Bob translation problem." The fix was straightforward. When Bob says, "That's one way to think about it. Another way to think about it would be to x, y, and z," we are going to translate that as follows: "Dude, do not do what you were thinking about, just do what Bob suggests!" After that, everything was fine.

We needed another cognitive recalibration related to Bob the following year at the American Educational Research Association/National Council on Measurement in Education meeting. There, some of my students and I mentioned working with the person we knew as "our friend Bob" only to find people nearly falling out of their seats. We knew he was prominent, but we did not know he was famous enough that the psychometric community already considered him a cultural treasure, like a world heritage site. That year, 20 years before he retired, he received a lifetime achievement award.

Bob was gracious to everyone. Decades later, he would ask after the people we had worked with: the high school teacher who wrote the activities, the engineer who conducted the training, the software developer who wrote the code. He always had time for the simplest question and made every questioner an important contributor to the larger process of unveiling essential realities.

It is sometimes hard to remember the way the world worked when I first met Bob. A world of paper where information networks were only recently deployed. Where paper was the primary mode of assessment and fixed response was the primary format. Bob helped my teams and many other teams break from these constraints by providing other computational and conceptual languages, allowing us to move from ideas of questions and answers on a test to task and performance features in rich classroom activities, complex simulations, and games. Bob's insights set the groundwork for so much of what happens in large-scale online mobile environments. He opened vistas for both technology-enhanced assessment as well as assessment-enhanced technology (DiCerbo & Behrens, 2012). We see these trends continued in the work of Dr. Sarah DeMark, a leader at Western Governors University, who was part of the original Cisco team, as well as that of Dr. Kristen DiCerbo, chief learning officer at Khan Academy, who worked closely with Bob in the later Cisco days on simulation and games. I see goals and approaches influenced by Bob replicated throughout many organizations in the education and human performance industries. Just this summer, a paper was published about using ECD to design benchmark systems for LLM validation. In my generative AI class, I teach ECD as the framework to not only assess human behavior but also assess system behavior.

These are just some examples of how this kind and generous man has impacted the world in only one of his many areas of impact. Bob was a great gift to all who knew him and to many millions of learners whose lives have been improved by the more appropriate knowledge he taught us to accumulate and critically examine. I have no doubt that his impact will be greater in the years ahead as new challenges catch up with the fundamental revelations he provided us.

References

DiCerbo, K., & Behrens, J. (2012, April 15). *From technology enhanced assessment to assessment enhanced technology* [Conference presentation]. Annual Meeting of the National Council on Measurement in Education, Vancouver, Canada.

Mislevy, R. J. (1994). Evidence and inference in educational assessment. *Psychometrika*, *59*(4), 439–483. https://doi.org/10.1007/BF02294388

Russell Almond

Associate Professor of Measurement and Statistics
Department of Educational Psychology and Learning Systems.
Florida State University

Sociocognitive Factors in Sociocultural Context

Robert (Bob) Joseph Mislevy was born June 28, 1950. He earned a Ph.D. in Research Methodology from the University of Chicago in 1981. He joined Educational Testing Service (ETS) in 1984. In 2001, he joined the faculty of the University of Maryland in the Department of Measurement, Statistics and Evaluation and returned to ETS in 2011 as the Frederic M. Lord Chair in Measurement and Statistics. He died on May 22, 2025, at age 74. He was predeceased by his wife Roberta (Robbie), and survived by two daughters and three grandchildren.

A complete account of Bob's work on assessment would take more space than is available in this brief tribute. It falls into roughly three categories: work applying ideas from Bayesian statistics to psychometrics, particularly his work with Darell Bock on Bilog and his suggestion to use plausible values in the National Assessment of Educational Progress; work on evidence-centered assessment design (ECD; Almond et al., 2015; Mislevy, 1994; Mislevy et al., 2003); and work on sociocognitive assessment (Mislevy, 2018; Oliveri et al., 2021). Many of his papers are available as research reports from the National Center for Research on Evaluation, Standards, and Students Testing (https://cresst.org/?post_type=publication&s=Mislevy), where Bob was a long-time fellow. As Bob was a thoughtful and careful writer, these works are worth revisiting; in repeated readings, the reader often sees new details. I assign many of these to students in my assessment design course.

Bob's impact went far beyond his skills as a philosopher and statistician. It came from his willingness to listen carefully and actively to people with different points of view. When Bob didn't understand something, his instinct was to ask questions, and to try and see the point of view of the other person. He was always respectful of his conversation partners, actively listening and never wanting to imply that his viewpoint was more important than theirs. This deep respect for others was central to Bob's achievement in both understanding language and sociocultural assessment.

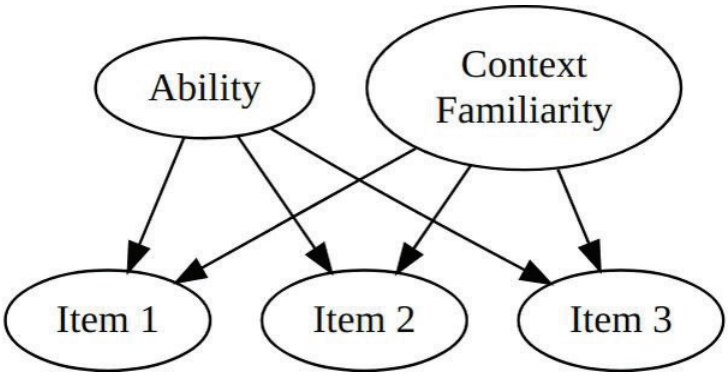
I met Bob Mislevy when I came to ETS in 1995. At the same time that Bob, Linda Steinberg, and I were having discussions that would lead to ECD, the Test of English as a Foreign Language (TOEFL) was going through a redesign effort. Bob and I sat in the back of the room, listening, and brought what we heard back to our discussions with Linda. One thing we observed was the subject matter experts (linguists and language experts) and the senior test developers were often talking past each other. The language experts were describing the knowledge, skills, and abilities (proficiencies) important in English communication while the test developers were describing tasks found to be useful. Our conclusion was that talking about evidence could bridge the gap: the language experts could talk about what kinds of tasks provide evidence for the posited proficiencies, and the test developers could talk about which of the posited proficiencies the tasks provide evidence for.

These discussions sparked Bob's interest in what the language experts were calling *communicative competence*. Bob did a deep dive into this subject, starting a series of discussions with Lyle Bachman and others about a modern view of language. Looking back on my notes from those early days, I see that *sociolinguistic* was one of the proficiencies we had identified in a toy language model we built. While for me this was merely a label given to a variable, Bob was struggling to understand it at a deeper level.

Mislevy (2003) lays out four philosophical bases that can be used for assessments: (1) trait scores—the assessment captures the correlation among item responses; (2) behaviorist—the assessment

captures the tendency/ability of the subject to engage in certain behaviors; (3) cognitive processing—the assessment identifies which knowledge, skills, and abilities the subject uses to process the items; and (4) sociocultural—the assessment measures the degree to which the subject's production follows appropriate social and cultural expectations. I remember discussing with Bob a meeting he attended on sociocultural assessment. He expressed some frustration that many participants were stuck in their respective philosophical framework—cognitive processing or sociocultural—and talking past each other. Bob, coming from the cognitive processing world, was trying hard to understand the sociocultural perspective; fortunately, Jim Gee was willing to engage with him from the other side. Their conversations helped Bob integrate the sociocultural perspective with the others, which, in turn, led to his 2018 book on the subject, *Sociocognitive Foundations of Educational Measurement*.

One task type which we discussed during the development of ECD was the reading passage testlet: a short passage followed by several items. The problem with this task type is that the content of the passage could break the local independence assumption of the items. The example Bob used was a reading passage about dinosaurs. A student who was a dinosaur enthusiast would have no di iculty with words like pterodactyl and paleolithic while a student new to dinosaurs might struggle with the unfamiliar words and topic. The model we built for this had two latent variables: one (part of the global proficiency model) for reading proficiency and one (local to the task) to capture familiarity with the context. Figure 1 shows this model. Note that because context familiarity provides an alternative explanation for performance ability, the amount of information about ability from this testlet is less than what could be obtained from three independent items. It is, however, a more realistic task.



In one of my last conversations with Bob, I brought this model up again. The way I had been thinking about sociocognitive factors is that the task would have a sociocultural context related to its expectation, and the subject would have a number of sociocultural contexts with which they were familiar. So the match between the contexts of the task and the subject would affect performance on the assessment. For Bob, his understanding was far more nuanced. When assessing communication, the context of the task includes both the sociocultural context of the writer/speaker and the sociocultural context of the audience. One of the marks of a skilled communicator is that they adapt their communication for the audience. This is a fundamental problem in communication and assessing communication, as each individual is unique—speaker and audience. This situation highlights the challenge of educators: to bring people from varied backgrounds to a place where they can share insights with each other. Bob's writings here are incredibly helpful: in particular, his book *Sociocognitive Foundations of Educational Measurement* (Mislevy, 2018) and his contributions to the 2021 special issue of this journal (e.g., Oliveri et al., 2021). As is often the case, we will find new insights with repeated readings.

We are all saddened that we will no longer have the chance to enjoy conversations with Bob filled with his deep insights and gentle humor; at the same time, he has left us with a wealth of ideas to read about, think about, and build upon.

References

Almond, R. G., Mislevy, R. J., Steinberg, L. S., Williamson, D. M., & Yan, D. (2015). *Bayesian networks in educational assessment* (CSE Report No. 837). Educational Testing Service.

Mislevy, R. J. (1994). Evidence and inference in educational assessment. *Psychometrika*, *59*, 439–483. https://doi.org/10.1007/BF02294388

Mislevy, R. J. (2003). Substance and structure in assessment arguments. *Law, Probability, and Risk*, *2*(4), 237–258. https://doi.org/10.1093/lpr/2.4.237

Mislevy, R. J. (2018). *Sociocognitive foundations of educational measurement*. Routledge.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessment. *Measurement: Interdisciplinary Research and Perspective*, *1*(1), 3–62.

Oliveri, M. E., Slomp, D., Rupp, A. A., & Mislevy, R. J. (2021). Principled development workplace English communication part 3: An integrated design and appraisal framework. *The Journal of Writing Analytics, 5*(1), 109–141. https://doi.org/10.37514/JWA-J.2021.5.1.04

Duanli Yan

Research Scientist
Measurement Incorporated

Brilliant Researcher, Generous Teacher, Role Model

Bob was the Fred Lord Chair at Educational Testing Service and, more importantly, a truly empathetic and supportive mentor who was deeply invested in the success of others. Despite his impressive achievements, he remained remarkably humble and approachable. He had a way of patiently guiding you toward finding answers for yourself, building your confidence step by step.

Bob hired me in 1989, right after I had completed my dual master's degrees in statistics and operations research. During the interview, I nervously admitted, "I don't know anything about education." He smiled and reassured me, "That's not a requirement." From that moment on, he took me under his wing and became a mentor who supported me for the next 35 years.

In 1990, Bob shared his insights on assessment frameworks and led us in developing domain analysis, domain modeling, and the architecture design of the four processes of assessment design and delivery. He also spearheaded the development of *StatShop*, a suite of statistical tools for assessment analysis and scoring. Under his guidance, I learned evidence-centered design and Bayesian networks—tools for modeling complex relationships among variables while incorporating domain knowledge, building interpretable models, handling incomplete data, and performing probabilistic reasoning for tasks such as diagnosis, prediction, and decision-making under uncertainty. Together with colleagues, we published influential works including "Bayes nets in educational assessment: Where the numbers come from" (Mislevy et al., 1999), "Models for conditional probability tables in educational assessment" (Almond et al., 2001), "Design and analysis in a cognitive assessment" (Yan et al., 2003), and several others that have significantly impacted in the field of educational assessment.

As early as 1990, Bob also envisioned AI applications in education. He involved me in his Bayesian modeling for the Hydrive project, an intelligent tutoring system designed to train F-14 fighter jet technicians. The system provided personalized instruction and feedback that emulated a human tutor, offering scalable one-on-one learning experiences and tracking student progress. Remarkably, it reduced training time from three months to just three weeks. Our work was later documented in *Design Recommendations for Intelligent Tutoring Systems* (Mislevy & Yan, 2017), and we coauthored *Bayesian Networks in Educational Assessment* (Almond et al., 2015).

Bob was not only a brilliant researcher but also a generous teacher and role model. When I was deciding where to pursue my Ph.D., he wrote me a strong recommendation letter, invited me to tour his campus, and personally introduced me to his department chair at the University of Maryland. He also created the National Council on Measurement in Education tutorial on Bayesian networks in educational assessment, which we have presented annually since 2002. Inspired by him, I now serve as an adjunct professor at Rutgers and Fordham, sharing my knowledge and experiences with my students and colleagues.

When my coedited volume *Computerized Multistage Testing: Theory and Applications* (Yan et al., 2014) was published, Bob wrote a thoughtful review in the *Journal of Educational and Behavioral Statistics* (Mislevy, 2015). His visionary work on AI also shaped *Automated Scoring of Complex Tasks in Computer-Based Testing* (Williamson et al., 2006), which inspired my later volume *Handbook of Automated Scoring: Theory into Practice* (Yan et al., 2020). In addition to contributing his commentary, *An Evidentiary-Reasoning Perspective on Automated Scoring* (Mislevy, 2020), we coauthored the chapter "Automated scoring in intelligent tutoring systems (Mislevy et al., 2020).

In 2016, Alina, Charlie, and I were especially grateful for Bob's recommendation of our team for the American Educational Research Association Division D Award for Significant Contribution to Educational Measurement and Research Methodology.

Looking back, I feel incredibly fortunate to have had Bob as my mentor for so many years. His guidance, encouragement, and vision have shaped my career and my life in ways I could never have imagined. I would not be where I am today without him, and I am eternally grateful for his mentorship and friendship.

References

Almond, R., Mislevy, R., Steinberg, L., Yan, D., & Williamson, D. (2015). *Bayesian networks in educational assessment*. Springer-Verlag. DOI: 10.1007/978-0-387-98138-3

Almond, R. G., DiBello, L., Jenkins, F., Mislevy, R. J., Senturk, D., Steinberg, L. S., & Yan, D. (2001). Models for conditional probability tables in educational assessment. In T. Jaakkola & T. Richardson (Eds.), *Artificial intelligence and statistics 2001* (pp. 137–143). Morgan Kaufmann.

Mislevy, R. J. (2015). A review of computerized multistage testing. *Journal of Educational and Behavioral Statistics, 40*(4), 425–431.

Mislevy, R. J. (2020). An evidentiary-reasoning perspective on automated scoring: Commentary on part I. In K. E. DiCerbo & J. T. Behrens (Eds.), *Handbook of automated scoring: Theory into practice*. Routledge.

Mislevy, R. J., & Yan, D. (2017). Evidence-centered assessment design and probability-based inference to support the generalized intelligent framework for tutoring. In R. Sottilare, A. Graesser, X. Hu, & G. Goodwin (Eds.), *Design recommendations for intelligent tutoring systems* (Vol. 5, pp. 101–124). US Army Research Laboratory. https://apps.dtic.mil/sti/pdfs/AD1160340.pdf

Mislevy, R., Almond, R., Yan, D, & Steinberg, L. (1999). Bayes nets in educational assessment: Where the numbers come from. In K. B. Laskey & H. Prade (Eds.), *Uncertainty in artificial intelligence '99* (pp. 437–446). Morgan-Kaufmann.

Mislevy, R. J., Yan, D., Gobert, J., & Sao Pedro, M. (2020). Automated scoring in intelligent tutoring systems. In K. E. DiCerbo & J. T. Behrens (Eds.), *Handbook of automated scoring: Theory into practice*. Routledge.

Williamson, D. M., Mislevy, R. J., & Bejar, I. (2006). *Automated scoring of complex tasks in computer-based testing*. Routledge.

Yan, D., Almond, R. G., & Mislevy, R. J. (2003). *Empirical comparisons of cognitive skills diagnosis models* [Conference session]. National Council on Measurement in Education 2003 Annual Meeting, Chicago, IL.

Yan, D., Mislevy, R. J., & Almond, R. G. (2003). Design and analysis in a cognitive assessment. *ETS Research Report Series, 2003*(2). https://doi.org/10.1002/j.2333-8504.2003.tb01924.x

Jiangang Hao

Research Director
Educational Testing Service

Mentor, Collaborator, and Role Model

I first met Bob Mislevy in 2013, at a turning point in my career. At the time, I was conducting astrophysics research at the Fermi National Accelerator Laboratory but was beginning to seek a new professional direction. Around the same period, Alina von Davier launched the Center for Advanced Psychometrics at Educational Testing Service (ETS), with a bold vision of bringing together scholars from diverse disciplines to explore how data from interactive digital assessments could be harnessed to measure complex skills. Bob, who had just returned to ETS after a decade at the University of Maryland, was serving as an advisor to the new center. He played a pivotal role in shaping its mission, recruiting talent, and mentoring researchers. My first encounter with him came during my interview for the research scientist position at the center. Shortly thereafter, I joined ETS, and Bob became my mentor, guiding me into the world of educational measurement.

When I joined ETS, I brought advanced training in physics and statistics, years of research in astrophysics with extensive publications, and the lived experience of taking countless tests along the way. With this background, I assumed I understood testing well, an illusion not uncommon among scientists outside the field of educational measurement. But this illusion quickly dissolved. Through Bob's Socratic style of mentoring, I came to realize that while the mathematics of assessment may appear less complex than physics, the ideas and principles behind what to measure, why to measure it, and how to measure are both profound and exquisite. It took me years of active learning from Bob and other brilliant colleagues to truly understand, appreciate, and internalize these principles.

Bob lived in Maryland but visited ETS for several days each month. Each visit gave us opportunities for long, wide-ranging conversations, which became crucial in shaping not only my understanding of assessment but also my approach to scientific thinking. As a well-trained scientist who naïvely thought of assessment as merely a collection of questions, I carried with me some long-standing puzzles I had wrestled with but never resolved. In our very first meeting, I started with one that had long troubled me: If Isaac Newton were given a modern graduate physics exam, he would probably fail, but did that mean Newton's understanding of physics was bad?

Bob smiled, recognizing it as the perfect opportunity to clarify my misconceptions. A test, he explained, is not merely a collection of questions for anyone to answer. It is a measurement, defined by what you want to measure, how you measure it, and how you interpret the results. In articulating these components, he emphasized the core values of measurement must be respected: validity (does the assessment measure what it is intended to measure?), reliability (does it do so with acceptable error?), comparability (are scores consistent across similar assessments?), generalizability (can results extend beyond the specific conditions of the test?), and fairness (do scores avoid advantaging or disadvantaging individuals based on irrelevant demographic characteristics?). Only after these issues were addressed, he explained, can an assessment be considered sound and trustworthy. So, my question about testing Newton, when examined through these dimensions, failed at the very first step: defining the construct and using the right instrument. This was the most charming part of talking with Bob: no matter how complex the question you raised, he could parse it with great skill and lead you toward a clear answer.

On a different occasion, I raised another question that had perplexed me for a long time: Every test has some degree of measurement error. This means that small differences in scores, a few points here or there, may not reflect true differences in ability. Yet in practice, we often make consequential decisions,

such as admitting one student over another, based on exactly those small differences. How can this be fair? In response, Bob introduced me to Paul Holland's concept of the dual interpretation of testing: A test can be seen both as a measurement and as a contest. From the measurement perspective, small score differences may lack statistical significance. But from the contest perspective, as long as the rules are clearly defined and applied equally, the outcome can still be considered fair. This framing struck me as a powerful way to reconcile the tension between measurement errors and real-world decision making. Once again, Bob had taken what seemed to be a very challenging problem and clarified it with remarkable simplicity, providing not only an answer but also a new lens through which to view assessment.

These conversations, and many others over the years, more than answered my technical questions; they reshaped how I thought about assessment. I came to see the field not as a problem that can be solved with mathematics and statistics, but as one requiring both technical rigor and deep reflection on human values. What made Bob such an extraordinary mentor was not only his knowledge but also his way of explaining. He did not lecture or impose. Instead, he listened carefully, took my questions seriously, and guided me toward insights that felt like my own discoveries. In doing so, he modeled the very essence of what it means to be an educator.

Looking back now, I realize how pivotal those exchanges were in my intellectual journey. They helped me transition from physics and statistics into educational measurement with a deeper appreciation for the field's complexity and significance.

Over the past decade, I also had the privilege of collaborating with Bob on several technical projects. We collaborated on psychometric considerations for game-based assessment (Mislevy et al., 2014), and we later introduced the evidence trace file (Hao & Mislevy, 2018; Hao et al., 2016) and the evidence identification–centered data design (Hao & Mislevy, 2022) as extensions of Bob's evidence-centered design framework, which provided for the systematic extraction of evidence from complex real-world data. We also developed methods to characterize communication patterns in collaborative tasks, advancing the assessment of collaborative problem-solving (Hao & Mislevy, 2019). Later, with Alina von Davier, we co-edited *Computational Psychometrics*, a volume that helped define a new discipline to bridge psychometric principles with methods from data science and machine learning/AI (von Davier et al., 2022a, 2022b). This work was later honored with the National Council on Measurement in Education Annual Award for Exceptional Achievement in Educational Measurement in 2024.

For me, Bob was a mentor, collaborator, and role model. His legacy endures not only in the field of educational measurement but also in the countless individuals, like me, who were fortunate enough to learn from him. I feel deeply grateful to have known him, and I know that his influence will continue to shape our field and our lives for many years to come.

References

Hao, J., Smith, L., Mislevy, R., von Davier, A. A., & Bauer, M. (2016). Taming log files from game/simulation-based assessments: Data models and data analysis tools. *ETS Research Report Series, 2016*(1), 1–17. https://doi.org/10.1002/ets2.12096

Hao, J., & Mislevy, R. J. (2018). The evidence trace file: A data structure for virtual performance assessments informed by data analytics and evidence-centered design. *ETS Research Report Series, 2018*(1), 1–16. https://doi.org/10.1002/ets2.12201

Hao, J., & Mislevy, R. J. (2019). Characterizing interactive communications in computer-supported collaborative problem-solving tasks: A conditional transition profile approach. *Frontiers in Psychology, 10*, 1011. https://doi.org/10.3389/fpsyg.2019.01011

Hao, J., & Mislevy, R. J. (2022). A data science perspective on computational psychometrics. In A. A. von Davier, R. J. Mislevy, & J. Hao (Eds.), *Computational psychometrics: New methodologies for a new generation of digital learning and assessment: With examples in R and Python* (pp. 133–158). Springer. https://doi.org/10.1007/978-3-030-99139-2_7

Mislevy, R. J., Oranje, A., Bauer, M. I., von Davier, A. A., Hao, J., Corrigan, S., Hoffman, E., DiCerbo, K. E., John, M., & Behrens, J. T. (2014). *Psychometric considerations in game-based assessment*. GlassLab Games.

von Davier, A. A., Mislevy, R. J., & Hao, J. (2022a). Introduction to computational psychometrics: Towards a principled integration of data science and machine learning techniques into psychometrics. In A. A. von Davier, R. J. Mislevy, & J. Hao (Eds.), *Computational psychometrics: New methodologies for a new generation of digital learning and assessment: With examples in R and Python* (pp. 1–6). Springer.

von Davier, A. A., Mislevy, R. J., & Hao, J. (2022b). *Computational psychometrics: New methodologies for a new generation of digital learning and assessment: With examples in R and Python*. Springer.

Diego Zapata-Rivera

Distinguished Presidential Appointee
Educational Testing Service

Assessment Innovations

I remember many interesting conversations with Bob about issues involving educational assessment. These conversations varied from general issues about the future of educational assessment to discussions about modeling evidence sources and the use of Bayesian networks for student modeling.

Bob had a remarkable ability to inspire confidence and curiosity in those around him. He listened patiently and asked great questions. More than an exchange of ideas, talking with him felt like embarking on a shared journey toward deeper understanding. His questions often made me realize aspects of the work that needed further consideration. He offered excellent suggestions and advice. Such discussions frequently resulted in research projects that have influenced the trajectory of my professional career. Below are several instances illustrating Bob's impact across multiple areas of research.

Bob's influence on my work can be traced back to my time as a graduate student working on human interaction with Bayesian student models. I remember reading his papers on Bayesian networks and Intelligent tutoring systems (e.g., Mislevy & Gitomer, 1996). When I joined Educational Testing Service (ETS), I had the opportunity to work on projects involving the application of evidence-centered design (ECD; Mislevy et al., 2003) in assessment. During that time, I remember having conversations with Bob about our projects when he visited us. One of these projects focused on applying ECD principles to identify and integrate evidence of students' knowledge, skills, and other attributes (KSAs) in games. We had a conversation with Bob about the idea of identifying relevant evidence of students' KSAs from patterns of game actions. Bob was open to this idea and encouraged our exploration (Shute et al., 2007). This work led to research exploring the role of games in educational assessment (e.g., Mislevy et al., 2016; Shute & Ventura, 2015; Zapata-Rivera & Bauer, 2012).

Other work that Bob greatly influenced involved the use of ECD principles to analyze accessibility features, provide support for English language learners, interact with open learner models, and characterize learning and assessment in digital environments (Arieli-Attali et al., 2019; Feng et al., 2009; Hansen et al., 2005; Hansen et al., 2009; Hansen et al., 2018; Zapata-Rivera et al., 2007a; 2007b; 2013). We talked about the design of assessment situations that provided students with the opportunity to demonstrate what they know or can do, situations that may have required support for non-focal skills.

More recently, we continued exchanging messages regarding our work on (a) conversation-based assessment (CBA), where short, written, or spoken conversations with virtual characters are designed to elicit evidence of target constructs (Forsyth et al., 2024; Zapata-Rivera et al., 2023, 2024); (b) caring assessment, in which several aspects of the learner are modeled including cognitive, noncognitive, cultural, and contextual aspects to create appropriate and engaging assessment situations (Sparks et al., 2024; Zapata-Rivera, 2017); and (c) personalized assessment for multilingual students, to which Bob contributed as a consultant in 2024. Key topics of discussion included the notion of conditional fairness, the challenges of gathering the necessary information to implement adaptive and caring assessments, tradeoffs between local usefulness and the broader comparability of assessments, and approaches for documenting use cases for personalized assessments for multilingual learners. As usual, his remarks offered valuable perspectives, prompting thoughtful discussion and shedding light on complex topics, thereby contributing to advancing these areas of research.

I am thankful for having the opportunity to work and learn from Bob. Bob will remain a guiding light and source of inspiration for many of us.

References

Arieli-Attali, M., Ward, S., Thomas, J., Deonovic, B., & von Davier, A. A. (2019). The expanded evidence-centered design (e-ECD) for learning and assessment systems: A framework for incorporating learning goals and processes within assessment design. *Frontiers in Psychology, 10*, 853.

Feng, M., Hansen, E. G., & Zapata-Rivera, D. (2009, April). *Using evidence centered design for learning (ECDL) to examine the ASSISTments system* [Conference presentation]. Annual Meeting of the American Educational Research Association (AERA), San Diego, CA.

Forsyth, C. M., Zapata-Rivera, D., Graf, E. A., & Jiang. Y. (2024). Complex conversations: LLM vs. knowledge engineering conversation-based assessment. *Proceedings of the 17th International Conference on Educational Data Mining* (pp. 868–871). https://doi.org/10.5281/zenodo.12729976

Hansen, E. G., Mislevy, R. J., Steinberg, L. S., Lee, M. J., & Forer, D. C. (2005). Accessibility of tests for individuals with disabilities within a validity framework. *System, 33*(1), 107–133.

Hansen, E. G., Zapata-Rivera, D., & Feng, M. (2009, April). *Beyond accessibility: Evidence centered design for improving the efficiency of learning-centered assessments* [Conference presentation]. Annual Meeting of the National Council on Measurement in Education, San Diego, CA.

Hansen, E. G., Zapata-Rivera, D., & White, J. (2018). Framework for the design of accessible intelligent tutoring systems. In S. D. Craig (Ed.), *Tutoring and intelligent tutoring systems* (pp. 69–101). Nova Science Publishers.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives, 1*(1), 3–62.

Mislevy, R. J., Corrigan, S., Oranje, A., DiCerbo, K., Bauer, M. I., von Davier, A. A., & John, M. (2016). Psychometrics and game-based assessment. In A. A. Rupp & J. P. Leighton (Eds.), *Technology and testing: Improving educational and psychological measurement* (pp. 23–48). Guilford Press.

Mislevy, R. J., & Gitomer, D. H. (1996). The role of probability-based inference in an intelligent tutoring system. *User-Modeling and User-Adapted Interaction, 5*, 253–282.

Shute, V. J., Ventura, M., Bauer, M., & Zapata-Rivera, D. (2009). Melding the power of serious games and embedded assessment to monitor and foster learning: Flow and grow. In U. Ritterfeld, M. J. Cody, & P. Vorderer (Eds.), *Serious games: Mechanisms and effects* (pp. 295–321). Routledge.

Shute, V. J., & Ventura, M. (2015). Stealth assessment. *The SAGE encyclopedia of educational technology* (pp. 675–676). Sage.

Sparks, J. R., Lehman, B., & Zapata-Rivera, D. (2024). Caring assessments: Challenges and opportunities. *Frontiers in Education, 9*, 1216481.

Zapata-Rivera, D. (2017). Toward caring assessment systems. *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization (UMAP '17)* (pp. 97–100). ACM. https://doi.org/10.1145/3099023.3099106

Zapata-Rivera, D., & Bauer, M. (2012). Exploring the role of games in educational assessment. In M. C. Mayrath, J. Clarke-Midura, D. H. Robinson, & G. Schraw (Eds.), *Technology-based assessments for twenty-first-century skills: Theoretical and practical implications from modern research* (pp. 147–169). Information Age Publishing.

Zapata-Rivera, D., Hansen, E., Shute, V. J., Underwood, J. S., & Bauer, M. (2007a). Evidence-based approach to interacting with open student models. *International Journal of Artificial Intelligence in Education*, *17*(3), 273–303.

Zapata-Rivera, D., Liu, L., Katz, I. R., & Vezzu, M. (2013). Exploring the use of game elements in the development of innovative assessment tasks for science. *Cognitive Technology, 18*(1), 43–50.

Zapata-Rivera, D., Forsyth, C. M., Graf, A., & Jiang, Y. (2024). Designing and evaluating evidence-centered design based conversations for assessment with LLMs. *Joint Proceedings of the Human-Centric eXplainable AI in Education and the Leveraging Large Language Models for Next Generation Educational Technologies Workshops (HEXED-L3MNGET 2024) co-located with 17th International Conference on Educational Data Mining (EDM 2024)*. https://ceur-ws.org/Vol-3840/L3MNGET24_paper3.pdf

Zapata-Rivera, D., Sparks, J. R., Forsyth, C. M., & Lehman, B. (2023). Conversation-based assessment: Current findings and future work. In R. J. Tierney, F. Rizvi, & K. Ercikan (Eds.), *International encyclopedia of education* (4th ed, pp. 504–518). Elsevier.

Zapata-Rivera, D., Vanwinkle, W., Shute, V., Underwood, J. S., & Bauer, M. (2007b). English ABLE. *Frontiers in Artificial Intelligence and Applications*, *158*, 323.

Ida Lawrence

Senior Vice President, Research & Development
Educational Testing Service (Retired)

Delivered at Robert J. Mislevy Memorial

Educational Testing Service, Princeton, NJ

September 25, 2025

Dinner and the Lord Chair

Good afternoon to colleagues, family, and friends of Bob, and thanks for the opportunity to say a few words at the conclusion of this time together to remember Bob. In addition to numerous people who have already shared their thoughts about Bob's enormous contributions to Educational Testing Service (ETS), and to the field of educational measurement, I am pleased to add mine.

In my role leading R&D, I had the distinct privilege in 2011 of strategizing with several colleagues to lure Bob back to ETS for a second stint. John Mazzeo will remember a particular dinner that John and I had with Bob at a lovely and fancy restaurant in Princeton. Our purpose in having this dinner with Bob was to put the icing on the cake and convince him to vacate his tenured faculty position at the University of Maryland to accept the ETS Lord Chair, a position formerly held by Paul Holland. My recollection is that our success in getting Bob to agree on such a significant career change took considerable convincing (thank you, John, for helping me through that dinner). Ultimately, we were successful, and Bob told us he was pleased to accept the offer to re-join ETS. He mentioned that a particular goal in returning to ETS was to finish the book he had been working on, *Sociocognitive Foundations of Educational Measurement*, ultimately published in 2018.

John and I are pretty sure that what sealed the deal was the special hamburger Bob ordered for dinner. This particular burger contained a center of Foie Gras. The meal was so rich that Bob mentioned to me the next morning that he was still digesting it. Years later, when he gave me a copy of the sociocognitive book, he entered an inscription into my copy. This is what he wrote:

> *"To Ida, Thank you for your support all these years and your friendship.*
> *The hamburger is at last digested! Best, Bob" (3/26/18).*

Randy Elliot Bennett

Norman O. Frederiksen Chair in Assessment Innovation
Educational Testing Service (Retired)

Radically New Thinking

The first time I heard Bob Mislevy's name was a few months before I met him. It was toward the end of 1983, and I was taking a workshop at Educational Testing Service (ETS) on item response theory (IRT). I remember the room, P-016, in the old Program Building, later Anrig Hall, which is now gone.

The workshop organizer, Martha Stocking, came into the seminar room one morning in a particularly enthusiastic and excited mood.

She said, "We've just hired a rising star to work on our IRT research. His name is Bob Mislevy."

So even as a relatively young researcher, his reputation was already preceding him.

I did get to know and work with him on various projects over the ensuring 42 years, so I had many opportunities to watch him grow and to learn from him—which I gladly did. From his landmark contributions to NAEP, to the beginnings of evidence-centered design (ECD), to his applications of it in various performance assessment projects, to the *Sociocultural Foundations* book, he proved himself to be unique in his ability to take ideas from other fields and apply them in novel ways, bringing new—sometimes, radically new—thinking to educational measurement.

As outstanding and groundbreaking as his many accomplishments were, they were more than matched in my eyes by his humility, kindness, generosity, collegiality, and decency. He was a scholar and gentleman in every sense of those words.

I am going to miss him—greatly.

Saad Khan

(Formerly) Educational Testing Service

Passionate Technologist

Working with Bob at Educational Testing Service from 2014–2017 was a memorable and rewarding experience for me. He was one of the world's foremost experts in the field of psychometrics and a mentor who significantly enhanced our understanding of complex subjects. Bob was also a passionate technologist with a keen interest in machine learning and AI, recognizing their potential to transform education. He guided my exploration of AI and psychometrics, and together we designed innovative assessments for 21st-century communication skills, integrating rigorous psychometrics with AI possibilities. His sharp intellect, accessibility, and ability to foster curiosity were invaluable assets.

Beyond his professional expertise, Bob's jovial demeanor, warmth, and generosity were unforgettable. He readily shared knowledge, offered support, and celebrated the successes of others. This unique blend of intellect, generosity, and cheer made collaborating with him truly enriching and memorable, leaving a lasting legacy of impactful partnership and mentorship.

Eva L. Baker

Distinguished Research Professor
University of California, Los Angeles

Always Teaching

I've been lucky to be a long-time colleague and friend of Bob's. I am also a great admirer of him as a scholar, innovator, and person. Let me focus on when Bob came to Los Angeles to spend extended time as a visiting scholar at CRESST, an R&D center focused on assessment, learning, technology, and methodology headquartered then and now at UCLA. CRESST, funded in part by the federal government, was made up of partners from a small set of universities, research centers, and educational organizations, and Bob was our principal partner from Educational Testing Service. Happily, the agenda for CRESST was developed collaboratively, and Bob chose his own valuable path. Bob Linn, Joan Herman, and I were so fortunate to have him as a collaborator and adviser across the decades.

When Bob came to UCLA, he was engaged with other colleagues on evidence-centered design. At that time, design played a less important role in assessment than methodological approaches to reliability and validity. Its connection to learning—from the measurement side—was frequently missing. Rather than extolling that contribution, I'd rather focus on the role he played at CRESST. Whenever I walked by his office, which was very close to mine, Bob was in a deep discussion with a CRESST faculty, graduate student, or staff member. No matter the subject matter or question, Bob was always willing to meet at any time and individually with anyone, from the most senior professor to the newest student associate. I confess I was personally grateful for at least a few appointments with him. During these conversations, it became obvious that Bob was actually teaching, either about design or whatever topic we brought to him. When I asked how he was finding the staff, he profusely complimented them. He also noted that the topic of imputed values came up a lot.

From the staff, I learned that Bob was never in a hurry with them, and he was unfailingly clear and supportive in their small meetings. Bob became valued by our team as a person as much as an expert because of his generosity and kindness. Beyond our work together, I most admired him as a scholar who, with unusual range, moved fluidly from theory to framework design to applications and back again. He understood early on that his facility and expertise across types of programs, settings, different groups of respondents, and social contexts contributed enormously to the importance, quality, credibility, and utility of his work. All I can add is how grateful I am to have known him.

David Dorsey

Vice President
Human Resources Research Organization (HumRRO)

A Brilliant Soul

My experience working directly with Bob Mislevy was likely shorter in duration than many of those writing reminiscences for this volume. Yet the impact he had on me was both profound and lasting. My most recent opportunity to collaborate with him came as we served together on a National Academies committee studying the future of foreign language assessment for the US State Department's Foreign Service Institute. Simply being part of the committee was rewarding in its own right, but what I treasure most are the conversations I had with Bob during that period.

Those exchanges fundamentally reshaped how I thought about the future of validity arguments, particularly in light of the wave of AI-enabled assessments on the horizon. The clarity and originality with which Bob connected ideas were remarkable. He was not trying to be persuasive or forceful; instead, he would make a small observation, pose a thoughtful question, or draw a link between lines of reasoning. And in that understated way, he opened entire avenues of thought for me, avenues that later grew into a number of different presentations and publications. I often think he would not have realized the depth of his influence in those moments. For me, though, it was as if the clouds briefly parted on a mountain top, and wisdom came down in a way only Bob could impart.

What stands out equally, perhaps even more, was how he delivered his brilliance. Bob combined a razor-sharp intellect with a gentle and winsome presence. He was affable, approachable, and deeply willing to listen. You never felt as if you were being lectured or judged. Instead, you felt invited into a dialogue, one in which the quality of your own ideas rose simply by being in his company. That combination of intellectual rigor and genuine kindness is rare. It is easy to admire someone for their intellect; it is exceedingly less likely to be changed by the way their presence frames that intellect.

Looking back, I realize that the mentors and intellectual heroes who shaped me most deeply all shared this same quality. Their legacy was as much about who they were as people as it was about what they wrote or said. It was about a way of being, about showing that brilliance and humility, insight and kindness could inhabit the same person. In Bob, I saw a clear expression of that ideal.

For this reason, when I think of Bob Mislevy, I remember not only his scholarly contributions, which have unquestionably shaped the field of assessment and measurement, but also the way he carried himself: the patience, the openness, and the humility of a person who could have commanded attention yet chose instead to offer encouragement. That is a legacy as important as his intellectual one. And it is why I will always place him at the top of my list of kind and brilliant souls whose impact will echo for generations to come.

Sandip Sinharay

Educational Testing Service

A Pioneering Scholar and an Extraordinary Gentleman

I joined Educational Testing Service (ETS) in 2001. I was fortunate to be part of a division that housed several excellent psychometricians and statisticians. I soon learned that the most accessible of them was, other than my supervisor Paul Holland, the gentleman who was in the office right next to me—Robert Mislevy. When I look back on that time, I consider myself so lucky to have been assigned an office next to that of one of the greatest scholars of our field, even though Bob was employed at the University of Maryland at that time and used to come to ETS only once every couple weeks.

I was working on a project regarding the statistical and psychometric methods used in National Assessment of Educational Progress (NAEP). Those methods are quite complicated, to put it gently, and I often got confused about them. But then I found that Bob's research (documented in two pioneering papers in 1984 and 1985 and in other later papers) was the basis of much of those methods. I also heard stories from people about how, initially, the algorithm used to fit the item response theory (IRT) model in NAEP would not converge, and then Bob suggested the use of prior distributions to solve the problem. Someone even compared his use of prior distributions to stabilize an unstable algorithm to the consumption of Pepto-Bismol to cure an unstable stomach. Still, I did not even think about talking to Bob because of his stature in the field, but I soon learned from my colleagues that he was very helpful and would gladly talk to me about any psychometric or statistical topic. So I set up an appointment with Bob and had a question-and-answer session on NAEP methods. That was so helpful in understanding the intricacies of NAEP methods. I also noticed how passionate Bob was about the methods even though he was not working on them at that time.

After that initial meeting, I regularly had discussions with Bob on various topics. An interesting fact is that Bob had done pioneering research on pretty much whatever topic I worked on in my first few years at ETS—Bayesian methods, Bayesian networks, IRT model fit, expected response function, analysis of missing data, etc. I often discussed aspects of these topics with him when I had any confusion or questions, and he was extremely generous with his time, sharing deep insights. One of his Ph.D. students, Roy Levy, received the Gulliksen fellowship at ETS and worked with Bob and me. It was a great opportunity for me to work with two extremely smart people, one early on in his career and the other at the pinnacle. Bob's suggestions and advice were crucial in the successful completion of our research and corresponding publication.

Later, I kept in touch with Bob and had occasional discussions with him on topics like intuitive test theory, game-based assessment, sociocognitive foundations of educational measurement, culturally responsive assessment, etc. I was asked to write a review of his book, *Sociocognitive Foundations of Educational Measurement*, but I did not have the courage to do so until I had several meetings and discussions with Bob. I often showed what turned out to be clear misunderstandings of some topics in the book, but he was very patient with me and even explained the reasons behind them, saying he recognized the challenges I faced as a statistician and non-native English speaker in understanding the philosophical discussions in his book.

He was still very active and very willing to work with ETS researchers on various topics like large-scale assessment and personalized assessment when we received the extremely sad news of his passing. Bob's contributions will continue to be recognized and valued within the field. Equally enduring for those who knew him as a colleague is the example he set through his generosity and kindness.

Howard T. Everson

Visiting Research Scholar, Educational Psychology
Professor (Adjunct) MS Program in Data Visualization
Graduate Center, City University of New York

Delivered at Robert J. Mislevy Memorial

Educational Testing Service, Princeton, NJ

September 25, 2025

A Thirty-Year Conversation

Let me begin by thanking my friends and colleagues at Educational Testing Service (ETS) and the Mislevy family for inviting me to take part in this wonderful event, a celebration of the life and work of my longtime friend and colleague Robert J. (Bob) Mislevy.

I have known Bob—as both a colleague and a friend—for more than 30 years. And I want to tell you how coming to know him influenced my career in educational measurement.

In the Fall of 1991, after working for a few short years with the leadership of the City University of New York (CUNY) to design and implement a large, university-wide placement testing program in math and English language arts, I decided I needed to learn more about the technical aspects of standardized testing if I was going to make any progress on improving CUNY's testing program. My doctoral work in cognition and instruction hadn't provided me with the technical background and knowledge of psychometrics I needed to modernize the University's testing program.

With the support and encouragement of my colleague, the late Roger Millsap, I applied—and was accepted for a post-doctoral fellowship at ETS in the Psychometric Research Group. ETS's Samuel Messick and William Angoff were my mentors, two luminaries in the field of educational measurement. I was very fortunate.

Hoping to make the most of my post-doc opportunity, I asked Sam and Bill who at ETS they thought was doing the most advanced, cutting-edge work in psychometrics. Sharing sideways glances with one another, they both said, without hesitation, it was time I met Bob Mislevy. I had read a few of Bob's early papers, but I had not met him.

As luck would have it, Bill Angoff had arranged for me to use Fred Lord's office in Thurstone Hall during my post-doc at ETS. (Lord was on extended medical leave at the time and wasn't expected to return to ETS any time soon.) Fred Lord's office was a few doors down the hall from Bob's.

Back then, I was struggling to learn more about model-based measurement and item response theory (IRT) and wanted to explore how I could apply these concepts to improve the validity of CUNY's testing program. BILOG was the software tool of choice at the time. Who better to consult with on these topics than Bob? After all, Bob and his mentor, Daryl Bock, had written the software that allowed for broader use of IRT methods.

Back then, I needed and wanted to learn more about model-based approaches to educational measurement. Bob's intellectual interests in educational testing were expanding into areas that included human cognition and its role in educational measurement. (My doctoral work had been in the area of cognition, instruction, and assessment, and had been influenced by the work of Richard Snow at Stanford.) Bob had deep knowledge of model-based measurement approaches. We discovered we had lots to talk about.

During those early, formative years, I found Bob to be kind, patient, and very generous with his time. He introduced me to a broad range of cutting-edge ideas—including advances in psychometric estimation methods, Bayes nets, and his emerging thoughts on applying evidentiary reasoning to strengthen test validity.

Those conversations continued, in one form or another, for more than 30 years, sometimes at ETS; sometimes in New York City with colleagues from Columbia University, CUNY, and NYU; and other times with faculty and students at College Park, Maryland. Time and distance didn't matter—it was all about the ideas!

Indeed, as recently as this past spring, Bob had agreed to present his work on the sociocognitive foundations of educational assessment to a group of interested scholars who met under the Edmund W. Gordon Seminar Series on Assessment in the Service of Learning banner. I never missed an opportunity to hear Bob speak.

Bob's influence on the field of educational measurement has been, and continues to be, profound. When Bob passed away in May, he left behind a body of work that will be studied by scholars for years to come.

I will end my remarks with an Indian proverb—a beatitude of sorts:

"Blessed is he who plants trees under whose shade he will never sit."

Robert Mislevy left us all many trees to sit under. My students and I will take every opportunity to enjoy the shade.

Howard Wainer

Itinerant Philosopher & Statistician
Pennington, NJ

Bob Mislevy, Helicopters, and the Optimization of Complex Processes The Story

I first met Bob Mislevy in the late 1970s at a meeting at The University of Chicago that was organized by the eminent psychometrician Darrell Bock. I myself had come to Chicago almost a decade earlier as a young assistant professor explicitly to work with and learn from Bock. I was then leaving for Washington to work in the newly elected Carter administration. Bock had told me of a remarkable young student named Bob Mislevy, whom he met when Bob was attending a workshop on multivariate analysis that he was teaching. According to Bock, no other student in the class was close enough to Bob to finish second. In Bock's words, "He sparkled." Bock was delighted that he was able to persuade Bob to leave his job and enroll in a Ph.D. program at Chicago under Bock's mentorship.

During that fateful meeting, a number of topics were discussed, among them various approaches to optimizing complex processes. Singled out were the manufacturing of paper, the testing of abilities with multiple components, and the running of a school system. My memory of that meeting, though certainly dimmed by time, was that the young Bob Mislevy was not sparkling. Indeed, he seemed distracted and uninterested.1 It was clear to me that he was unlikely to have absorbed any of the deep truths that were exposed. I was wrong.

Fast forward more than a decade. The newly arrived graduate student Bob Mislevy had been transmogrified into the academic shooting star at Educational Testing Service (ETS) whose work on the National Assessment of Educational Progress (NAEP) sampling, evidence-centered design, and psychometrics in general validated Bock's prescient judgement.

In 1990, a blue-ribbon committee at ETS, called "TOEFL 2000," was formed to develop a new and improved version of the current Test of English as a Foreign Language (TOEFL) that would take into account many new developments in testing and language learning. It was to be a multi-year task, and it was granted a generous budget commensurate with its importance and its membership. Bob was on the committee in the formal role of senior psychometrician and the informal role of in-house sage. The first meeting of this committee was held at a large conference room in Conant Hall on the ETS campus. The committee was composed of more than 20 scholars, drawn from ETS staff and august institutions far and wide. They were all seated around an enormous table, and the meeting began. The chair of the committee opened the proceedings with an outline of the scope and goals of the committee and then passed the microphone clockwise around the table to solicit views from each member on how they thought the committee should proceed—at least initially.2

During that initial meeting, Bob whispered to Linda Steinberg, his long-term collaborator (and my long-term wife), that he was not optimistic about the committee's prospects and quipped that they'd be better off if they disbanded the committee and bought a helicopter with the unexpended funds. He reasoned that this way they would at least have something to show for the money. Over the ensuing weeks, this story leaked out and quickly became a legend at ETS. Indeed, thereafter any suggestion of convening a blue-ribbon panel was often met with the quip, "It would be better if they bought a helicopter."

1 I later learned that during that meeting Bob was engaged in passing a kidney stone, which fully explained his distraction.

2 All members were prepared for a long first day, for they all knew from experience that such meetings didn't end after everything was said, but only after everyone had said it.

 https://doi.org/10.37514/JWA-J.2025.8.1.06

The Back Story

The topic of why blue-ribbon panels do not work was one focus of that long-ago Chicago meeting. This was the discussion to which I believed Bob had paid no attention. Yet his helicopter suggestion showed that he not only heard the discussion but also internalized its underlying ideas. Some of those ideas are:

Such committees don't work because the task is too hard—it is purely hubris to think that we are somehow smarter or wiser than our forebears, who devised the current program/procedure that we are trying to improve upon. Our only advantage over our predecessors is that we may have additional evidence to lean on. If we don't gather and pay attention to such evidence, we are unlikely to make things better.

What works is to institute a culture of continuous experimentation in which we (a) choose dependent variables that characterize how well the system is working, (b) identify various independent variables that we believe affect the outcome, and (c) vary those independent variables one-at-a-time by a little and see if things improve. If there is improvement, change that variable a little bit at a time, in the same direction, and keep doing so until outcomes start to fall off. Then back up to the point of maximal performance, choose a second independent variable, and repeat: keep cycling through until the system is optimized. In this way, when the future arrives, the system of the future is there to greet it.

So, suppose we are trying to optimize an education program. We might choose NAEP scores as an outcome variable to be maximized, and among the independent variables, we might have class size. An experiment would divide classes randomly into treatment and control and then reduce class size by a little in an experimental section and see how NAEP scores are affected. If performance improves, try reducing class size a little more and see what happens. Continue doing this until further reductions are impossible/impractical. Do this at different grade levels and for different subjects. Or, alternatively, to save money, we might increase class sizes a little at a time and see if student performance changes. If not, we can achieve the same outcomes with fewer resources. Changes should be made gradually to avoid shocking the system so much that its function is seriously impaired.

If we adopt a culture of continuous small experimentation, the complex process we want to optimize is always moving toward that goal.

With this outlook, it becomes clearer that there is a role, albeit a more modest one, that a blue-ribbon committee can play. Such a panel can (i) help identify sensible dependent variables that characterize overall performance, (ii) suggest various independent variables that might affect the outcome and can be manipulated, and (iii) lend legitimacy and support to both the process and the outcomes that the process yields.

But, as Bob pointed out so parsimoniously 35 years ago, if the committee is being tasked with actually coming up with the optimal structure for a complex process, you are better off using the money to buy a helicopter. The system won't improve, but you will be able to ferry the committee members to the airport more quickly.

T.J. Elliott

Playwright & author
Chief Learning Officer
Educational Testing Service (Retired)

Claims Will Always Matter

Recently, my attention to a personal history of testing has focused on rebutting the prevalent notion among the punditocracy of the left (center, far, handed) that the problem with the corruption, unconstitutionality, and general havoc of the current federal government our electorate chose came about because a certain large segment of our citizenry was bedeviled and bewitched by misinformation. That supposition then begat a notion blaring like a dead smoke alarm battery from these cognoscenti: if only we could get *them* the correct information, everything would change. My take is that the kinds of choices involved in elections—attendance at rallies, campaign contributions, voting for someone or not casting any ballot—arise more out of a person's attitude than from a reasoning process. Want things to change? Figure out how to change people's attitudes, which is very different from any other kind of learning.

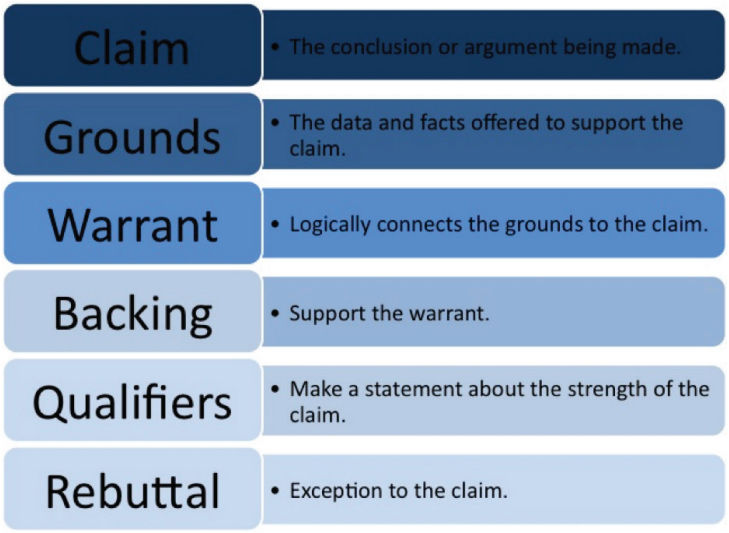
This stance does not preclude an appreciation for what reasoning can and does accomplish in our lives. An irony of our current sociopolitical circumstances is the absence of reasoning as the current administration seeks to dismantle both the process and effects of reasoning in our lives to achieve their objectives. A government report citing non-existent studies (McDuffie & Benadjaoud, 2025), a misleading tariff chart (Farley & Gore, 2025), and images of "genocide" in South Africa from the War in Congo (Chutel & Cvorak, 2025): you get the point. The veracity of a claim seems not to matter. Indeed, claims don't matter in this moment.

No. No. No. Claims do matter. And the person who generously, kindly, superbly taught me so much about claims sadly died on May 22, 2025: Bob Mislevy. We met during my time as Chief Learning Officer at Educational Testing Service (ETS). Bob returned there for a second stint in 2011, but his work was always central to the world of educational measurement in the 21st century. If anyone wanted to understand how assessments should go to be maximally helpful to learners and educators, they had to encounter and comprehend the key concept that Bob and other ETS colleagues, Russell Almond and Jan Lukas, devised called evidence-centered design (ECD), a way of thinking about testing founded on the ancient notion of *claims* (Mislevy et al., 2004).

If evidence sounds like an element more associated with criminal law or philosophy than with testing, then you're onto what Bob discerned. After all, evidence is what we need from the best designed tests, as well. And he used the best sources to create this new foundation for better assessments. John Henry Wigmore pioneered the analysis of legal evidence in trials to make claims of guilt or innocence. Evidence scholar David Schum studied the use of evidence in probabilistic reasoning. (Schum was particularly interesting since he used love letters as crucial evidence in a murder trial.) Stephen Toulmin developed a system of logic that could be used to evaluate a wide variety of arguments. The explanation Toulmin, a British philosopher, gives for a claim and what a claim needs to be able to pose an argument inspired Bob to take a similar approach to assessment. Toulmin wrote,

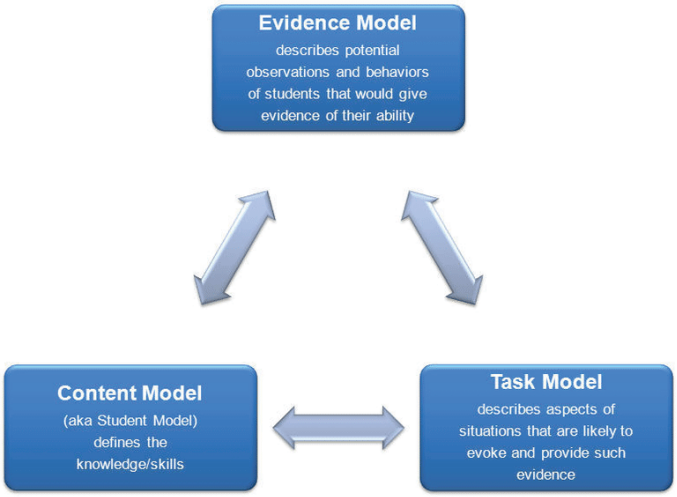
> *"A man who makes an assertion puts forward a claim—a claim on our attention and to our belief . . . just how seriously it will be taken depends, of course, on many circumstances. . . . whatever the nature of the particular assertion, in each case we can challenge the assertion, and demand to have our attention drawn to the grounds (backing, data, facts, evidence, considerations) on which the merits of the assertion are to depend. We can, that is, demand an argument" (p. 11).*

Toulmin's system of logic is illustrated below.



John Henry Wigmore looked at claims in a legal context. His 10-volume *Treatise on the Anglo-American System of Evidence in Trials at Common Law* (1904–05), usually called Wigmore on Evidence, is one of the world's most important books on law. If a court were to establish guilty or not guilty, liable or not liable, it was making a claim, and doing so meant gathering, examining, and judging evidence for one or the other of those positions. Evidentiary reasoning provides a coherent framework for making claims instead of making things up or just making a mess.

When it comes to the claims we make—or are made about us—in everyday life, evidence often fails to receive appropriate attention. We think much more about the score or outcome (getting into a particular school, snagging an A grade on an exam, being awarded a job) than we do about how it was derived. As Bob Mislevy noted in a 1994 paper discussing Wigmore at length, "There are multiple routes to an outcome, and observing the outcome alone does not indicate the route" (p. 486). In the most useful sorts of processes that end up producing claims, we should want evidence not only of what someone knows or can do but of how they are able to use that knowledge to accomplish the tasks that constitute the assessment. Below is an illustration of such evidence models.



Bob was able to extend and apply to assessment the concept of claims—their utility and value—from these disparate domains of logic, criminal courts, and probability. For example, Bob with Russell and Jan (and others later) started from the premise that the purpose of any test is to make a claim. As the trio wrote in the introductory paper on ECD, "What all educational assessments have in common is the desire to reason from particular things students say, do, or make, to inferences about what they know or can do more broadly" (Mislevy, 1994, p. 1). In other words, clarifying that an assessment was a claim was saying that we as a society needed to think differently about how we were testing kids of all ages. What were we claiming? Would it stand up as claims from law and science did?

Building upon another ETSer Sam Messick's 1994 construct-centered approach to assessment, Mislevy's work around ECD allowed ETS to develop unique expertise in creating frameworks to guide assessment development. Properly working as a non-profit, ETS shared this knowledge worldwide under the organizational leadership of R&D Senior Vice President Ida Lawrence and CEO Kurt Landgraf. These frameworks could be used to operationalize constructs to foster connections among assessments, teacher development, and student learning. They could even advance efforts distilled in my personal rallying cry: No test but for learning! Make the evidence to support a claim about a learner the *actual* work they did in class and not some snapshot in an end-of-year exam. Create a loop of information that would guide and shape the instruction for each student. It didn't happen then for a swarm of reasons including the expense involved to build out the infrastructure needed just to be clearer about what exactly were the claims schools would want to make. But that doesn't mean it couldn't happen. And if it ever does, Bob's work would be at the foundation of such a valuable invention.

Bob Mislevy and his co-author Michelle Risconcente (2005) were clear that more transparency was needed in testing. They wanted the educational establishment to make "explicit the structures of assessment arguments, the elements and processes through which they are instantiated, and the interrelationships among them" (p. 61). And he agreed with another ETS colleague and friend, the distinguished researcher Michael Kane, who once pointed out that there are limitations to a claim, misadventures that can result when we extrapolate from a claim. Bob understood that and throughout his professional life kept trying to get us to be more enlightened and innovative in our thinking about assessments. He spent years looking at how games and simulations might yield more useful evidence for particular claims about what someone knew or could do. With Almond and Lukas (2004), Bob wrote,

"Advances in cognitive and instructional sciences stretch our expectations about the kinds of knowledge and skills we want to develop in students, and the kinds of observations we need to evidence them (Glaser et al., 1987). Off-the-shelf assessments and standardized tests are increasingly unsatisfactory for guiding learning and evaluating students' progress." (p. 2)

He was a pro-learning guy, not necessarily a pro-testing guy.

In fact, he cared about helping the opt-out movement and the anti-testing "memers" better understand the nature of a claim. He told me that they would be advantaged in what they really cared about—the welfare and flourishing of their children—if they had this knowledge so that they could then go to their schools, to the superintendents, to the teachers, to their politicians and say, "You need to make better claims about our children. You are misusing the educational measurement system."

Bob used to emphasize to me and others the difference when assessing whether somebody can do some sort of physical task, whether they can analyze the situation, and whether they could be an air traffic controller. He made sure we kept in the forefront of design discussions the type of claim we were making; claims differ in both the static or dynamic nature of their language, the amount of other skills that are interdependent upon them, and the amount of other independent variables. Even the circumstances in which somebody would then have to apply their knowledge makes for differences within the claim. In every case, Bob patiently reminded us to follow the same sequence: identify the claim we want to make and *then* look at what tasks would provide evidence that would allow us to assert that this person can do those things or does know those things.

Bob became very involved in work that both inspired ECD and has informed it: games and their possible connections to assessment. Assessment design is compatible with game design because both build on the same principles of learning. In a book chapter he co-wrote with colleagues from Cisco and other organizations (Mislevy, 2012), one of those principles is stated as "Assessment design is compatible with game design because they build on the same principles of learning." Here is Bob and his colleagues:

"When we design a game or assessment, we are determining the kinds of situations people will encounter and how they can interact with them. The art of game design is creating situations, challenges, rules, and affordances that keep players at the leading edge of what they can do. Serious games do this so that what players must learn to do to succeed in the game are important things to know and be able to do in a domain such as genetics, history, network engineering, or land use planning. The art of assessment design is creating situations such that students' actions provide information about their learning, whether as feedback to themselves, their teachers or a learning system, or other interested parties such as researchers, school administrators, or prospective employers. Each of these considerations for a game or an assessment imposes constraints on the situations we design and what can happen in them. Game-based assessments need to address them all at the same time." (p. 5)

I could go on: Bob Mislevy's impact in so many areas was immense, and I have not even touched upon his extraordinary openness and mentoring. The world was better for Bob being in it. That is what we all aspire to attain, and he did so. Rest in peace.

References

Chutel, L., & Cvorak, M. (2025, May 3). Trump showed images of 'genocide' in South Africa. One was from the war in Congo. *The New York Times*. https://www.nytimes.com/2025/05/23/us/politics/trump-south-africa-white-farmers-congo-fact-check.html

Farley, R., & Gore, D. (2025, April 3). Trump's misleading tariff chart. FactCheck.org. https://www.factcheck.org/2025/04/trumps-misleading-tariff-chart/

Glaser, R., Lesgold, A., & Lajoie, S. (1987). Toward a cognitive theory for the measurement of achievement. In R. Ronning, J. Glover, J. C. Conoley, & J. Witt (Eds.), *The influence of cognitive psychology on testing and measurement: The Buros-Nebraska Symposium on measurement and testing* (Vol. 3) (pp. 41–85). Erlbaum.

McDuffie, W., & Benadjaoud, Y. (2025, May 29). RFK Jr.'s MAHA report cited nonexistent studies. https://abcnews.go.com/Politics/rfk-jrs-maha-report-contained-existent-studies/story?id=122321059

Mislevy, R.J. (1994). Evidence and inference in educational assessment. Psychometrika 59, 439–483. https://doi.org/10.1007/BF02294388

Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2004). A brief introduction to evidence-centered design. CSE Report 632. *US Department of Education*. https://files.eric.ed.gov/fulltext/ED483399.pdf

Mislevy, R. J., Behrens, J. T., DiCerbo, K. E., Frezzo, D. C., & West, P. (2012). Three things game designers need to know about assessment. In D. Ifenthaler, D. Eseryel, & X. Ge (Eds.), *Assessment in game-based learning: Foundations, innovations, and perspectives* (pp. 59–81). Springer.

Mislevy, R. J., & Riconscente, M. M. (2025). Evidence-centered assessment design. In T. M. Haladyna & S. M. Downing (Eds.), *Handbook of test development* (pp. 61–90). Erlbaum.

Toulmin, S. E. (2002). *The uses of argument*. Cambridge University Press.

Valerie J. Shute

Professor Emerita
Mack and Effie Campbell Tyner Professor of Education
Florida State University

For Bob—My Colleague, My Friend, My Hero

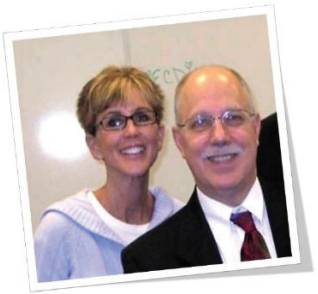
Whenever someone asks who my heroes are, Bob's name is always one of the first I say. We met in the mid-to-late 1980s—two different minds chasing the same truths. We were both working on intelligent tutoring systems in labs and classrooms and crossing paths at conferences around the world. I was in awe of him then, and even more so today.

Over the decades, we worked at some of the same places—though oddly, never at the same time. For example, I arrived at Educational Testing Service (ETS) about a month after Bob left for the University of Maryland (which I had attended back in the early 1970s). Then, when I later left ETS for Florida State University, Bob returned to ETS. My hope is that these job-hops were merely correlational, not causal!

As I described in my recent paper on the history of stealth assessment (https://myweb.fsu.edu/vshute/pdf/history.pdf), Bob's conceptualization of evidence-centered design (ECD) was key to my own work. Throughout the 1990s, I focused on identifying the cognitive, affective, and contextual factors that influence learning, building and testing intelligent tutoring systems like *Stat Lady* and *Smithtown*. But my components—developing content, measuring performance, making inferences about competency states, providing instructional support—were still separate pieces, loosely connected but lacking a unifying framework.

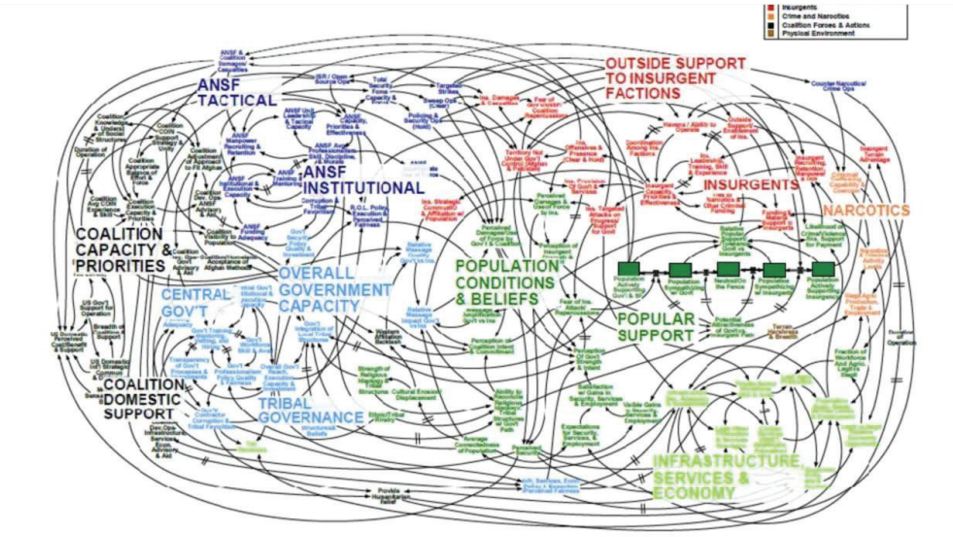
That changed in 2000 when I joined ETS and discovered exactly what I needed: Bob's framework of ECD. Just as I was searching for a way to integrate my work on stealth assessment, Bob (along with Russell Almond, Linda Steinberg, and others) was publishing groundbreaking work on ECD. After reading everything I could find—especially Mislevy, Steinberg, & Almond (2003)—I finally had a theoretical foundation to bring my ideas together and "model and support" student learning in a coherent way.

ECD was brilliant, but often a challenge to present to educational audiences (who sometimes choked on all the moving parts). So I wrote a short primer, *ECD for Dummies* (https://myweb.fsu.edu/vshute/ECD.pdf), hoping to broaden its reach. Bob kindly gave it his blessing.



Across the years, I always was excited to see Bob's name listed for attending some conference or meeting I was attending. We'd carve out some time for one-on-one conversations that *always* moved me forward (though I'm not sure I ever moved him anywhere). One memory stands out: a meeting in DC hosted by the Advanced Distributed Learning Initiative, attended mostly by military officers and a few civilian scientists. Here's a photo of us at that meeting:

I presented my ideas on using stealth assessment to measure and support systems thinking and argued why systems thinking skills were so important for military personnel. To illustrate, I compared how President Bush and President Obama thought about the War in Afghanistan. Bush's approach—simplified as "send in troops, kill bad guys, come home"—contrasted with the more complex picture of the Obama administration's view circa 2009 (see picture below).



I'm pretty sure I saw Bob cringe a little. Afterward, he gently suggested that, while he understood my point, framing it apolitically might be wiser—especially with a military audience. Classic Bob: thoughtful, kind, and always helping me sharpen my message without ever discouraging my passion.

What Bob gave me—gave all of us—was far more than theories or frameworks. As suggested above, his work on ECD made mine possible. The stealth assessments I embedded in games stand on the shoulders of his quiet genius. And it wasn't just his brilliant mind I admired. It was his kindness. His patience. That rare gift of making everyone around him want to rise. He never forgot that behind every test, every score was a human being—a story worth telling, *always in context*. The photo below is from my online retirement celebration in 2021, where Bob said some kind things about me. That meant the world.

Now I find myself missing Bob deeply, searching for words bright enough to honor his legacy, trying not to cry. But maybe it's enough to simply say Bob shaped not just how I think, but who I strive to be. I was so very lucky to know him. Rest well, my friend. My hero. My colleague.

Your light lives on—in every mind you touched, and every heart you inspired.

André A. Rupp

Senior Associate
Center for Assessment

Moments of Humanity

This year, we lost one of the truly great and most influential figures in our field—my colleague, friend, and mentor Robert (Bob) Mislevy. I got the news of his passing in a text on a Sunday morning while having breakfast with my lovely son and wife. The suddenness broke my heart—he went far too soon and left behind a wonderful family that deserved more time with him. It is heartening, though, that there have been multiple organized gatherings and outreach efforts by places and organizations he has touched to honor him, which shows the lasting humanistic impact he had.

As anyone who knew Bob can surely attest, he was deeply intellectual, thoughtful, patient, generous, and compassionate. He profoundly shaped how people across disciplines think about and design assessments, most recently in the context of evidence-centered design and sociocognitive assessment, though he also made many indelible marks through technical solutions in Bayesian statistics and psychometrics. He had a knack for writing in accessible terms about complex and generative ideas, adeptly synthesized from various disciplines.

I fondly remember chatting with Bob at the National Council on Measurement in Education (NCME) last year, laughing and reminiscing, and taking a happy photo together that I saved for fun in my "celebrity" folder. Little did I know that that was the last time I would see him in person. I owe Bob so much in my career and life—many insights, opportunities, and moments of humanity—and I know many others who feel similarly.

There are many things that Bob did for me and with me that left a profound impact on me. He invited me to be a co-author on publications, invited me to be part of grants with him, and encouraged me to apply for the faculty position at the University of Maryland where I eventually became his colleague. Without him, I surely would not have had several highlights of my career.

Beyond opportunities, he always supported me in very personal and incredibly powerful ways. He always had a kind smile and an open door for when I needed a bit of advice about how to manage my career. He would say things like, "Well, this is certainly something that one could do, but it is not something that I would choose to do," to nudge me gently toward more productive engagements. I recall his late wife telling me that his strongest expression of being upset was to say, "Well, that is different!" rather than "That is interesting!" I always found this funny, especially since I tend to curse much more at the technology and models I build when I am frustrated. Not Bob. He simply exuded kindness, patience, and grace.

On a more personal note, when I came to the United States as a new immigrant with my ex-wife, Bob spent a whole Saturday with me to help me find a used car because I urgently needed one where I lived. Since I did not have any credit history in the United States at that time, he offered to co-finance the car with me, even though he did not really know me that deeply at that point. I still cannot believe he did that. My current wife and I still drive that car today. After almost 350,000 miles, the "Bobmobile" is still going, and I think of him every day I drive it.

For these and many other reasons, all I can say is this: Bob, you will be sorely missed—as an academic, a colleague, a family member, and just a fabulous human being overall. Your legacy and spiritual trace will remain a guiding light in our field.

Drew H. Gitomer

Professor Emeritus
Former Rose and Nicholas Demarzo Chair in Education
Rutgers University

In Gratitude: The Gift of Bob Mislevy

I joined Educational Testing Service (ETS) in 1985, a few months after Bob did, and that is when I had the good fortune to join the multitude of researchers who were forever changed by their relationship with Bob. The brilliance and magnitude of his accomplishments and contributions are legion, beginning when he was a student at the University of Chicago and continuing through his time at ETS, then the University of Maryland, and again at ETS. In this brief tribute, I would like to highlight core attributes that I believe made Bob so very special, no matter what he was working on, who he was working with, or where he was working.

First, Bob asked questions that fundamentally altered the field's trajectory and those of researchers worldwide. He could have had a stellar career by taking his prodigious psychometric skills to address all manner of measurement problems. But he went much further than that by engaging with philosophy, legal reasoning, history, cognitive science, emerging technologies, teaching, and a range of workplaces to address questions that did not simply ask how can the field of psychometrics address measurement problems, but rather what does the field of psychometrics need to be and how does it need to transform to solve important problems in education and in the larger society.

Second, Bob was as humble a person of greatness as one could ever envision. Not only did that make everyone around him feel valuable, but it also made his own work much stronger. And that was because he took people's ideas and problems seriously, no matter their professional status or training. Whether it was a graduate student, junior researcher, data analyst, test developer, teacher, technician, or dental hygienist, Bob would ask questions, value their perspectives, and elevate their contributions for others to see. He would also draw on the knowledge and experience of others to challenge and evolve his own thinking, and ultimately that of the entire field.

Third, Bob was a model for all of us on how to bridge the tenuous dichotomy between applied and theoretical. Two noted examples include developing new estimation methods to address NAEP scoring problems and helping CISCO address training and certification problems for computer network engineers by advancing an assessment design theory. He never seemed to lose sight of either the specific applied problem to be solved or the theoretical implications represented by the problem.

Finally, Bob was the consummate teacher, one who could shift codes depending on who he was working with. He could talk in equations with the most technical of psychometricians, but he could also use pictures, stories, analogies, and metaphors to help the most non-technical people achieve a deep understanding of the meaning and implications of particular ideas. Indeed, his clarity of communication has enabled evidence-centered design to resonate widely across education.

It was a gift to know and work with and call Bob a friend for so long. He elevated us by helping us to be better scholars and better colleagues. Even in sadness at his death, thinking of this brilliant, wise, and gentle soul always leaves me with a smile.

Michael J. Feuer

Professor of Education
Professor of International Relations
The George Washington University

Soulful Psychometrician

Counting Bob Mislevy as one of my most beloved colleagues, I wanted very much to be present for this celebration of his life. Alas, my travel schedule was uncooperative.

I wandered into the world of measurement some 35 years ago, a public policy guy with training in economics and econometrics who therefore thought he knew everything relevant to advise the government on just about anything. Luckily, I was soon introduced to Bob Mislevy, who actually did know most of what mattered related to issues in testing and assessment, and well beyond that, and whose approach to the advancement of rational thinking about tests and their uses became an indispensable source of inspiration.

How and why did Bob stand out in a field of accomplished scholars and methodologists? In a few ways. Of course, he had an exquisite grasp of models of human cognition and cognitive measurement, and his contributions to the scholarly literature are justifiably legendary. Perhaps as or more important, though, was his appreciation for the broader context of this work and its potential for actually making a difference in people's lives. That is, he was concerned with the uses (and misuses) of assessment well before such concerns became fashionable, and he demonstrated empathy long before that term became a goal of psychometric science.

Indeed, Bob was one of the most *soulful* psychometricians I have ever known, someone whose generosity of spirit and agility of mind made him a role model for rational discourse even among people who modestly thought they had all the answers (not to mention all the questions). I remember once watching Bob at a National Academy of Sciences meeting and thinking, wow, here is living proof that knowledge does not always inversely correlate with humility. He was among the most prodigious volunteers to the work of that august body; he was always available, always eager to talk, loved to hear what people were thinking, relished the idea that consensus-seeking might promote new knowledge, and never wanted the spotlight that he *truly* deserved (which many others modestly thought *they* deserved).

I am grateful for my good luck in having had Bob Mislevy as a friend, and I know that his memory will be a blessing for our whole community.

Alina A von Davier

Chief of Assessment
Duolingo
Honorary Research Fellow, AERA, University of Oxford, Carnegie Mellon University

The Wave: Remembering Bob Mislevy and His Lasting Impact on Assessment Science

Bob Mislevy, a pioneering psychometrician who passed away on May 22, 2025, was not only a giant in our field but also a generous mentor and close collaborator, shaping my work and outlook in profound ways. We co-edited (with Jiangang Hao) the 2022 volume of *Computational Psychometrics*, which brought together methods for making sense of complex, digitally generated data—an effort that reflected Bob's lifelong habit of connecting principled assessment design with new technological affordances.

At Educational Testing Service (ETS), he advised the Computational Psychometrics Research Center and consistently backed my most novel ideas, giving me both courage and a rigorous scaffold for pursuing them. Our collaboration extended into game-based assessment through SimCityEDU with GlassLab, where Bob's evidence-centered design (ECD) provided the backbone for turning rich gameplay into defensible inferences about systems thinking and problem solving—work that helped the community understand how to align game design with assessment arguments. Together, we also co-wrote (with other colleagues) a concise book distilling lessons from our game-based efforts—what learners reveal through their processes and how designers can capture those signals responsibly—seeking to translate research insights into practical guidance for classrooms and product teams (Mislevy, 2024).

Beyond ETS, Bob and I continued to collaborate after I left. We finished the edited volume and then turned to an idea that thrilled both of us—a Bayesian framework for comparability in digital tests that integrated content features and process information, so score meaning could be stabilized even as tasks, interfaces, and data streams evolve. Although we did not complete that project, its spirit is pure Bob: principled arguments, explicit claims, and evidence pathways that keep pace with innovation.

His ECD work also directly influenced my teams. At ACTNext, it inspired our extended ECD (e-ECD/xECD) line of work (Arieli-Attali et al., 2019), which we used to formalize learning-related aspects in the student, evidence, and task models and to articulate the learning–assessment relationship in complex, interactive environments. Later at Duolingo, those same ECD principles helped us outline the Duolingo English Test (DET) "assessment ecosystem," a digitally informed chain of inferences that connects construct definition, design, measurement, security, and test-taker experience into a coherent validity argument (Burstein et al., 2022; Langenfeld et al., 2022).

In every one of these endeavors, Bob's voice nudged us back to first principles: what claim are you making about the learner, and what evidence—elicited where, when, and how—justifies that claim? That clarity powered our SimCityEDU studies, our computational psychometrics agenda, and our unfinished work on Bayesian comparability; it also lives on in the frameworks my teams use today.

The intellectual lineage from ECD to e-ECD to the DET ecosystem is, to me, the clearest testament to Bob's lasting impact: an elegant argument structure sturdy enough to support new data types, new interactions, and new threats, without losing sight of learners themselves. More personally, Bob's mentorship—his steady encouragement, incisive questions, and unfailing kindness—made room for my own voice. I will miss him dearly, and I will keep building with the tools he taught us to use.

References

Arieli-Attali, M., Ward, S., Thomas, J., Deonovic, B., & von Davier, A. A. (2019). The expanded evidence-centered design (e-ECD) for learning and assessment systems: A framework for incorporating learning goals and processes with assessment design. *Frontiers in Psychology, 10*, 853. https://doi.org/10.3389/fpsyg.2019.00853

Burstein, J., LaFlair, G., Kunnan, A. J., & von Davier, A. A. (2022). A theoretical assessment ecosystem for a digital-first assessment—*The Duolingo English Test*. Duolingo Research Report DRR-21-04: 1-32. Duolingo.

Langenfeld, T., Burstein J., & von Davier, A. A. (2022). Digital-first learning and assessment systems for the 21st century. *Frontiers in Education*, *7*, 857604. https://doi.org/10.3389/feduc.2022.857604

Mislevy, R. J. (2024). Sociocognitive and argumentation perspectives on psychometric modeling in educational assessment. *Psychometrika*, *89*(1), 64–83. https://doi.org/10.1007/s11336-024-09966-5

von Davier, A. A., Mislevy, R. J., & Hao, J. (Eds.). (2022). *Computational psychometrics: New methodologies for a new generation of digital learning and assessment: With examples in R and Python*. Springer.

David M. Williamson

Vice President of Psychometrics
College Board (Retired)

Delivered at Robert J. Mislevy Memorial

Educational Testing Service, Princeton, NJ

September 25, 2025

A Special Kind of Challenge

I confess this is the most difficult journal submission I have ever attempted, despite the assurance of publication. I know that to eulogize Bob when I am still in denial about his passing is going to fall upsettingly short of representing who he was. Nonetheless, I am thankful to the *Journal of Writing Analytics* for arranging this remembrance and hopeful that others can share what I have omitted and, collectively, we can do justice to who he was.

In sharing a remembrance of Bob, I will omit all references to his substantial academic contributions. They stand as a testament to his genius, immortalized in numerous journal articles and books that future scholars will be able to study, apply, and build upon. Instead, I will share a few anecdotes that reveal his remarkable kindness, generosity, and goodwill in the hope that those who didn't know him can gain insight into a part of his greatness that isn't as well-documented and is sometimes lacking in those who are academically brilliant.

I came to know Bob in 1996 as a graduate student at Fordham University. I had completed my coursework and was in search of a dissertation topic while working full-time at a subsidiary of Educational Testing Service (ETS), where Bob was employed. I had taken an interest in the challenges of scoring complex computerized simulations, specifically architectural designs, and that trail led me to Bob and his colleague Russell Almond (to whom many of these same compliments apply). Over the course of four years, encompassing many hours and many meetings, Bob mentored me in Bayesian networks and in the emerging concepts of evidence-centered design necessary to pair complex observations as evidence. He served as an official external co-mentor on my dissertation and attended my defense. None of this was part of his job responsibilities, just something he simply made time for in addition to his regular work. Later, as he was leaving ETS to join the University of Maryland, he cited his experiences with me as one of his motivations for becoming a professor. In doing so, he gifted me with one of the greatest professional compliments I ever received.

Bob's incredible generosity wasn't reserved for those in psychometrics. Those in the writing community who worked with him will attest that he was always eager to collaborate, share knowledge, and *learn* from anyone who expressed an interest. These were often beyond his job responsibilities and required making space in his off time to balance these engagements with his assigned work. It is a special kind of challenge to collaborate across disciplines, for different fields have different perspectives, values, and goals, and even the language of expertise itself can be a barrier to successful cross-disciplinary collaboration. Bob embraced these differences and approached collaborations from a position of shared respect and appreciation, even when it was challenging to do so. I recall a time when we were casually chatting, and a colleague asked how a certain cross-disciplinary collaboration was going. Bob acknowledged that there were some challenges and said, "I'm starting to realize that I like to solve problems, and they like to problematize solutions." Then he laughed. In his jest, he acknowledged both that solutions are rarely perfect and that tensions always exist in choosing the best solution, especially in cross-disciplinary efforts. Incidentally, this is also the most negative thing I have ever heard Bob say. I would further argue that it wasn't intended to be negative or derogatory at all, merely an acknowledgement of the different value systems at play.

Bob's incredible kindness wasn't restricted to the academic realm. When one of our colleagues had newborn twins, he went to their house to congratulate them. While there, he earnestly asked if he could help in any way and ended up spending the afternoon doing several loads of their laundry. No one found this surprising.

These are just a few insights into who Bob Mislevy was, beyond what appears in print. I hope that these tales, paired with those from others who are contributing, can help provide a richer view of what it was like to know him. There is much to lament in Bob's passing, and my denial of his departure is not entirely without merit. In fact, I embrace the denial. The original source is elusive, but it is said that each person dies twice: once at their death and again the last time someone remembers them. Accordingly, Bob Mislevy will be alive for a very, very long time.

Edmund W. Gordon

John M. Musser Professor of Psychology, Emeritus, Yale University
Richard March Hoe Professor, Emeritus of Psychology and Education, Teachers College, Columbia University

A Humanistic Behavioral Scientist (with a Philosophic Bent)

My name is Edmund W. Gordon. I have, perhaps, lived longer than any person in this room. I earned my first graduate degree from Howard University under the mentorship of Professors Alaine Lock and Howard Thurman at Howard University 80 years ago, at a time when Black people were not permitted to teach in tier-one or tier-two universities. But I have subsequently been taught and mentored by some of the finest scholars the world has produced.

Among them are W.E.B DuBois, Kenneth and Mamie Clark, Ralph Bunche, Doxey Wilkerson, all Black scholars at Howard. At Harvard I learned from Jack Carrol and Jerry Brunner; at Yale, James Comer, John Dollard, W.M. McGuire, and Edward Zigler; at Stanford, I studied with Cronbach and Snow, I collaborated with Calfee, and I mentored Hakuta; at the University of Pittsburgh, I studied and worked with Glazer, Greeno, and Resnick. Here at Educational Testing Service (ETS), I was a colleague with and learned from Benett, Messick, and Robert Mislevy.

These, as you will acknowledge, are giant scholars in our field.

Now, of course, I am bragging about my relationships with them, but my point is that the man we honor today stood shoulder to shoulder with these giant scholars, these mentors of mine. And I was determined to be here today to remind you that Robert Mislevy was also a gentle person. He was always kind and a very decent human being. On my scale of human decency, no one stands higher than he.

Bob was a humanistic behavioral scientist with a philosophic bent. Bob worried about how human behavior works—what are the mechanisms by which behavior functions? But Bob also worried about what these mechanisms of behavior mean for the behaving person, and for the recipients of the behavior so enacted. Most behavior is intentional and has meaning.

Robert Mislevy stands tall among these giants with whom I was privileged to study and work.

Rest well, my friend Bob. You deserve your seat in heaven, at the right hand of the Gods.

You have earned this reward.

Amen!

Robert J. Mislevy Curriculum Vitae

ROBERT J. MISLEVY

EDUCATION

University of Chicago	1981	Ph.D.	Research Methodology

Dissertation: *A general linear model for the analysis of Rasch item threshold estimates.*
Advisor: R. Darrell Bock

Northern Illinois University	1974	M.S. (Honors)	Mathematics
Northern Illinois University	1972	B.S. (Summa cum laude)	Mathematics, Business

EMPLOYMENT HISTORY

2011-2021	Frederic M. Lord Chair in Measurement & Statistics, Educational Testing Service, Princeton, NJ
2011-2025	Professor Emeritus, Department of Measurement, Statistics, and Evaluation, University of Maryland, College Park, MD
2001-2011	Professor, Department of Measurement, Statistics, and Evaluation, Affiliate Professor, Joint Program in Survey Methodology (2007-2011), Affiliate Professor, Second Language Acquisition (2004-2011), University of Maryland, College Park, MD
1996-2001	Distinguished Research Scholar, Educational Testing Service, Princeton, NJ
1989-1996	Principal Research Scientist, Division of Statistics and Psychometrics Research, Educational Testing Service, Princeton, NJ
1986-1989	Senior Research Scientist and Chair, Model-Based Measurement Group, Division of Statistical and Psychometric Research and Services, Educational Testing Service, Princeton, NJ
1984-1986	Research Scientist, Division of Statistical and Psychometric Research and Services, Educational Testing Service, Princeton, NJ
1982-1984	Research Associate, National Opinion Research Center, Chicago
1977-1978	Graduate Teaching Assistant, University of Chicago. Courses in Multivariate Statistics, Dept. of Behavioral Sciences; supervisor R.D. Bock.
1978-1982	Senior Research Analyst, International Educational Services
1976-1977	Adjunct Professor, National College of Education, Downtown Chicago Campus. Course in Experimental Design.
1974-1978	Research Analyst, Institute for Educational Research, Glen Ellyn, IL

PROFESSIONAL ACTIVITIES

Honors and Awards

Annual Award for Exceptional Achievement in Educational Measurement (with Alina von Davier and Jiangang Hao) for *Computational Psychometrics: New Methodologies for a New Generation of Digital Learning and Assessment (*published 2021, Springer), National Council on Measurement in Education (2024).

Keynote Speaker. Further remarks on evidence and inference in educational assessment. International Meeting of the Psychometric Society. Bologna, Italy (2022)

Lifetime Career Achievement Award. Psychometric Society (2022).

Keynote Speaker. Beyond Results 2021 Workshop: From Log Data to Valid Inferences. International Association for the Evaluation of Educational Achievement. Hamburg (2021)

Award for Significant Contribution to Educational Measurement and Research Methodology, for *Sociocognitive Foundations of Educational Measurement* (published March 2018, Routledge), Division D of the American Educational Research Association (2020)

Annual Award for Exceptional Achievement in Educational Measurement, for *Sociocognitive foundations of educational measurement* (published March 2018, Routledge), National Council on Measurement in Education (2019).

Keynote Speaker. International Meeting of the Psychometric Society (IMPS 2018), July 12, 2018, Columbia University, New York.

Invited plenary address. International Objective Measurement Workshop (IOMW 2018), April 11, 2018, New York University, New York.

Keynote Speaker. Learning Analytics and Knowledge (LAK16). University of Edinburgh (2016).

Keynote Speaker. International Association for Computerized Adaptive Testing (IACAT) summit. Princeton, NJ (2014).

Buros Center for Testing / Educational Testing Service, Seventh Annual Lecture. University of Nebraska, Lincoln (April, 2013)

Elected *Fellow of the American Educational Research Association* (2012).

International Keynote Speaker. Annual Workshop on Item Response Theory and Educational Measurement, University of Twente, The Netherlands (2011).

Anne Anastasi Memorial Lecturer. Fordham University (2011).

Robert L. Linn Distinguished Address Award. Awarded by the American Educational Research Association, Division D. (2009).

Best Paper Award, The Fifth International Conference on Networking and Services (ICNS 2009), for "Psychometric and Evidentiary Approaches to Simulation Assessment in Packet Tracer Software" (with D. Frezzo, J. Behrens, P. West, & K. DiCerbo, Cisco Systems).

Keynote Speaker. 34th International Association for Educational Assessment (IAEA) Conference, Cambridge University (2008).

Frank B. Womer Endowed Lecture in Testing, Measurement, and Evaluation. Awarded by the School of Education, University of Michigan. (2008).

Elected to the *National Academy of Education* (2007).

E.F. Lindquist Award for outstanding applied or theoretical research in the field of testing and measurement, by the American Educational Research Association and the American College Testing Program (2007).

Samuel J. Messick Memorial Lecture Award, International Language Testing Association (2007).

Outstanding Paper Award, E-Learn 2003, for "Using Evidence-Centered Design to Develop Advanced Simulation-Based Assessment and Training" (with M. Bauer & D. Williamson, ETS, and J. Behrens, Cisco Systems).

Award for Career Contributions to Educational Measurement, National Council on Measurement in Education (2003).

Triennial Award for Technical Contribution to Educational Measurement (with L.S. Steinberg and Russell G. Almond, for "evidence-centered assessment design"), National Council on Measurement in Education (2000).

New Product Development Award for Most Outstanding Breakthrough (with L.S. Steinberg and Russell G. Almond, for the Portal assessment-design object model and tool system), Educational Testing Service (1998).

ETS Senior Research Scientist Award (1993).

Triennial Award for Application of Educational Measurement Technology (with R. D. Bock, for contributions to the design and analysis of educational assessment), National Council on Measurement in Education (1990).

Raymond B. Cattell Early Career Award for Programmatic Research, American

Educational Research Association (1988).

Triennial Award for Technical Contribution to Educational Measurement (with A. E. Beaton, E. G. Johnson, and K. M. Sheehan, for "plausible values" methodology in the National Assessment for Educational Progress), National Council on Measurement in Education (1988).

Visiting Scholar

University of California at Berkeley (March, 2013).

College Board (2001-2002).

UCLA/CRESST (July, 1996)

ACT (1995).

Central Institute for Test Development (Cito), The Netherlands (1985).

Offices Held

Mentor, Edmund W Gordon MacArthur Foundation Fellowship for 21st Century Learning and Assessment, 2013-

Member, Research Activities Committee, National Academy of Education, 2013-2015

Member, Career Award Committee, NCME, 2013-2015

Member, Robert L. Linn Award committee, 2009; Chair, 2010

Member, AERA Lindquist Award committee, 1999-2002

Program Committee, User Modeling Conference, 1997, 1999

President, Psychometric Society, 1993-1994

Board of Trustees, Psychometric Society, 1993-1996

Program Chair, Division D, AERA, 1988

Advisory Panels

Committee on Foreign Language Assessment for the U.S. Foreign Service Institute. National Academies of Science. Committee produced *A Principled Approach to Language Assessment: Considerations for the U.S. Foreign Service Institute* (2020; Nat'l Academies Press).

Study Group on Adaptive Educational Technologies, National Academy of Education.

Gordon Commission on the Future of Educational Assessment. Member; Executive Board.

Advisory Panel, "AutoMentor: Virtual Mentoring and Assessment in Computer Games for STEM Learning" to the University of Wisconsin under a grant from NSF.

Key Advisor, "Domain-Specific Assessment: Bringing the Classroom into Community College Accountability" to SRI International, under grant for the Institute of Education Sciences.

Technical Panel on 12th Grade Preparedness Research, National Assessment Governing Board.

Technical Advisory Group for the Programme for the International Assessment for Adult Competencies (PIAAC), for ETS under a contract from OECD.

Task Force on Computerized Adaptive Testing for Longitudinal Study, National Center for Educational Statistics.

Idea of Testing (Spencer Foundation). Study group produced *Assessment, equity, and opportunity to learn* (Moss et al., eds; 2008; Cambridge University Press).

Defense Language Testing Advisory Board.

National Research Council's Committee on the Foundations of Assessment. Committee produced *Knowing What Students Know: The Science and Design of Educational Assessment* (2001; Nat'l Academies Press).

Institute for Modern Mental Testing (Trustee).

The Japan Institute for Educational Measurement.

National Research Council Committee on Alternatives for Assessing Adult Education and Literacy Programs (Chair).

National Assessment Governing Board's Design/Feasibility Team. Committee produced *Design/Feasibility Team: Report to the National Assessment Governing Board* (1996; NAGB).

National Research Council Committee on Setting Standards for Literacy.

National Academy of Science's Board on Testing and Assessment (BOTA).

Contractor's Technical Advisory Group (TAG) for design and analysis of the National Assessment of Educational Progress.

National Academy of Sciences / Mathematical Sciences Education Board committee on mathematics assessment. Committee produced *Measuring What Counts: A Conceptual Guide for Mathematics Assessment.* (1993; Nat'l Academies Press).

National Board of Medical Examiners advisory panel on Standardized Patients Assessment.

Editorial Boards of Journals

Educational Psychology Review, Associate Editor.

Measurement: Interdisciplinary Research and Perspectives, Associate Editor.

Journal of Learning, Technology, and Assessment, Associate Editor.

Psychometrika, Associate Editor (1995), Guest Associate Editor (2004)

Affiliations

American Educational Research Association (program chair for Division D)

National Council on Measurement in Education

American Statistical Association

Psychometric Society (past president; former trustee, bylaws committee)

Northeastern Educational Research Association

COURSES TAUGHT

Foundations of assessment.

Cognitive psychology and educational assessment.

Theory-based task design.

Bayesian inference and measurement models.

Experimental design.

TRAINING SESSIONS TAUGHT

Designing scenario-based assessment items using an evidence-centered design framework (AERA)

Principled assessment design for inquiry (AERA)

Item response theory (Scientific Software).

Factor analysis and structural equations modeling (Scientific Software).

Bayes nets in educational assessment (AERA, NCME)

DOCTORAL DISSERTATIONS SUPERVISED

Chin-Fang Weng (2014): *First order autoregressive mixed effects zero inflated Poisson model for longitudinal data – A Bayesian approach.*

Ting Zhang (2012): *The role of reading comprehension in large-scale subject-matter assessments.* (co-chair, with Prof. Judith Torney-Purta)

Taslima Rahman (2012): *Reading comprehension and its assessment: Aligning operationalization with conceptualization of the construct.* (co-chair, with Prof. Patricia Alexander)

Tiandong Li (2012): *Randomization-based inference about latent variables from complex samples: The case of two-stage sampling.*

Younyoung Choi (2011): *Dynamic Bayesian inference networks and hidden Markov models for modeling learning progressions over multiple time points.*

Daisy Wise Rutstein (2011*): Measuring learning progressions using Bayesian modeling in complex assessments.*

Futoshi Yumoto (2011*): Effects of unmodeled latent classes on multilevel growth mixture estimation in value-added modeling.* (co-chair, with Prof. Gregory Hancock)

William Dardick (2010): *Reweighting data in the spirit of Tukey: Using Bayesian posterior probabilities as Rasch residuals for studying misfit.*

Marc Kroopnic (2010): *Exploring unidimensional proficiency classification accuracy from multidimensional data in a vertical scaling context.*

Feifei Li (2009): *An information correction method for testlet-based test analysis: From the perspectives of item response theory and generalizability theory.*

Yunyun Dai (2009): *A mixture Rasch model with a covariate: A simulation study via Bayesian Markov chain Monte Carlo estimation.*

Dongyang Li (2009): *Developing a common scale for testlet model parameter estimates under the common- item nonequivalent groups design.*

Hua Wei (2008): *Multidimensionality in the NAEP science assessment: substantive perspectives, psychometric models, and task design*

Karen Douglas (2007): *A general method for estimating the classification reliability of complex decisions based on configural combinations of multiple assessment scores.*

Roy Levy (2006): *Posterior predictive model checking for multidimensionality in item response theory and Bayesian networks.* [Received National Council of Measurement in Education's Brenda Loyd dissertation award]

Chun-Wei Huang (2003): *Psychometric analyses based on evidence-centered design and cognitive science of learning to explore students' problem-solving in physics.*

Ilona Arnold-Berkovits (2002): *Structural modeling with ordered polytomous and continuous variables: a simulation study comparing full-information Bayesian estimation to correlation covariance methods* (co-chair, with Prof. Gregory Hancock)

UNIVERSITY SERVICE

Campus

General Education Assessment Committee.

College of Education

College Senate.

College Senate Steering Committee.

Summer Reorganization Committee.

College of Education Organization Plan Committee.

Advancement, Promotion, & Tenure Committee, Chair.

Department of Measurement, Statistics and Evaluation

Examination Committee, chair.

Admissions Committee, chair.

Search Committees, as member and as chair.

PUBLICATIONS

Books

Wainer, H., Dorans, N. J., Flaugher, R., Green, B. F., Mislevy, R. J., Steinberg, L., & Thissen, D. (2000). *Computerized adaptive testing: A primer* (2nd ed.). Lawrence Erlbaum Associates.

Mislevy, R.J. (2018). *Sociocognitive foundations of educational measurement*. Routledge. (Received NCME's 2019 Award for Exceptional Achievement in Educational Measurement and AERA Division D's 2020 Award for Significant Contribution to Educational Measurement and Research Methodology.)

Mislevy, R.J., Haertel, G.D., Riconscente, M.M., Rutstein, D.W., & Ziker, C. (2017). *Assessing model-based reasoning using evidence-centered design.* Springer Nature.

Levy, R., & Mislevy, R.J. (2016). *Bayesian psychometric modeling.* Chapman & Hall/CRC.

Almond, R.G., Mislevy, R.J., Steinberg, L.S., Williamson, D.M., & Yan, D. (2015). *Bayesian networks in educational assessment.* Springer.

Mislevy, R.J., Corrigan, S., Oranje, A., DiCerbo, K., John, M., Bauer, M.I., Hoffman, E., von Davier, A.A., Hao, J. (2014). *Psychometric considerations in game-based assessment* (GlassLab Technical Report). Institute of Play. URL: https://web.archive.org/web/20160320151604/http:/www.instituteofplay.org/wp-content/uploads/2014/02/GlassLab_GBA1_WhitePaperFull.pdf

Wainer, H., Dorans, N.J., Flaugher, R., Green, B.F., Mislevy, R.J., Steinberg, L., & Thissen, D. (1990). *Computerized adaptive testing: A primer.* Lawrence Erlbaum Associates.

Mislevy, R.J. (1993). *Linking educational assessments: Concepts, issues, methods, and prospects* (ETS Policy Information Center Report). Educational Testing Service.

Edited volumes

Von Davier, A.A., Mislevy, R.J., & Hao, J. (Eds.). (2021). *Computational psychometrics: New methods for a new generation of educational assessment.* Springer.

Williamson, D.M., Mislevy, R.J., & Bejar, I.I. (Eds.). (2006). *Automated Scoring of complex performances in computer-based testing.* Lawrence Erlbaum Associates.

Mislevy, R.J., & Knowles, K. (Eds.). (2002). *Performance Assessments for Adult Education: Exploring the Measurement Issues.* National Academies Press.

Frederiksen, N., Mislevy, R.J., & Bejar, I.I. (Eds.). (1993). *Test theory for a new generation of tests.* Erlbaum.

Software

Zimowski, M., Muraki, E., Mislevy, R., & Bock, R. D. (2002). *BILOG-MG II: Multiple group item analysis and test scoring with binary logistic models*. Mooresville, IN: Scientific Software.

Zimowski, M., Muraki, E., Mislevy, R., & Bock, R. D. (1997). *BILOG-MG: Multiple group item analysis and test scoring with binary logistic models*. Mooresville, IN: Scientific Software.

Zimowski, M., Muraki, E., Mislevy, R., & Bock, R. D. (1993). *BIMAIN: Multiple group item analysis and test scoring with binary logistic models*. Mooresville, IN: Scientific Software.

Mislevy, R.J., & Bock, R.D. (1983). *BILOG: Item analysis and test scoring with binary logistic models* [computer program]. Mooresville, IN: Scientific Software, Inc.

Patents

Hao, J., Smith III, L., Mislevy, R., & von Davier, A.A. (2014). *Systems and methods for designing, parsing and mining of game log files.* (U.S. Patent Application No. 14/527,591).

R.J. Mislevy, G.D. Haertel, L.A. Hamel, C.A. Kennedy, M. Wilson. (2005). *System and method for assessment design.* (U.S. Patent Application No. 20050221266).

R.G. Almond, S. Sinharay, L. Steinberg, R.J. Mislevy. (2014). *Method and system for calibrating evidence models.* (U.S. Patent Application No. 8798518).

E.G. Hansen, R.J. Mislevy, L.S. Steinberg. (2007). *Accessibility of testing within a validity framework* (U.S. Patent No. 7217134).

Steinberg, L. S., Mislevy, R. J., & Almond, R. G. (2005). *Portal assessment design system for educational testing* (U.S. Patent No. 434350000).

Research Papers in Refereed Journals

Uher, J., Arnulf, J. K., Barrett, P. T., Heene, M., Heine, J.-H., Martin, J., Mazur, L. B., McGann, M., Mislevy, R.J., Speelmann, C., Toomela, A. & Weber, R. (2025). Psychology's questionable research fundamentals (QRFs): Key problems in quantitative psychology and psychological measurement beyond questionable research practices (QRPs). *Frontiers in Psychology,16,* Article 1553028. https://doi.org/10.3389/fpsyg.2025.1553028

Mislevy, R.J. (2024). Are sum scores a great accomplishment of psychometrics or intuitive test theory? *Psychometrika.* https://doi.org/10.1007/s11336-024-10003-8

Mislevy, R.J. (2024). Sociocognitive and argumentation perspectives on psychometric modeling in educational assessment. *Psychometrika.* https://doi.org/10.1007/s11336-024-09966-5

Gong, T., Shuai, L., & Mislevy, R.J. (2023). Sociocognitive processes and item response models: A didactic example. *Journal of Educational Measurement.* https://doi.org/10.1111/jedm.12376

Choi, Y., & Mislevy, R.J. (2022). An Evidence-Centered Design framework and dynamic Bayesian networks for modeling learning progression in a formative assessment system. *Frontiers in Psychology,* 13, Article 742956. https://doi.org/10.3389/fpsyg.2022.742956

Oliveri, M. E., Mislevy, R., & Slomp, D. (2021). Principled development of workplace English communication, Part 1: A sociocognitive framework. *Journal of Writing Analytics, 5,* 34-70. https://doi.org/10.37514/JWA-J.2021.5.1.02

Oliveri, M. E., Slomp, D., Rupp, A.A., & Mislevy, R.J. (2021). Principled development of workplace English communication, Part 2: Expanded Evidence-Centered Design and Theory of Action frameworks. *Journal of Writing Analytics, 5,* 71-108. https://doi.org/10.37514/JWA-J.2021.5.1.03

Oliveri, M. E., Slomp, D., Rupp, A.A., & Mislevy, R.J. (2021). Principled development of workplace English communication, Part 3: An integrated design and appraisal framework. *Journal of Writing Analytics, 5,* 109-141. https://doi.org/10.37514/JWA-J.2021.5.1.04

Oliveri, M.E., Slomp, D.H., Elliot, N., Rupp, A.A., Mislevy, R.J., Vezzu, M., Tackitt, A., Nastal, J., Phelps, J., & Osborn, M. (2021). Introduction: Meeting the challenges of workplace English communication in the 21st Century. *Journal of Writing Analytics, 5,* 1-33. https://doi.org/10.37514/JWA-J.2021.5.1.01

Mislevy, R.J. (2020). Statistical theoreticians and educational assessment: Comments on Shelby Haberman's NCME Career Contributions Award. *Journal of Educational Measurement, 57,* 386-396.

Mislevy, R.J., & Oliveri, M. E. (2019). Digital Module 09: Sociocognitive Assessment for Diverse Populations. *Educational Measurement: Issues and Practice*, *38*(4), 110-111.

Mislevy, R.J. (2019). Advances in the science of measurement and cognition. *Annals of the American Academy of Politics & Social Science, 683*(1), 164-182.

Hao, J., & Mislevy, R.J. (2019). Characterizing interactive communications in computer-supported collaborative problem-solving tasks: A conditional transition profile approach. *Frontiers in psychology, 10,* 1011.

Oliveri, M.E., Lawless, R.R., & Mislevy, R.J. (2019). Using evidence-centered design to support the development of culturally and linguistically sensitive collaborative problem-solving assessments. *International Journal of Testing, 19*(3), 270-300. https://doi.org/10.1080/15305058.2018.1543308

Oliveri, M. E. & Mislevy, R. (2019). Introduction to challenges and opportunities in the design of 'next-generation assessments of 21st century skills [Special Issue]. *International Journal of Testing*, *19*, 97-102.

Bauer, M., Wylie, C., Jackson, T., Mislevy, R.J., Hoffman-John, E., John, M. (2017). Why video games can be a good fit formative assessment. *Journal of Applied Testing and Technology, 18*(1), 19-31.

Andrews, J. J., Kerr, D., Mislevy, R.J., von Davier, A.A., Hao, J., & Liu, L. (2017). Modeling collaborative interaction patterns in a simulation-based task. *Journal of Educational Measurement*, *54*(1), 54-69.

Mislevy, R.J. (2016). How developments in psychology and technology challenge validity argumentation. *Journal of Educational Measurement, 53*, 265-292.

Zhang, T., Torney-Purta, J., & Mislevy, R. (2015). Understanding civic cognitive assessment tasks: Associations between linguistic features and student performance. *Citizenship Teaching and Learning, 11*, 29-47.

Dardick, W.R., & Mislevy, R.J. (2015). Reweighting data in the spirit of Tukey: Using Bayesian posterior probabilities as Rasch residuals for studying misfit. *Educational and Psychological Measurement, 75,* 1-26.

Mislevy, R.J., & Duran, R.P. (2014). A sociocognitive perspective on assessing EL students in the age of common core and Next Generation Science Standards. *TESOL Quarterly*, *48*, 560-585.

Mislevy, R.J. (2013). Measurement is a necessary but not sufficient frame for assessment. *Measurement: Interdisciplinary Research & Perspectives, 11*, 47-49.

Mislevy, R.J. (2013). Evidence-centered design for simulation-based assessment. *Military Medicine, 178,* 107-114. Mislevy, R.J., Haertel, G., Cheng, B.H., Ructtinger, L., DeBarger, A., Murray, E., Rose, D., Gravel, J., M. Colker, A.M., Rutstein, D., & Vendlinski, T. (2013). A "conditional" sense of fairness in assessment. *Educational Research and Evaluation, 19,* 121–140.

Mislevy, R.J., & Zwick, R.J. (2012). Scaling, linking, and reporting in a periodic assessment system. *Journal of Educational Measurement*, *49*, 148–166.

Mislevy, R.J., Behrens, J.T., DiCerbo, K., & Levy, R. (2012). Design and discovery in educational assessment: Evidence centered design, psychometrics, and data mining. *Journal of Educational Data Mining, 4,* 11-48. Available online at http://www.educationaldatamining.org/JEDM/images/articles/vol4/issue1/MislevyEtAlVol4Issue1P11_48.pdf

Rupp, A.A., Levy, R., DiCerbo, K.E., Sweet, S., Crawford, A.V. Caliço, T., Benson, M., Fay, D., Kunze, K.L., Mislevy, R.J., & Behrens, J.T. (2012). Putting ECD into practice: The interplay of theory and data in evidence models within a digital learning environment. *Journal of Educational Data Mining, 4,* 49-110. Available online at http://www.educationaldatamining.org/JEDM/images/articles/vol4/issue1/RuppEtAlVol4Issue1P49_110.pdf

Mislevy, R.J. (2012). The case for informal argument. *Measurement: Interdisciplinary Research & Perspectives*, *10*, 93–96.

Mislevy, R,J. (2010). Design under constraints: The case of large-scale assessment systems. *Measurement: Interdisciplinary Research and Perspectives, 8,* 199-203.

Mislevy, R,J. (2010). Some implications of cognitive psychology for educational assessment. *Research Papers in Education*, *25*, 253-270.

Douglas, K. M., & Mislevy, R.J. (2010). Estimating classification accuracy for complex decision rules based on multiple scores. *Journal of Educational and Behavioral Statistic*s, 35, 280-306.

Rupp, A., Gushta, M., Mislevy, R.J., & Shaffer, D.W. (2010). Evidence-centered design of epistemic games: Measurement principles for complex learning environments. *Journal of Technology, Learning, and Assessment*, *8*(4).

Mislevy, R.J., Behrens, J.T., Bennett, R.E., Demark, S.F., Frezzo, D.C., Levy, R., Robinson, D.H., Rutstein, D.W., Shute, V.J., Stanley, K., & Winters, F.I. (2010). On the roles of external knowledge representations in assessment design. *Journal of Technology, Learning, and Assessment, 8*(2). http://napoleon.bc.edu/ojs/index.php/jtla/article/viewFile/1621/1465

Frezzo, D.C., Behrens, J.T., & Mislevy, R.J. (2009). Design patterns for learning and assessment: facilitating the introduction of a complex simulation-based learning environment into a community of instructors. *The Journal of Science Education and Technology.* Springer Open Access http://www.springerlink.com/content/566p6g4307405346/

Mislevy, R.J., & Yin, C. (2009). If language is a complex adaptive system, what is language testing? *Language Learning,* 59, Supplement 1, 249-267.

Levy, R., Mislevy, R.J., & Sinharay, S. (2009). Posterior predictive model checking for multidimensionality in item response theory. *Applied Psychological Measurement, 33*, 519-537.

Frezzo, D.C., Behrens, J.T., & Mislevy, R.J. (2009). Activity theory and assessment theory in the design and understanding of the Packet Tracer ecosystem. *The International Journal of Learning and Media*, *2*. http://ijlm.net/knowinganddoing/10.1162/ijlm.2009.0015

Shaffer, D.W., Hatfield, D., Svarovsky, G.N., Nash, P., Nulty, A., Bagley, E., Frank, K., Rupp, A.A., & Mislevy, R.J. (2009). Epistemic Network Analysis: A Prototype for 21st-Century Assessment of Learning. *The International Journal of Learning and Media*, *2*, 33-53. http://ijlm.net/fandf/doi/abs/10.1162/ijlm.2009.0013

Mislevy, R.J. (2008). How cognitive science challenges the educational measurement tradition. *Measurement: Interdisciplinary Research and Perspectives, 6,* 124. Available online at http://bearcenter.berkeley.edu/measurement/docs/CommentaryHaig_Mislevy.pdf

Mislevy, R.J. (2007). Validity by design. *Educational Researcher, 36,* 463-469.

Mislevy, R.J., & Haertel, G. (2006). Implications of evidence-centered design for educational assessment. *Educational Measurement: Issues and Practice, 25,* 6-20.

Hansen, E.G., Mislevy, R.J., Steinberg, L.S., Lee, M.J., & Forer, D.C. (2005). Accessibility of tests within a validity framework. *System: An International Journal of Educational Technology and Applied Linguistics, 33,* 107-133.

Mislevy, R.J. (2004). Can there be reliability without "reliability"? *Journal of Educational and Behavioral Statistics, 29,* 241-244.

Behrens, J.T., Mislevy, R.J., Bauer, M., Williamson, D.M., & Levy, R. (2004). Introduction to evidence centered design and lessons learned from its application in a global E-Learning program. *International Journal of Testing, 4,* 295-301.

Williamson, D.M., Bauer, M., Steinberg, L.S., Mislevy, R.J., Behrens, J.T., & DeMark, S. (2004). Design rationale for a complex performance assessment. *International Journal of Testing, 4,* 303-332.

Levy, R., & Mislevy, R.J. (2004). Specifying and refining a measurement model for a simulation-based assessment. *International Journal of Testing, 4,* 333-369.

Mislevy, R.J. (2003). Substance and structure in assessment arguments. *Law, Probability, and Risk, 2,* 237-258.

Mislevy, R.J. (2003). Rehabilitating psychometrics: Commentary on Pellegrino and Chudowsky "The Foundations of Assessment." *Measurement: Interdisciplinary Research and Perspectives*, *1*, 162-165.

Mislevy, R.J., Steinberg, L. S., & Almond, R. A. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, 1, 3-67.

Mislevy, R.J., Steinberg, L. S., Breyer, F. J., Johnson, L., & Almond, R. A. (2002). Making sense of data from complex assessments. *Applied Measurement in Education, 15, 363-378.*

Mislevy, R.J., Steinberg, L.S., & Almond, R.A. (2002). Design and analysis in task-based language assessment. *Language Testing, 19,* 477-496.

Almond, R.G., Steinberg, L.S., & Mislevy, R.J. (2002). Enhancing the design and delivery of assessment systems: A four-process architecture. *Journal of Technology, Learning, and Assessment, 1*(5). http://www.bc.edu/research/intasc/jtla/journal/v1n5.shtml

Mislevy, R.J. (2000). Modeling in complex assessments. *The NERA Researcher, 38,* 7-17.

Mislevy, R.J., & Chang, H.H. (2000). Does adaptive testing violate local independence? *Psychometrika, 65*(2), 149–156. https://doi.org/10.1007/BF02294370

Cameron, C.A., Beemsterboer, P.L., Johnson, L.A., Mislevy, R.J., Steinberg, L.S., & Breyer, F.J. (2000). A cognitive task analysis for dental hygiene. *Journal of Dental Education, 64,* 333-351.

Mislevy, R.J., Steinberg, L.S., Breyer, F.J., Almond, R.G., & Johnson, L. (2000). A cognitive task analysis, with implications for designing a simulation-based assessment system. *Computers and Human Behavior, 15,* 335-374.

Almond, R.G., & Mislevy, R.J. (1999). Graphical models and computerized adaptive testing. *Applied Psychological Measurement,* 23, 223-237.

Mislevy, R.J. (1998). Implications of market-basket reporting for achievement level setting. *Applied Measurement in Education*, *11*, 49-63.

Mislevy, R.J. (1996). Test theory reconceived. *Journal of Educational Measurement, 33,* 379-416.

Mislevy, R.J., & Gitomer, D.H. (1996). The role of probability-based inference in an intelligent tutoring system. *User-Modeling and User-Adapted Interaction, 5,* 253-282.

Mislevy, R.J., & Wilson, M.R. (1996). Marginal maximum likelihood estimation for a psychometric model of discontinuous development. *Psychometrika, 61,* 41-71.

Béland, A., & Mislevy, R.J. (1996). Probability-based inference in a domain of proportional reasoning tasks. *Journal of Educational Measurement, 33,* 3-27.

Mislevy, R.J. (1995). What can we learn from international assessments? *Educational Evaluation and Policy Analysis, 17,* 419-437.

Mislevy, R.J. (1995). Test theory and language learning assessment. *Language Testing, 12,* 341-369.

Mislevy, R.J. (1994). Evidence and inference in educational assessment. *Psychometrika, 59*, 439-483.

Mislevy, R.J. (1994). Mathematics assessment and mathematical thinking. *Assessment in Practice, 1*, 3-4, 7.

Mislevy, R.J., Sheehan, K.M., & Wingersky, M.S. (1993). How to equate tests with little or no data. *Journal of Educational Measurement, 30*, 55-78, 1993.

Mislevy, R.J. (1993). Should "multiple imputations" be treated as "multiple indicators?" *Psychometrika, 58*, 79-85.

Mislevy, R.J. (1993). Some formulas for use with Bayesian ability estimates. *Educational and Psychological Measurement, 53*, 315-328.

Wainer, H., Johnson, E.G., Lewis, C., & Mislevy, R.J. (1993). Some research problems encountered at the Educational Testing Service. *Journal of Official Statistics, 9*, 189-201.

Mislevy, R.J., Beaton, A.E., Kaplan, B., & Sheehan, K.M. (1992). Estimating population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement, 29,* 133-161.

Mislevy, R.J., Johnson, E.G., & Muraki, E. (1992). Scaling procedures in the National Assessment for Educational Progress. *Journal of Educational Statistics*, *17,* 131-154.

Mislevy, R.J. (1991). Randomization-based inference about latent variables from complex samples. *Psychometrika, 56*, 177-196.

Mislevy, R.J., Wingersky, M.S., Irvine, S.H., & Dann, P.L. (1991). Resolving mixtures of strategies in spatial visualization tasks. *British Journal of Mathematical and Statistical Psychology, 44,* 265-288.

Sheehan, K.M., & Mislevy, R.J. (1990). Integrating cognitive and psychometric models in a measure of document literacy. *Journal of Educational Measurement, 27*, 255-272.

Mislevy, R.J., & Verhelst, N. (1990). Modeling item responses when different subjects employ different solution strategies. *Psychometrika*, *55*, 195-215.

Mislevy, R.J., & Stocking, M.L. (1989). A consumer's guide to LOGIST and BILOG. *Applied Psychological Measurement*, *13*, 57-75.

Mislevy, R.J., & Sheehan, K.M. (1989). Information matrices in latent-variable models. *Journal of Educational Statistics*, *14*, 335-350.

Mislevy, R.J., & Sheehan, K.M. (1989). The role of collateral information about examinees in item parameter estimation. *Psychometrika*, *54*, 661-679.

Bock, R.D., & Mislevy, R.J. (1988). Comprehensive educational assessment for the states: The duplex design. *Educational Evaluation and Policy Analysis, 10,* 89-105.

Mislevy, R.J. (1988). Exploiting auxiliary information about items in the estimation of Rasch difficulty parameters. *Applied Psychological Measurement, 12*, 281-296.

Mislevy, R.J. (1987). Exploiting auxiliary information about examinees in the estimation of item parameters. *Applied Psychological Measurement, 11*, 81-91.

Mislevy, R.J. (1987). Recent developments in item response theory. *Review of Research in Education, 15*, 239-275.

Mislevy, R.J. (1986). Bayes modal estimation in item response models. *Psychometrika, 51*, 177-195.

Mislevy, R.J. (1986). Estimation of latent group effects. *Journal of the American Statistical Association, 80*, 993-997.

Mislevy, R.J. (1986). Recent developments in the factor analysis of categorical data. *Journal of Educational Statistics, 11*, 3-31.

Mislevy, R.J. (1984). Estimating latent distributions. *Psychometrika, 49*, 359-381.

Mislevy, R.J. (1983). Item response models for grouped data. *Journal of Educational Statistics, 8*, 271-288.

Bock, R.D., & Mislevy, R.J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement, 6*, 431-444.

Mislevy, R.J., & Bock, R.D. (1982). Biweight estimates of latent ability. *Educational and Psychological Measurement, 42*, 725-737.

Bock, R.D., Mislevy, R.J., & Woodson, C.E.M. (1982). The next stage in educational assessment. *Educational Researcher, 11*, 4-11, 16.

Bock, R.D., & Mislevy, R.J. (1981). An item response curve model for matrix-sampling data: The California grade three assessment. *New Directions for Testing and Measurement, 10*, 65-90.

Chapters in Edited Volumes

Mislevy, R.J., Oliveri, M.E., Slomp, D., Crop Eared Wolf, A., & Elliot, N. (2025). An evidentiary-reasoning lens for socioculturally responsive assessment. In R. E. Bennett, L. Darling-Hammond, & A. Badrinarayan (Eds.). *Socioculturally responsive assessment: Implications for theory, measurement, and systems-level policy* (pp. 199-241). Routledge.

Mislevy, R.J. (2025). An evidentiary-reasoning perspective on culturally responsive assessment: Commentary on Section 2, Large-scale assessment. In C.M. Evans & C.S. Taylor (Eds.), *Culturally responsive assessment in classrooms and large-scale contexts: Theory, research, and practice* (pp. 228-242). Routledge.

Mislevy, R.J. (2023). Foreword. In L. Mari, M. Wilson, & A. Maul. (2023). *Measurement across the sciences: Developing a shared concept system for measurement* (pp. ix-xviii). Cham, Switzerland: Springer Nature.

Yin, C., & Mislevy, R.J. (2023). Evidence-centered Design. In X. Liu, D. McCaffrey, & R. Lawless (Eds.), *Quantitative research* (pp. 163-170). Oxford, UK: Elsevier Ltd. This volume is part of R.J. Tierney, F. Rizvi, & K. Ercikan (Eds.), *International Encyclopedia of Education, 4th Edition.* https://doi.org/10.1016/B978-0-12-818630-5.10023-5

Yin, C., & Mislevy, R.J. (2022). Evidence-centered design in language testing. In G. Fulcher & L. Harding (Eds.). *Routledge Handbook of Language Testing* (pp. 289-305). London: Routledge.

Bell, C., & Mislevy, R.J. (2021). Practice, feedback, argument, and measurement: Metaphors for analyzing assessment of teaching. In M. Blikstad-Balas (Ed.), *Analyzing teaching quality: Perspectives, principles and pitfalls* (pp. 21-52). Oslo: Scandinavian University Press.

Von Davier, A.A., Mislevy, R.J., & Hao, J. (2021). Introduction to computational psychometrics: Towards a principled integration of data science and machine learning techniques into psychometrics. In A.A. von Davier, R.J. Mislevy, & J. Hao (Eds.). *Computational psychometrics: New methods for a new generation of educational assessment* (pp. 1-6). Springer.

Mislevy, R.J. (2021). Next generation learning and assessment: What, why and how. In A.A. von Davier, R.J. Mislevy, & J. Hao (Eds.). *Computational psychometrics: New methods for a new generation of educational assessment* (pp. 9-24). Springer.

Andrews-Todd, J., Mislevy, R.J., LaMar, M., & de Klerk, S. (2021). Virtual performance-based assessments. In A.A. von Davier, R.J. Mislevy, & J. Hao (Eds.). *Computational psychometrics: New methods for a new generation of educational assessment* (pp. 45-60). Springer.

Mislevy, R.J., & Bolsinova, M. (2021). Concepts and models from psychometrics. In A.A. von Davier, R.J. Mislevy, & J. Hao (Eds.). *Computational psychometrics: New methods for a new generation of educational assessment* (pp. 81-107). Springer.

Hao, J., & Mislevy, R.J. (2021). A data science perspective on computational psychometrics. In A.A. von Davier, R.J. Mislevy, & J. Hao (Eds.). *Computational psychometrics: New methods for a new generation of educational assessment* (pp. 133-158). Springer.

Levy, R., & Mislevy, R.J. (2021). History of Bayesian inference in educational measurement. In B.E. Clauser & M.B. Bunch (Eds.), *The history of educational measurement* (pp. 292-317). Routledge.

Mislevy, R.J. (2021). Construct representation and implications of a sociocognitive perspective: Assessing writing. In D. Kelly-Riley & N. Elliot (Eds.). *Improving outcomes: Disciplinary writing, local assessment, and the aim of fairness* (pp. 53-66). Modern Language Association.

Mislevy, R.J. (2020). An evidentiary-reasoning perspective on automated scoring: Commentary on Part I. In D. Yan, A.A. Rupp, & P. W. Foltz (Eds.). (2020). *Handbook of automated scoring: Theory into practice* (pp. 151-168). Boca Raton, FL: CRC Press.

Mislevy, R.J., Yan, D., Gobert, J., & Sao Pedro, M. (2020). Automated scoring in intelligent tutoring systems. In D. Yan, A.A. Rupp, & P.W. Foltz (Eds.). (2020). *Handbook of automated scoring: Theory into practice* (pp. 403-422). Boca Raton, FL: CRC Press.

Oliveri, M.E., Mislevy, R.J., & Elliot, N. (2020). After admissions: What comes next in higher education? In C. Wendler & M.E. Oliveri (Eds.). *Higher Education Admission Practices: An International Perspective* (pp. 347-375). Cambridge, UK: Cambridge University Press.

Mislevy, R.J., & Elliot, N. (2020). Ethics, psychometrics, and writing assessment: A conceptual model. In J. Duffy & L. P. Agnew (Eds.). *After Plato: Rhetoric, ethics, and the teaching of writing* (143-162). Logan, UT: Utah State University Press.

Oranje, A., Mislevy, B., Bauer, M. I., & Jackson, G. T. (2019). Summative game-based assessment. In D. Ifenthaler & Y. J. Kim (Eds.). *Game-based assessment revisited* (pp. 37-65). Cham: Springer.

Mislevy, R.J. (2018). On integrating psychometrics and learning analytics in complex assessments. In H. Jiao & Lissitz, R. W. (Eds.). *Test Data Analytics and Psychometrics: Informing Assessment Practices.* Charlotte, NC: Information Age Publishing.

Mislevy, R.J. (2018). On measurement in educational assessment. In C. Secolsky & B. Denison (Eds.). *Measurement, assessment, and evaluation in higher education* (2nd Ed.) (pp. 11-31). Routledge.

Behrens, J.T., Piety, P., DiCerbo, K.E., Mislevy, R.J. (2018). Inferential foundations for learning analytics in the digital ocean. In D. Niemi, R.E. Clark, & B. Saxberg (Eds.), *Learning analytics in education* (pp. 1-48) Charlotte, NC: Information Age Publishing.

Yan, D., & Mislevy, R.J. (2017). ECD and probability-based inference to support Generalized Intelligent Framework for Tutoring (GIFT). In R.A. Sottilare, A.C. Graesser, X. Hu, & G.A. Goodwin (Eds.), *Design recommendations for Intelligent Tutoring Systems: Volume 5* (pp. 101-124). Orlando, FL: US Army Research Laboratory.

Mislevy, R.J. (2017). Resolving the paradox of rich performance tasks. In H. Jiao & Lissitz, R. W. (Eds.). *Test fairness in the new generation of large-scale assessment* (pp. 1-46). Charlotte, NC: Information Age Publishing.

Kane, M.T., & Mislevy, R.J. (2017). Validity evidence based on examinee response processes: meaning and inference. In K. Ercikan & J.W. Pellegrino (Eds.). *Validation of score meaning in the next generation of assessments* (pp. 11-24). Routledge.

Kerr, D., Andrews, J.J., & Mislevy, R.J. (2016). The In-task Assessment Framework for behavioral data. In A. Rupp & J. Leighton (Eds.). *Handbook of Cognition and Assessment* (pp. 472-507). Hoboken, NJ: Wiley-Blackwell.

Bejar, I.I., Mislevy, R.J., Rupp, A.A., & Zhang, M. (2016). Automated scoring with validity in mind. In A. Rupp & J. Leighton (Eds.). *Handbook of Cognition and Assessment* (pp. 226-246). Hoboken, NJ: Wiley-Blackwell.

González-Brenes, J., Mislevy, R.J., Levy, R., Behrens, J.T., & DiCerbo, K.E. (2016). Bayesian inference networks. In A. Rupp & J. Leighton (Eds.). *Handbook of Cognition and Assessment* (pp. 328-353). Hoboken, NJ: Wiley-Blackwell.

Mislevy, R.J. (2016). Missing responses in item response theory. In W. J. van der Linden (Ed.), *Handbook of Modern Item Response Theory, 2nd Edition,* Volume 2 (pp. 171-194). Chapman & Hall/CRC Press.

von Davier, A.A., & Mislevy, R.J. (2016). Design and modeling frameworks for 21st Century simulations & game-based assessments. In C. Wells & M. Faulkner-Bond (Eds.), *Educational measurement: From foundations to the future* (pp. 239-258). NY: Guilford.

DiCerbo, K.E., Mislevy, R.J., & Behrens, J.T. (2016). Inference in game-based assessment. In H. O'Neil, E.L. Baker, & R.S. Perez (Eds.). *Using games and simulations for teaching and assessment: Key issues* (pp. 253-279). Routledge.

Riconscente, M.M., Mislevy, R.J., & Corrigan, S. (2015). Evidence-centered assessment design. In S. Lane, T.M. Haladyna, & M. Raymond (Eds.). *Handbook of Test Development, 2nd Ed* (pp. 40-63). Informa / Taylor & Francis / Routledge.

DiCerbo, K., Bertling, M., Stephenson, S., Jia, Y., Mislevy, R.J., Bauer, M., & Jackson, T. (2015). The role of exploratory data analysis in the development of game-based assessments. In C.S. Loh, Y. Sheng, & D. Ifenthaler (Eds.). *Serious games analytics: Methodologies for performance measurement, assessment, and improvement* (pp. 319-342). Springer.

Mislevy, R.J., Corrigan, S., Oranje, A., DiCerbo, K., Bauer, M.I., von Davier, A.A., John, M. (2015). Psychometrics for game-based assessment. In F. Drasgow (Ed.*), Technology and testing: Improving educational and psychological measurement* (pp. 23-48). Washington, DC: The National Council on Measurement in Educational.

Mislevy, R.J., Behrens, J.T., DiCerbo, K.E., Frezzo, D.C., & West, P. (2012). Three things game designers need to know about assessment. In D. Ifenthaler, D. Eseryel, & X. Ge (Eds.). *Assessment in game-based learning: Foundations, innovations, and perspectives* (pp. 59-81). Springer.

Behrens, J.T., Mislevy, R.J., DiCerbo, K.E., & Levy, R. (2012). An evidence centered design for learning and assessment in the digital world. In M.C. Mayrath, J. Clarke-Midura, & D. Robinson (Eds.). *Technology-based assessments for 21st century skills: Theoretical and practical implications from modern research* (pp. 13-54). Charlotte, NC: Information Age.

Mislevy, R.J. (2012). Modeling language for assessment. In C. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics.* Hoboken, NJ: Wiley-Blackwell. Available online at http://onlinelibrary.wiley.com/doi/10.1002/9781405198431.wbeal0770/full

Levy, R., Mislevy, R.J., & Behrens, J. T. (2011). Markov chain Monte Carlo in educational research. In A. Gelman, G. Jones, X. L. Meng, & S. Brooks (Eds.). *Handbook of Markov chain Monte Carlo: Methods and applications* (pp. 531-546). Boca Raton: Chapman & Hall/CRC Press.

Mislevy, R.J., & Sabatini, J. (2012). How research on reading and research on assessment are transforming reading assessment (or if they aren't, how they ought to). In J. Sabatini & E. R. Albro (Eds.). *Assessing Reading in the 21st Century: Aligning and Applying Advances in the Reading and Measurement Sciences* (pp. 119-134). Lanham, Maryland: R & L Education.

Mislevy, R.J., & Yin, C. (2012). Evidence-centered design in language testing. In G. Fulcher & F. Davidson (Eds.). *Routledge Handbook of Language Testing* (pp. 208-222). London: Routledge.

West, P., Wise Rutstein, D., Mislevy, R.J., Liu, J., Levy, R., DiCerbo, K.E., Crawford, A., Choi, Y., Chapple, K., Behrens, J.T. (2012). A Bayesian network approach to modeling learning progressions. In A.C. Alonzo & A.W. Gotwals (Eds.). *Learning progressions in science* (pp. 255–291). Rotterdam, The Netherlands: Sense Publishers.

Haertel, G., Wentland, E., Yarnall, L., & Mislevy, R., (2011). Evidence centered task design in test development. In C. Secolsky (Ed.) *Measurement, assessment, and evaluation in higher education.* Routledge.

Mislevy, R.J., Bejar, I.I., Bennett, R.E., Haertel, G.D., & Winters, F.I. (2010). Technology supports for assessment design. In B. McGaw, E. Baker, & P. Peterson, (Eds.). *International Encyclopedia of Education,* 3rd Edition, Volume 8 (pp. 56-65). Amsterdam: Elsevier.

Huang, C.-W., & Mislevy, R.J. (2010). An application of the polytomous Rasch model to mixed strategies. In M. Nering & R. Ostini (Eds.). *Handbook of Polytomous Item Response Theory Models* (pp. 213-230). London: Routledge Academic.

von Davier, M., Gonzalez, E., & Mislevy, R.J. (2009). What are plausible values and why are they useful? *IERI Monograph Series, Volume 2* (pp. 9-36). Princeton, NJ: IEA-ETS Research Institute.

Mislevy, R.J. (2009). Validity from the perspective of model-based reasoning. In R.L. Lissitz (Ed.), *The concept of validity: Revisions, new directions and applications* (pp. 83-108). Charlotte, NC: Information Age Publishing.

Mislevy, R.J. (2008). Issues of structure and issues of scale in assessment from a situative/sociocultural perspective. In P. A. Moss, D. Pullin, E. H. Haertel, J. P. Gee, & L. J. Young (Eds.). *Assessment, equity, and opportunity to learn* (pp. 259-294). Cambridge University Press.

Mislevy, R., Chapelle, C. A., Chung, Y.-R., & Xu, J. (2008). Options for adaptivity in computer-assisted language learning and assessment. In C. A. Chapelle, Y.-R. Chung, & J. Xu (Eds.). *Towards adaptive CALL: Natural language processing for diagnostic language assessment* (pp. 9-24). Ames, IA: Iowa State University.

Mislevy, R.J., Levy, R., Kroopnick, M., & Rutstein, D. (2008). Evidentiary foundations of mixture item response theory models. In G.R. Hancock & K.M. Samuelsen (Eds.). *Advances in Latent Variable Mixture Models.*(pp. 149-175). Charlotte, NC: Information Age Publishing.

Mislevy, R.J., Gee, J.P., & Moss, P.A. (2008). On qualitative and quantitative reasoning about assessment validity. In K. Ercikan & W.-M. Roth (Eds.). *Generalizing from educational research: Beyond the quantitative–qualitative opposition* (pp. 67-100). Mahwah, NJ: Erlbaum.

Behrens, J.T., Frezzo, D., Mislevy, R.J., Kroopnick, M., & Wise, D. (2008). Structural, functional and semiotic symmetries in simulation-based games and assessments. In E. Baker, J. Dickieson, W. Wulfeck, & H. O'Neill (Eds.). *Assessment of problem solving using simulations* (pp. 59-80) . Mahwah, NJ: Lawrence Erlbaum Associates.

Mislevy, R.J., & Levy, R. (2007). Bayesian psychometric modeling from an evidence-centered design perspective. In C. R. Rao and S. Sinharay (Eds.). *Handbook of Statistics, Volume 17* (pp. 839-865). North-Holland: Elsevier.

Rupp, A.A., & Mislevy, R.J. (2007). Cognitive foundations of structured item response models. In J.P. Leighton & M. J. Gierl (Eds.). *Cognitive Diagnostic Assessment: Theories and Applications.* Cambridge: Cambridge University Press.

Mislevy, R.J. (2006). Cognitive psychology and educational assessment. In R.L. Brennan (Ed.), *Educational Measurement* (Fourth Edition) (pp. 257-305). Phoenix, AZ: Greenwood.

Bao, H., Gotwals, A.W., Songer, N.B., & Mislevy, R.J. (2006). Using structured item response theory models to analyze content and inquiry reasoning skills in BioKIDS. In X. Liu & W. Boone (Eds.). *Applications of Rasch measurement in science education.* Maple Grove, MN: Journal of Applied Measurement Press.

Mislevy, R.J., & Huang, C-W. (2006). Measurement models as narrative structures. In M. von Davier & C.H. Carstensen (Eds.). *Multivariate and Mixture Distribution Rasch Models.* Springer.

Mislevy, R.J., & Riconscente, M.M. (2006). Evidence-centered assessment design: Layers, concepts, and terminology. In S. Downing & T. Haladyna (Eds.). *Handbook of Test Development* (pp. 61-90). Mahwah, NJ: Erlbaum

Levy, R., Behrens, J.T., & Mislevy, R.J. (2006). Variations in adaptive testing and their online leverage points. In D.D. Williams, S.L. Howell, & M. Hricko (Eds.). *Online assessment, measurement, and evaluation* (pp. 180-202). Hershey, PA: Information Science Publishing.

Hansen, E.G., & Mislevy, R.J. (2006). Accessibility of computer-based testing for individuals with disabilities and English language learners within a validity framework. In M. Hricko (Ed.). *Online assessment and measurement: foundations and challenges* (pp. 212-259). Hershey, PA: Information Science Publishing.

Hendrickson, A.B., & Mislevy, R.J. (2005). Item response theory (IRT): Cognitive Models. In B.S. Everitt & D.C. Howell (Eds.). *Encyclopedia of Statistics in Behavioral Science, Volume 2* (pp. 978–982). Chichester: Wiley.

Braun, H. I., & Mislevy, R. (2005). Intuitive test theory. *Phi Delta Kappan, 86*(7), 488-497.

Mislevy, R.J., Steinberg, L.S., Almond, R.G., Haertel, G., & Penuel, W. (2003). Leverage points for improving educational assessment. In B. Means & G. Haertel (Eds.). *Evaluating the effects of technology in* education (pp. 149–180). Teachers College Press.

Mislevy, R.J., Wilson, M.R., Ercikan, K., & Chudowsky, N. (2003). Psychometric principles in student assessment. In T. Kellaghan & D. Stufflebeam (Eds.). *International Handbook of Educational Evaluation* (pp. 489-531). Dordrecht, the Netherlands: Kluwer Academic Press.

Almond, R.G., Steinberg, L.S., & Mislevy, R.J. (2003). A framework for reusing assessment components. In H. Yanai, A. Okada, K. Shigemasu, Y. Kano, & J.J. Meulman (Eds.). *New developments in psychometrics* (pp. 281-288). Tokyo: Springer.

Frase, L.T., Chudorow, M., Almond, R.G., Burstein, J., Kukich, K., Mislevy, R.J., Steinberg, L.S., & Singley, K. (2003). Technology and assessment. In H.F. O'Neil & R. Perez (Eds.). *Technology applications in assessment: A learning view* (pp. 213-244). Mahwah, NJ: Erlbaum.

Mislevy, R.J., Almond, R.G., & Steinberg, L.S. (2002). On the roles of task model variables in assessment design. In S. Irvine & P. Kyllonen (Eds.). *Generating items for cognitive tests: Theory and practice* (pp. 97-128) Hillsdale, NJ: Erlbaum.

Almond, R.G., Dibello, L., Jenkins, F., Mislevy, R.J., Senturk, D., Steinberg, L.S. and Yan, D. (2001). Jaakkola and Richardson (Eds.). *Models for Conditional Probability Tables in Educational Assessment: Artificial Intelligence and Statistics 2001* (137–143). San Francisco: Morgan Kaufmann.

Williamson, D., Mislevy, R.J., & Almond, R.G. (2000). Model criticism of Bayesian networks with latent variables. In C. Boutilier & M. Goldszmidt (Eds.). *Uncertainty in artificial intelligence 16*, pp. 634-643. San Francisco: Morgan Kaufmann.

Almond, R.G., Herskovits, E., Mislevy, R.J., and Steinberg, L.S. (1999). Transfer of information between system and evidence models. In D. Heckerman & J. Whittaker (Eds.). *Artificial Intelligence and Statistics 99* (pp. 181-186). San Francisco: Morgan Kaufmann.

Mislevy, R.J., Almond, R.G., Yan, D., & Steinberg, L.S. (1999). Bayes nets in educational assessment: Where do the numbers come from? In K.B. Laskey & H. Prade (Eds.). *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence* (437-446). San Francisco: Morgan Kaufmann.

Mislevy, R.J. (1997). Postmodern test theory. In A. Lesgold, M. J. Feuer, & A. M. Black (Eds.). *Transition in work and learning: Implications for assessment*, pp. 180-199. Berkeley, CA: McCutchan.

Mislevy, R.J. (1997). Assessing student learning. In H. J. Walberg & G. D. Haertel (Eds.). *Educational Psychology: Effective Practices and Policies*, pp. 176-195. Berkeley, CA: McCutchan.

Gitomer, D.H., Steinberg, L.S., & Mislevy, R.J. (1995). Diagnostic assessment of trouble-shooting skill in an intelligent tutoring system. In P. Nichols, S. Chipman, & R. Brennan (Eds.). *Cognitively diagnostic assessment* (pp. 73-101). Hillsdale, NJ: Erlbaum.

Mislevy, R.J. (1997). Probability-based inference in cognitive diagnosis. In P. Nichols, S. Chipman, & R. Brennan (Eds.). *Cognitively diagnostic assessment*, pp. 43-71. Hillsdale, NJ: Erlbaum.

Mislevy, R.J. (1993). A framework for studying differences between multiple-choice and free-response test items. In R.E. Bennett & W.C. Ward (Eds.). *Construction versus choice in cognitive measurement* (pp. 75-106). Hillsdale, NJ: Erlbaum.

Mislevy, R.J. (1993). Foundations of a new test theory. In N. Frederiksen, R.J. Mislevy, & I. I. Bejar (Eds.). *Test theory for a new generation of tests*. Hillsdale, NJ: Erlbaum, 1993.

Masters, G.N., & Mislevy, R.J. (1993). New views of student learning: Implications for educational measurement. In N. Frederiksen, R.J. Mislevy, & I. Bejar (Eds.). *Test theory for a new generation of tests* (pp. 219-242). Hillsdale, NJ: Erlbaum.

Mislevy, R.J., Yamamoto, K, & Anacker, S. (1992). Toward a test theory for assessing student understanding. In R.A. Lesh & S. Lamon (Eds.). *Assessments of authentic performance in school mathematics* (pp. 293-318). Washington, DC: American Association for the Advancement of Science.

Mislevy, R.J., & Bock, R.D. (1988). A hierarchical item response model for educational assessment. In R.D. Bock (Ed.), *Multilevel Analysis of Educational Data* (pp. 57-74). Academic Press.

Mislevy, R.J. (1987). Scaling and scoring procedures. In I. S. Kirsch & A. Jungeblut (Eds.). *Literacy: Profiles of America's Young Adults*. Princeton, NJ: National Assessment of Educational Progress.

Selected Technical Reports

Hao, J., & Mislevy, R.J. (2018). Evidence Trace File: A data analytics-oriented structure for capturing evidence from virtual performance assessments. (*Research Report ETS RR-18-28).* Princeton, NJ: Educational Testing Service.

Rahman, T., & Mislevy, R.J. (2016). Integrating cognitive views into psychometric models for reading comprehension assessment. (*Research Report ETS RR-17-35).* Princeton, NJ: Educational Testing Service.

Hao, J., Smith, L., Mislevy, R., von Davier, A., & Bauer, M. (2016). Taming log files from game and simulation based assessment. (*Research Report ETS RR-16-10).* Princeton, NJ: Educational Testing Service.

Hauck, M. C., Wolf, M. K., & Mislevy, R. (2013). *Creating a next-generation system of K–12 English learner (EL) language proficiency assessments.* Princeton: Educational Testing Service. https://www.ets.org/s/research/pdf/24473_K12_EL_Paper.pdf

Mislevy, R.J., Schank, P., Feng, M., Fried, R., Haertel, G.D., & Colker, A.M. (2013). A "Wizard" for authoring scenario-based tasks, using evidence-centered design principles and structures. *(Large-Scale Assessment Technical Report 14).* Menlo Park, CA: SRI International.

Mislevy, R.J. (2013). On the proportion of missing data in classical test theory. *(Research Memorandum ETS RM-13-06).* Princeton, NJ: Educational Testing Service.

Gorin, J.S., & Mislevy, R.J. (2013). *Inherent measurement challenges in the Next Generation Science Standards for both formative and summative assessment.* Princeton, NJ: K-12 Center at ETS.

Vendlinski, T., Haertel, G., Cheng, B., DeBarger, A., Rutstein, D., Fried, R., Snow, E., Zalles, D., Mislevy, R.J., Cho, Y., Fulkerson, D., McCarthy, K., &Finkelstein, D. (2013). Using the Principled Assessment Design in Inquiry (PADI) System: Some frequently asked questions. *(Large-Scale Assessment Technical Report 12).* Menlo Park, CA: SRI International.

Zalles, D., Haertel, G., & Mislevy, R. (2010). Using evidence-centered design to support assessment, design and validation of learning progressions *(Large-Scale Assessment Technical Report 10*). Menlo Park, CA: SRI International.

Mislevy, R. (2012). Four metaphors we need to understand assessment. Commissioned paper for The Gordon Commission on the Future of Assessment in Education. Princeton, NJ: Educational Testing Service. Retrieved January 13, 2013, from http://www.gordoncommission.org/rsc/pdfs/mislevy_four_metaphors_understand_assessment.pdf

Mislevy, R.J. (2011). Evidence-centered design for simulation-based assessment. CSE Technical Report 800. Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA.

Zwick, R. & Mislevy, R.J. (2011). Scaling and linking through-course summative assessments. Princeton, NJ: Center for K-12 Assessment and Performance Management, Educational Testing Service.

Behrens, J.T., Mislevy, R.J., DiCerbo, K.E., & Levy, R. (2010). An evidence centered design for learning and assessment in the digital world. CSE Technical Report 778. Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA.

Zalles, D., Haertel, G., & Mislevy, R. (2010). Using evidence-centered design to support assessment, design and validation of learning progressions *(Large-Scale Assessment Technical Report 10*). Menlo Park, CA: SRI International.

West, P., Wise Rutstein, D., Mislevy, R.J., Liu, J., Levy, R., DiCerbo, K.E., Crawford, A., Choi, Y., Chapple, K., Behrens, J.T. (2010). A Bayesian network approach to modeling learning progressions. CSE Technical Report 776. Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA.

Colker, A. M., Liu, M., Mislevy, R., Haertel, G., Fried, R., & Zalles, D. (2010). A design pattern for experimental investigation (*Large-Scale Assessment Technical Report 8*). Menlo Park, CA: SRI International.

Cheng, B. H., Ructtinger, L., Fujii, R., & Mislevy, R. (2010). Assessing Systems Thinking and Complexity in Science *(Large-Scale Assessment Technical Report 7).* Menlo Park, CA: SRI International.

Liu, M., Mislevy, R., Colker, A. M., Fried, R., & Zalles, D. (2010). A Design Pattern for Experimental Investigation *(Large-Scale Assessment Technical Report 8).* Menlo Park, CA: SRI International.

Zhang, T., Mislevy, R., Haertel, G., Javitz, H., Murray, E., & Gravel, J. (2010). A Design Pattern for a Spelling Assessment for Students with Disabilities *(Assessment for Students with Disabilities Technical Report 2).* Menlo Park, CA: SRI International.

Haertel, G., Haydel DeBarger, A., Cheng, B., Blackorby, J., Javitz, H., Ructtinger, L., Snow, E., Mislevy, R.J., Zhang, T., Murray, E., Gravel, J., Rose, D., Mitman Colker, A., & Hansen, E. G. (2010). Using Evidence-Centered Design and Universal Design for Learning to Design Science Assessment Tasks for Students with Disabilities (*Assessment for Students with Disabilities Technical Report 1*). Menlo Park, CA: SRI International.

Mislevy, R.J., Riconscente, M.M., & Rutstein, D.W. (2009). Design patterns for assessing model-based reasoning *(PADI-Large Systems Technical Report 6).* Menlo Park, CA: SRI International.

Mislevy, R., Liu, M., Cho, Y., Fulkerson, D., Nichols, P., Zalles, D., Fried, R., Haertel, G., Cheng, B., DeBarger, A., Villalba, S., Colker, A., Haynie, K., & Hamel, L., (2009). A Design Pattern for Observational Investigation Assessment Tasks *(PADI-Large Systems Technical Report 2).* Menlo Park, CA: SRI International.

Fulkerson, D., Nichols, P., Haynie, K., Mislevy, R. (2009). Narrative Structures in the Development of Scenario-Based Science Assessments *(PADI-Large Systems Technical Report 3).* Menlo Park, CA: SRI International.

Brecht, J., Cheng, B., Mislevy, R., Haertel, G., & Haynie, K. (2009). *The PADI System as a Complex of Epistemic Forms and Games* (PADI Technical Report 21). Menlo Park, CA: SRI International.

Seeratan, K., & Mislevy, R.J. (2009). Design patterns for assessing internal knowledge representations (*PADI Technical Report 22*). Menlo Park, CA: SRI International.

Mislevy, R.J., & Rahman, T. (2009). A design pattern for assessing cause and effect reasoning in reading comprehension (*PADI Technical Report 20*). Menlo Park, CA: SRI International.

Li, T., & Mislevy, R.J. (2008). Multiple imputation for latent variables in classical test theory for cluster samples. *In JSM Proceedings, Statistical Computing Section.* Alexandria, VA: American Statistical Association.

Hamel, L., Mislevy, R., & Winters, F. (2008). Design rationale for an assessment task authoring system: a wizard for creating "mystery inquiry" assessment tasks *(PADI Technical Report 19).* Menlo Park, CA: SRI International.

Hansen, E. G., Mislevy, R.J., & Steinberg, L. S. (2008). Evidence-centered assessment design for reasoning about accommodations for individuals with disabilities in NAEP reading and math. (*Research Report RR-08-38).* Princeton, NJ: Educational Testing Service.

Hansen, E. G., & Mislevy, R.J. (2008). Design patterns for improving accessibility for test takers with disabilities. (*Research Report RR-08-49).* Princeton, NJ: Educational Testing Service.

Wei, H., Mislevy, R., & Kanal, D. (2008). An introduction to design patterns in language assessment *(PADI Technical Report 18).* Menlo Park, CA: SRI International.

Mislevy, R., & Haertel, G. (2008). Implications of evidence-centered design for educational testing *(Draft PADI Technical Report 17).* Menlo Park, CA: SRI International.

Mislevy, R.J., Behrens, J,T., Bennett, R.E., Demark, S.F., Frezzo, D.C., Levy, R., Robinson, D.H., Rutstein, D.W., Shute, V.J., Stanley, K., Winters, F.I. (2007). On the roles of external knowledge representations in assessment design. CSE Technical Report 722. Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA.

Seibert, G., Hamel, L., Haynie, K., Mislevy, R., & Bao, H. (2006). Mystery powders: An application of the PADI design system using the four-process delivery system. *PADI Technical Report 15.* Menlo Park, CA: SRI International.

Bao, H., Gotwals, A.W., & Mislevy, R. (2006). Assessing local item dependence in building explanation tasks. *PADI Technical Report 14.* Menlo Park, CA: SRI International.

Mislevy, R.J. (2005). Issues of structure and issues of scale in assessment from a situative/sociocultural perspective. *CSE Technical Report 668.* Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA.

Mislevy, R.J., & Riconscente, M. (2005). Evidence-Centered Assessment Design: Layers, structures, and terminology. *PADI Technical Report 9.* SRI International, Menlo Park, CA.

Riconscente, M., Mislevy, R.J., & Hamel, L. (2005). An introduction to PADI task templates. *PADI Technical Report 3.* SRI International, Menlo Park, CA.

Baxter, G. & Mislevy, R.J. (2004). The case for an integrated design framework for assessing science inquiry. *CSE Technical Report 638.* Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA.

Cromley, J.G. & Mislevy, R.J. (2004). Task templates based on misconception research. *CSE Technical Report 646.* Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA.

Mislevy, R.J., Almond, R.G., & Lukas, J. (2004). A brief introduction to evidence-centered design. *CSE Technical Report 632.* Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA.

Braun, H.I., & Mislevy, R.J. (2004). Intuitive test theory. *CSE Technical Report 631.* Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA.

Levy, R., & Mislevy, R.J. (2004). Specifying and refining a measurement model for a simulation-based assessment. *CSE Technical Report 619.* Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA.

Yan, D., Almond, R., & Mislevy, R. A. (2004). A comparison of two models for cognitive diagnosis. (*Research Report RR-04-02).* Princeton, NJ: Educational Testing Service.

Yan, D., Mislevy, R.J., & Almond, R. G. (2003). Design and analysis in a cognitive assessment. (*Research Report RR-03-32).* Princeton, NJ: Educational Testing Service.

Almond, R.G., Steinberg, L.S., & Mislevy, R.J. (2003). Four-process architecture for assessment delivery, with implications for assessment design. *CSE Technical Report 616.* Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA.

Steinberg, L.S., Mislevy, R.J., Almond, R.G., Baird, A.B., Cahallan, C., Dibello, L.V., Senturk, D., Yan, D., Chernick, H., & Kindfield, A. C. H.(2003). Introduction to the Biomass Project: An illustration of evidence-centered assessment design and delivery capability. *CSE Technical Report 609.* Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA.

Mislevy, R.J., et al. (The PADI Research Group). (2003). Design patterns for assessing inquiry in science. PADI Research Report #1. Menlo Park, CA: SRI International.

Mislevy, R.J. (2003). Argument structure and argument substance in educational assessment. *CSE Technical Report 605.* Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA.

Mislevy, R.J. (2003). On the structure of educational assessments. *CSE Technical Report 597.* Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA.

Mislevy, R.J. (2003). Evidentiary relationships among data-gathering methods and reporting scales in surveys of educational achievement. *CSE Technical Report 595.* Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA.

Mislevy, R.J., Almond, R.G., DiBello, L., Jenkins, F., Steinberg, L.S., Yan, D., & Senturk, D. (2002). Modeling conditional probabilities in complex educational assessments. *CSE Technical Report 580.* Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA.

Mislevy, R.J., Wilson, M.R., Erkican, K., & Chudowsky, N. (2002). Psychometric principles in student assessment. *CSE Technical Report 583.* Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA.

Almond, R.G., Steinberg, L.S., & Mislevy, R.J. (2002). Design and analysis in task-based language assessment. *CSE Technical Report 579.* Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA.

Scrams, D. J., Mislevy, R.J., & Sheehan, K. M. (2002). An analysis of similarities in item functioning within antonym and analogy variant families. (*Research Report RR-02-13).* Princeton, NJ: Educational Testing Service.

Sheehan, K. M., & Mislevy, R.J. (2001). An inquiry into the nature of the sentence-completion task: implications for item generation. (*Research Report RR-01-13).* Princeton, NJ: Educational Testing Service.

Almond, R.G., Steinberg, L.S., & Mislevy, R.J. (2001). A sample assessment using the four process framework. *CSE Technical Report 543.* Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA.

Mislevy, R.J., Steinberg, L.S., Breyer, F.J., Johnson,. L., & Almond, R.A. (2001). Making sense of data from complex assessments. *CSE Technical Report 538.* Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA.

Mislevy, R.J., Steinberg, L.S., Almond, R.G., Haertel, G.D., & Penuel, W.R. (2001). Leverage points for improving educational assessment. *CSE Technical Report 534.* Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA.

Mislevy, R.J., Almond, R.G. , Yan, D., & Steinberg, L.S. (2000). Bayes nets in educational assessment: Where do the numbers come from? *CSE Technical Report 518.* Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA.

Mislevy, R.J., Steinberg, L.S., & Almond, R.G. (1999). On the roles of task model variables in assessment design. *CSE Technical Report 500.* Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA.

A sample assessment using the four process framework (with R. Almond & L. Steinberg). White paper for the IMS Working Group on Question and Test Inter-Operability. Princeton, NJ: Educational Testing Service, 1999.

A cognitive task analysis, with implications for designing a simulation-based assessment system (with L. Steinberg, J. Breyer, L. Johnson, & R. Almond). *CSE Technical Report 487.* Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA, 1998.

A note on knowledge-based model construction in educational assessment (with R. Almond & L. Steinberg). *CSE Technical Report 480.* Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, University of California at Los Angeles, 1998.

Graphical models and computerized adaptive testing (with R. Almond). *CSE Technical Report 434.* Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA, 1997.

Graphical models and computerized adaptive testing (with R. Almond). *TOEFL Technical Report #14.* Princeton: Educational Testing Service, 1998.

Measuring cognitive skills (with S. Lazer, K. Whittington, & W.C. Ward). In E.G. Johnson, S. Lazer, & C.Y. O'Sullivan, *NAEP reconfigured: An integrated redesign of the National Assessment of Educational Progress* (4.1-4.52). Princeton, NJ: Educational Testing Service, 1997.

Forsyth, R., Hambleton, R., Linn, R., Mislevy, R., & Yen, W. (1996). *Design/Feasibility Team: Report to the National Assessment Governing Board.* Washington, D.C.: National Assessment Governing Board.

Evidence and Inference in Educational Assessment. CSE Technical Report 414. Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA, 1996.

The role of probability-based inference in an intelligent tutoring system (with D. Gitomer). *CSE Technical Report 413.* Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA, 1996.

Missing responses and Bayesian IRT ability estimation: Omits, choice, time limits, and adaptive testing (with P-K. Wu). (Research Report RR-96-30-ONR). Princeton: Educational Testing Service, 1996.

Myford, C.M., & Mislevy, R.J. (1995). Monitoring and Improving a Portfolio Assessment System. *CSE* Technical *Report 402.* Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA.

Virtual representation of IID observations in Bayesian belief networks. (ETS Research Memorandum 94-13-ONR) Princeton, NJ: Educational Testing Service, 1994.

Information-decay pursuit of dynamic parameters in student models. (ETS Research Memorandum 94-14-ONR) Princeton, NJ: Educational Testing Service, 1994.

Dealing with uncertainty about item parameters: Expected response functions (with M. S. Wingersky & K. M. Sheehan). (ETS Research Report 94-28-ONR) Princeton, NJ: Educational Testing Service, 1994.

Marginal maximum likelihood estimation for a psychometric model of discontinuous development (with M. R. Wilson). (ETS Research Report 92-74-ONR) Princeton, NJ: Educational Testing Service, 1992.

Item-by-form variation in the 1984 and 1986 NAEP reading surveys. In A. E. Beaton & R. Zwick (Eds.). *The effect of changes in the national assessment: Disentangling the NAEP 1985-86 reading anomaly.* Princeton, NJ: Educational Testing Service, 1990.

Inferring examinee ability when some item responses are missing *(with P-K. Wu). (Research Report RR-88-48-ONR). Princeton: Educational Testing Service, 1988*

Marginal estimation procedures (with K. M. Sheehan). In A. E. Beaton (Ed.), *The NAEP 1983-84 Technical Report*. Princeton, NJ: National Assessment of Educational Progress, 1987.

Scale score reporting of national assessment data (with M. R. Reiser & M. Zimowski). Final Report. Chicago: International Educational Services, 1982.

The profile of American youth: Data quality analyses of the Armed Services Vocational Aptitude Battery (with R. D. Bock). Chicago: National Opinion Research Center, 1981.

Reviews

Review of D. Yan, A.A. von Davier, & C. Lewis (Eds.). *Computerized Multistage Testing. Journal of Educational and Behavioral Statistics,* 40(4), 425-431, 2015.

Review of P. Newton, J-A. Baird, H. Goldstein, H. Patrick & P. Tymms (Eds.). *Techniques for monitoring the comparability of examination standards. Assessment in Education,16*(2), 263–269, 2009.

Review of S.E. Embretson & S. Reise's *Item response theory for psychologists. Contemporary Psycholo*gy, 46(6), 629-632, 2001.

Review of R.J. Little & D. B. Rubin's *Statistical analysis with missing data. Journal of Educational Statistics, 16*, 150-155, 1991.

Review of S. A. Raizen & L. V. Jones's (Eds.). *Indicators of precollege education in science and mathematics. Journal of the American Statistical Association, 82*, 680-681, 1987.

SELECTED PRESENTATIONS

Further remarks on evidence and inference in educational assessment. Keynote Address at IMPS2022, the International Meeting of the Psychometric Society. Bologna, Italy (2022).

A Sociocognitive Perspective on Educational Measurement. Keynote Address at the workshop Beyond Results 2021: From log data to valid inferences. Hosted by IEA / DIPF / ZIB, September 30, 2021, Hamburg, Germany.

Advances in the Sciences of Cognition and Measurement. Presented at the Workshop on Educational Assessment as Useful and Usable Evidence, hosted by the National Academy of Education and the American Academy of Political and Social Science, September 14, 2018, University of Pennsylvania.

An Urgent Assessment Question and a Proposed Answer (with an eye toward Bayesian Inference). Keynote address at the International Meeting of the Psychometric Society (IMPS 2018), July 12, 2018, Columbia University, New York.

Assessment Arguments, Process Data, and Validity. Presented in the session "Challenges, Issues and Opportunities in Using Response Process Data in Improving Measurement," at the annual meeting of the National Council on Measurement in Education (NCME), New York, April 16, 2018.

Distributed and Local Assessment Paradigms: Can They Co-Exist in Symbiotic Ways? (with Maria Elena Oliveri and Norbert Elliot). Presented at the annual meeting of the National Council on Measurement in Education (NCME), New York, April 14, 2018.

An Urgent Assessment Question and a Proposed Answer (with an eye toward measurement). Invited plenary address at the International Objective Measurement Workshop (IOMW 2018), April 11, 2018, New York University, New York.

Assessing Diverse Learners (with Frogs!). (with Maria Elena Oliveri, Guadalupe Carmona, and René Lawless). Presented in the vice-presidential session "Culturally and Linguistically Responsive Assessment for Instruction," annual meeting of the National Council on Measurement in Education, San Antonio, TX, April 29, 2017.

Evidence-Centered Design for Research and Development (with Malcolm Bauer). IES Contractors' Conference, Washington D.C., December 16, 2016.

On Integrating Psychometrics and Learning Analytics in Complex Assessments. Invited presentation at the Sixteenth Annual Maryland Conference: Data Analytics and Psychometrics: Informing Assessment Practices, November 3-4, 2016, University of Maryland at College Park.

On Integrating Psychometrics and Learning Analytics in Complex Assessments. Invited presentation at the Insight Analytics & Emerging Technologies Symposium, ACT, Iowa City, November 16-17, 2016.

A Psychometric Framework for New Forms of Assessment. Invited presentation at The Workshop on Computational Models for Learning Systems and Educational Assessment (CMLA 2016), Las Vegas, NV, June 26, 2016.

A Dispatch from the Psychometric Front. Keynote address at Learning Analytics and Knowledge (LAK2016) at the University of Edinburgh, Scotland, UK, April 29, 2016.

An application of evidence-centered-design to assess collaborative problem solving in higher education (with Maria Elena Oliveri and Rene Lawless). Presented at the annual meeting of the National Council on Measurement in Education (NCME), Washington DC, April 11, 2016.

Resolving the Paradox of Rich Performance Tasks, with Implications for Fairness. Invited presentation at the Fifteenth Annual Maryland Conference: Test Fairness in the New Generation of Large-scale Assessment, October 29-30, 2015, University of Maryland at College Park.

The Case for Psychometrics in Game- and Simulation-Based Assessment. Presented at the 2015 CRESST Conference, August 19-20, 2015, Redondo Beach, CA.

How Developments in Psychology and Technology Challenge Assessment. Presentation to the National Academy of Sciences Board on Behavioral, Cognitive, and Sensory Sciences, August 3, 2015, Washington DC.

How Technology and Psychology Challenge Educational Measurement: Statistics Ain't the Half of It. Presented at the Quantitative Methods Seminar, University of Chicago, May 29, 2015, Chicago IL.

Some Implications of Expertise Research for Educational Assessment. Keynote presentation at the Affiliated Group Meeting of the Valid Assessment of Learning Outcomes in Higher Education project (KoKoHs), held in conjunction with the Annual Meeting of the American Educational Research Association, Chicago, IL, April 18, 2015.

Reflections on the Nature of Assessment and Implications for Its Practice. Presented in the Division H Invited Symposium, "Has the next generation of state assessment programs and assessment platforms considered linguistic, social, and cultural inclusivity?" at the Annual Meeting of the American Educational Research Association, Chicago, IL, April 18, 2015.

How I Have Come to Think about Assessment. Plenary address at the Conference on Language, Learning, and Culture, April 10, 2015, Virginia International University, Fairfax VA.

What is Computerized Adaptive Testing? Toward a Taxonomy of CAT. Keynote speech at IACAT 2014, October 8-10, 2014, Princeton NJ.

The Challenge of Game, Learning, and Assessment Integration. (With Kristen DiCerbo, Erin Hoffman, Yue Jia, Geneva Haertel, Malcolm Bauer, Shonte Stephenson, Terry Vendlinski, Michael John, and John Murray.) Presented at Games, Learning, and Society 10, Madison WI, June 10-13, 2014.

Flexibility, Adaptivity, and Measurement: Psychometric Paradigms for Game Design. (With Shonte Stephenson, Andreas Oranje, Maria Bertling, and Yue Jia.) Presented at the annual meeting of the National Council on Measurement in Education (NCME), Philadelphia, PA, April 4, 2014.

Modeling Reading Difficulty of Subject-Matter Assessments. (With Ting Zhang & Xueli Xu.) Presented at the annual meeting of the National Council on Measurement in Education (NCME), Philadelphia, PA, April 5, 2014.

Best Practices in Developing Technology-Enhanced, Scenario-Based Science Assessments. Presented at the annual meeting of the National Council on Measurement in Education (NCME), Philadelphia, PA, April 6, 2014.

Three things game developers need to know about assessment. Presented at the 2014 CRESST Conference, Redondo Beach, CA, April 29-30, 2014.

Evidence-Centered Design (ECD) and game-based assessment: Experience from GlassLab. Presented at the Consortium for Research and Evaluation of Advanced Technologies in Education (CREATE), December 10, 2013, NYU Steinhardt, New York, NY.

The future is already here; it's just very evenly distributed. Invited address at the conference "Testing Then and Now," celebrating the 125th anniversary of Teachers College, December 9, 2013, Columbia University, New York, NY

Bayes nets in educational assessment. Presented at the NCME's Graduate Student Invited Symposium, Annual Meeting of the National Council on Measurement in Education, San Francisco, April 2013.

Three things game designers need to know about assessment. Presented at the Annual Meeting of the National Council on Measurement in Education, San Francisco, April 2013.

A learning progression for psychometrics. Presented at the ETS Research & Development Forum, Washington DC, April 11, 2013.

Some conceptions about evidence-centered assessment design. Buros for Testing / Educational Testing Service, Seventh Annual Lecture. University of Nebraska, Lincoln NE, April 8, 2013.

Evidence-centered design, with an eye toward simulation-based assessment. Visiting Scholar workshop, University of California at Berkeley, March 20, 2013.

Some implications of expertise research for educational assessment. Berkeley Educational Assessment Research lecture, University of California at Berkeley, March 19, 2013.

Validity from the perspective of model-based reasoning. Visiting Scholar workshop, University of California at Berkeley, March 18, 2013.

Conceptions and misconceptions about evidence-centered assessment design. Presented to the Assessment Guild at SRI International, Menlo Park, CA, September 21, 2012.

A learning progression for psychometrics. Presented at the EDMS Measurement and Statistics Monday Seminar, College of Education, University of Maryland, College Park, MD. December 10, 2012.

Evidence Centered Design of complex assessment systems. With J.T. Behrens (presenter) & P. West. Presented at the Annual Meeting of the National Council on Measurement in Education, Vancouver, BC, April 16, 2012.

Bayesian analysis of IRT (Item Response Theory) parameters and amount of information. With T. Rahman (presenter). Presented at the Annual Meeting of the American Educational Research Association, Vancouver, BC, April 17, 2012.

Four metaphors you need to understand assessment. Presented at the National Academy of Education's Adaptive Educational Technologies Project Summit, Washington, D.C., December 1-2, 2011.

Some implications of expertise research for educational assessment. 2011 Anne Anastasi Memorial Lecture, presented at Fordham University, New York, October 25, 2011.

Toward an articulation between educational measurement and a sociocognitive psychological perspective. Invited presentation at the Workshop on Item Response Theory and Educational Measurement, sponsored by the Research Centre for Examination and Certification (RCEC). University of Twente, The Netherlands, October 13-14, 2011.

Scaling and linking through-course summative assessments (with R. Zwick). Presented at the Through-course Summative Assessments Research Symposium, Center for K-12 Assessment and Performance Management, Educational Testing Service, Princeton, NJ, February 11, 2011.

The critical role of design patterns in large-scale assessment. Presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA, April 9, 2011.

Measuring learning progressions using Bayesian modeling in complex assessments. Presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA, April 8, 2011.

What is the construct in task-based language assessment? Presented in the invited colloquium "Reprising the role of tasks in language assessment" organized by John Norris and Steven Ross at the Second Language Research Forum 2010, October 14-17, 2010, University of Maryland, College Park, MD.

Integrating Measurement and Sociocognitive Perspectives in Educational Assessment. Robert L. Linn Distinguished Address. Presented at the Annual Meeting of the American Educational Research Association, Denver, CO, May 1, 2010.

Validity from the Perspective of Model-Based Reasoning. Presented at the PIER Educational Research Speaker Series, Carnegie-Mellon University, Pittsburgh PA, February 16, 2009.

Psychological Foundations. Presented in the symposium "An Integrated Research Program for E-Learning and Assessment in a Complex Domain," at the Annual Meeting of the American Educational Research Association, April 13, 2009, San Diego, CA.

Assessment Arguments and Students with Special Needs (with T. Zhang & E. Hansen). Presented at the Annual Meeting of the American Educational Research Association, April 16, 2009, San Diego, CA.

Design Patterns (with M. Liu). Presented at the Annual Meeting of the American Educational Research Association, April 13, 2009, San Diego, CA.

Some Terminology and Concepts from Evidence-Centered Design. Presented at the meeting of the project ""A 21st Century Assessment Project for Situated and Sociocultural Approaches to Learning," Arizona State University, May 15-16, 2009.

The Intersection of Psychometrics with Simulation Scenarios. Invited address at the Psychometrics of Simulation/Games Workshop, sponsored by the Office of Naval Research, Redondo Beach, CA, July 15-16, 2009.

A Bayes Net Approach to Modeling Learning Progressions and Task Performances. Presented at the Learning Progressions in Science Conference, June 24-26, 2009, University of Iowa, Iowa City, IA.

Psychometric and Evidentiary Approaches to Simulation Assessment in Packet Tracer Software (with D. Frezzo, J. Behrens, P. West, & K. DiCerbo, Cisco Systems). Presented at the Fifth International Conference on Networking and Services (ICNS 2009), April 20-25, 2009, Valencia, Spain (Best Paper Award).

Using Evidence-centered Design in Building States' Large-scale Assessments. Presented at the CCSSO National Conference on Student Assessment, Orlando, Florida, June 16, 2008.

On "Measuring" Proficiency in Cross-Cultural Communication. Presented at the Roundtable Conference on Cross-Cultural Communication in a Globalized World, University of Maryland, College Park, MD, September 24, 2008.

Assessment Arguments and Accelerated Learning. Presented at the Accelerated Learning Workshop, Institute for Defense Analysis, Alexandria, VA, July 22, 2008.

Some Implications of Expertise Research for Educational Assessment. Keynote address at the 34th International Association for Educational Assessment (IAEA) Conference, Cambridge University, September 8, 2008.

Validity from the Perspective of Model-Based Reasoning. Presented at the conference "The Concept of Validity: Revisions, New Directions and Applications," University of Maryland, College Park, MD October 9-10, 2008.

If Language is a Complex Adaptive System, What is Language Assessment? Presented at "Language as a Complex Adaptive System", an invited conference celebrating the 60th Anniversary of Language Learning, at the University of Michigan, Ann Arbor, MI, November 7-9, 2008.

Implications of Expertise Research for Educational Assessment. Inaugural presentation of the Frank B. Womer Lecture Series, School of Education, University of Michigan, November 10, 2008.

Validity from the Perspective of Model-Based Reasoning. Presented at the Survey Research Center, University of Michigan, Ann Arbor, MI, November 12, 2008.

Frontiers in assessment research. Presented at the Assessing Reading In The 21st Century Conference, Philadelphia, PA, April 16-19, 2008

Estimating classification accuracy for educational decisions based on multiple scores (with K. Douglas, presenter). Presented at the annual meeting of the American Educational Research Association, New York, March, 2008.

Some remarks on quantitative vs. qualitative reasoning in educational research. Presented at the Interactive Symposium Session "Generalizing From Educational Research: Beyond the Quantitative–Qualitative Opposition" at the annual meeting of the American Educational Research Association, New York, March, 2008.

A taxonomy of adaptive testing. Keynote address presented at the Fifth Annual Technology for Second Language Learning Conference, September 21-22, 2007, Iowa State University, Ames, Iowa, USA.

Some terminology and concepts for simulation-based assessment. Presented at the Lindquist Center at the University of Iowa, Iowa City, IA, September 19, 2007.

Cognitive diagnosis as evidentiary argument. Invited address at Pearson Educational Measurement, Iowa City, IA. September 20, 2007.

Some terminology and concepts for games-based assessment. Presented at the Gaming and Simulation for Analyst Education Workshop, Sherman Kent Center for Intelligence Analysis, Alexandria, VA, July 23, 2007.

Toward a test theory for the interactionalist era. Samuel J. Messick Memorial Lecture. Presented at the Language Testing Research Colloquium, Barcelona, Spain, June 9, 2007.

Chapter 8: Cognitive Psychology and Educational Assessment. Presented in the symposium on the Fourth Edition of *Education Measurement,* at the annual meeting of the National Council on Measurement in Education, Chicago, IL, April, 2007.

Implications of evidence-centered design for educational testing: Lessons from the PADI project (with G.D. Haertel). Presented at Invited Symposium K3, Assessment Engineering: An Emerging Discipline at the annual meeting of the National Council on Measurement in Education, Chicago, IL, April, 2007.

Leverage Points for Technology in Educational Assessment. Presented in the Symposium Implications of Cognitive and Sociocultural Studies for the Practice of Assessment: A Dialogue across Different Perspectives, at the Annual Meeting of the American Educational Research Association, Chicago, IL, April, 2007.

Design patterns for learning and assessment: Facilitating the introduction of a complex, simulation-based learning environment into a community of instructors (with D. Frezzo & J,. Behrens, presenter). Annual meeting of the American Educational Research Association, Chicago, IL, April, 2007.

Posterior predictive model checking for multidimensionality in item response theory (with R. Levy, presenter, & S. Sinharay). Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, CA, April, 2006.

A Brief Introduction to Evidence-Centered Design, with an eye toward performance tests and simulation-based assessment (with J.T Behrens). Invited presentation at the Association for Performance Testing, September 30, 2006, Washington DC.

Adventures in Pasteur's Quadrant (with J.T Behrens). Presentation at the IERI Contractors Meeting, August 30, 2006, Washington DC.

Prospectus for the PADI design framework in language testing. Presentation at the conference ECOLT 2006 (the East Coast Organization of Language Testers), October 15, 2006, Washington DC.

New Models for Assessment. Invited presentation at Games+Learning+Society 2.0, June 16, 2006, Madison, WI.

Some implications of expertise research for educational assessment. Invited presentation, Conference on Expertise and the Measurement of Competence, University of Fribourg, Switzerland, July 6, 2005.

Cognitive diagnosis as evidentiary argument. Leadoff presentation, Fourth Spearman Conference, Philadelphia, PA, October 15, 2004.

Intuitive test theory. CRESST Conference 2004, UCLA, Los Angeles, CA, September 9, 2004.

Cognitive Psychology and Measurement Models. Keynote address at SMABS 2004. Friedrich Schiller University, Jena, Germany, August 17, 2004.

Test use for specific populations (with E. Hansen). Annual meeting of the American Educational Research Association, San Diego, CA, April 12-16, 2004.

The challenge of context. Presented at the NCME Graduate Students Invited Colloquium. Annual meeting of the National Council on Measurement in Education, Chicago, IL, April 13-15, 2004.

Some Observations on Cognitive Psychology and Educational Assessment. NCME 2003 Career Award Address. Annual meeting of the National Council on Measurement in Education, Chicago, IL, April 13-15, 2004.

An overview of the Principled Assessment Design for Inquiry (PADI) project. Annual meeting of the American Educational Research Association, San Diego, CA, April 12-16, 2004.

A Structural Perspective on Accommodations and Validity Arguments. Validity and Accommodations: Psychometric and Policy Perspectives. University of Maryland, August 4-5, 2003.

Educational Assessments as Evidentiary Arguments: What Has Changed, and What Hasn't. Invitational Conference on Inference, Culture, and Ordinary Thinking in Dispute Resolution. Benjamin N. Cardozo School of Law, New York, April 27-29, 2003.

Design Patterns for Assessing Scientific Inquiry. Annual meeting of the American Educational Research Association, Chicago, IL, April 21-26, 2003.

Object Modeling in Assessment Design and Delivery. Annual meeting of the American Educational Research Association, Chicago, IL, April 21-26, 2003.

Leverage points for improving educational assessment." Annual meeting of the American Educational Research Association, Chicago, IL, April 21-26, 2003.

Specifying and refining a complex measurement model." Annual meeting of the National Council on Measurement in Education, Chicago, IL, April 24-26, 2003.

Empirical comparisons of cognitive diagnosis models." Annual meeting of the National Council on Measurement in Education, Chicago, IL, April 24-26, 2003.

Evidentiary logic in assessment of diverse learners. Annual meeting of the National Council on Measurement in Education, Chicago, IL, April 24-26, 2003.

Cognition and assessment: From theory to practice. Keynote presentation at a conference of the same name, College Park, MD., August 17, 2001.

On the language of assessment. Presented at the Satellite Conference of the International Meeting of the Psychometric Society, Osaka, Japan, July 15, 2001.

Assessing the effects of technology on learning: Implications for assessment instruments. Presented at the CILT2000 conference, Center for Learning Technologies, Maclean, VA, October 26-29, 2000.

The challenge of context. Plenary address at the 2000 CRESST conference, Los Angeles, CA, September 2000.

What is assessment really about, and how must it change? *Featured presentation at the Future of Education conference, Northwestern University, Evanston, IL, May 25-26, 2000.*

Making sense of data from complex assessments. Presentation to the Department of Educational Psychology, Columbia University, New York, NY, March 1, 2000.

Leverage points for improving educational assessment. *Workshop on the Evaluation of Technology in Assessment, sponsored by the U.S. Department of Education, held at SRI International, Menlo Park, CA, February 24-27, 2000.*

Evidentiary relationships among data-gathering methods and reporting scales in surveys of educational achievement. *Invited presentation to the National Academy of Sciences Committee on Marketbasket Reporting for the National Assessment of Educational Progress, Washington, D.C., February 7-8, 2000.*

Making sense of data from complex assessments. Keynote address at the annual conference of the Northeastern Educational Research Association, Ellenville, NY, October 1999.

Design and analysis of complex assessments. Plenary address at the 1999 CRESST conference, Los Angeles, CA, September 1999.

Bayes nets in educational assessment: Where the numbers come from (with R. Almond, D. Yan, & L. Steinberg). Presented at Uncertainty in Artificial Intelligence 99, Stockholm, Sweden, August, 1999.

A cognitive task analysis, with implications for designing a simulation-based assessment system (with L. Steinberg, F.J. Breyer, R. Almond, & L. Johnson). Presented at the annual meeting of the American Educational Research Association, Montreal, Canada, April, 1999.

Evidentiary considerations in performance assessment (with R. Almond & L. Steinberg). Presented at the annual meeting of the National Council of Measurement in Education, Montreal, Canada, April, 1999.

On the roles of task model variables in assessment design (with R. Almond & L. Steinberg). Presented at the conference "Generating items for cognitive tests: Theory and practice", co-sponsored by Educational Testing Service and the United States Air Force Laboratory and held at the Henry Chauncey Conference Center, Educational Testing Service, Princeton, NJ, November, 1998.

Leverage points for improving educational assessment (with R. Almond & L. Steinberg). Presented to the National Academy of Science's Committee on National Statistics, Washington, D.C., October, 1998.

Evidence-centered test design. Plenary address at the 1998 CRESST conference, Los Angeles, CA, September 1998.

Task design, student modeling, and evidentiary reasoning in complex educational assessments. (with R. Almond & L. Steinberg). Poster presentation for the Section on Bayesian Statistical Science at the Annual Meeting of the American Statistical Association, Anaheim, CA, August, 1997.

Using prototype-instance hierarchies to model global dependence (with R. Almond & L. Steinberg). AMS Summer Research Conference on Graphical Markov Models, Influence Diagrams, and Bayesian Belief Networks. July 1997, Seattle, WA.

On the consequences of ignoring certain conditional dependencies in cognitive diagnosis (with R. Patz). Presented at the Annual Meeting of the American Statistical Association, Orlando, FL, August, 1995.

Probability-based inference in cognitive diagnosis. Presented at the Office of Naval Research Contractors' meeting on Teaching and Learning, Evanston, IL, September 1994.

Test theory and language learning assessment. Plenary address to the Center for the Advancement of Language Learning Invitational Conference on Aptitude Measurement, Washington, DC, September 1994.

Evidence and inference in educational assessment. Presidential address to the Psychometric Society, Champaign, IL, June 1994.

Probability-based inference in cognitive diagnosis. Presented at the annual meeting of the Psychometric Society, Champaign, IL, June 1994.

Test theory reconceived. Invited address at the annual meeting of the National Council of Measurement in Education, Atlanta, GA, April 1993.

Policy and technical issues in the national testing program. Presented at the annual meeting of the American Psychological Association, Washington, DC, August 1992.

Scaling procedures in the National Assessment (with E. Johnson). Presented at the annual meeting of the American Statistical Association, Boston, MA, August 1992.

Common themes in problems from cognitive diagnosis and item response theory. Presented at the Office of Naval Research contractors' meeting on Cognitive Diagnosis, Champaign, IL, May 1992.

COMPETITIVELY-FUNDED PROJECTS

Project Director

2009-2010 AutoMentor: Virtual mentoring and assessment in computer games for STEM learning. (with University of Wisconsin). National Science Foundation.

2008-2010 Assessment with games and simulations. MacArthur Foundation, through a subcontract to Arizona State University.

2008-2010 *Toward a synthesis of educational measurement with cognitive, situative, and sociocultural perspectives on learning.* Spencer Foundation. $300K.

2008-2010 *University of Maryland Component of National Technology Center.* U.S. Department of Education.

2007- *Application of evidence-centered-design to states' large-scale science assessment.* National Science Foundation.

2006- *Principled assessment designs for special education.* U.S. Department of Education.

2001-2006 *Principled Assessment Design for Inquiry in Science* (with SRI). National Science Foundation / U.S. Department of Education, IERI grant.

2001-2005 *Valid Assessment for English Language Learners.* U.S. Department of Education.

2001- *Evidence-centered assessment design.* Cisco Learning Institute. $200K per annum.

Prototype for simulation-based assessment. Cisco Learning Institute.

2000-2001 *Planning grant for schema-based assessment in science* (with SRI). National Science Foundation / U.S. Department of Education, IERI grant.

1999-2000 *Foundations of a new test theory—Continuation Award.* U.S. Department of Education.

1998-1999 *Markov chain Monte Carlo estimation.* Educational Testing Service.

1996-1998 *New technology in adaptive testing using collateral information about test items.* Graduate Records Examinations Board.

1996-1997 *Methodological foundations for assessing communicative competence.* Test of English as a Foreign Language.

1996-1997 *Foundations of a new test theory—Continuation Grant.* U.S. Department of Education.

1993-1995 *Two approaches to inference involving choice in assessment.* Program Research Planning Committee of Educational Testing Service.

1993-1994 *Test theory reconceived.* Educational Testing Service.

1993-1994 *Primary school children's attitudes toward science.* Educational Testing Service.

1993-1994 *Preparation of presidential address for the Psychometric Society.* Educational Testing Service.

1993-1994 *Foundations of a new test theory.* U.S. Department of Education.

1992-1994 *Explorations of issues and technical procedures in portfolio analysis.* Program Research Planning Committee of Educational Testing Service.

1992-1993 *Linking educational tests.* Educational Testing Service.

1991-1994 *Diagnosis of cognitive skills and expertise.* Office of Naval Research.

1991-1993 *Foundations of a new test theory.* National Opinion Research Center.

1990-1991 *Equating with little or no data.* Educational Testing Service.

1989-1990 *Statistical foundations of adaptive tests.* Program Research Planning Committee of Educational Testing Service.

1989-1990 *Continuing research—Modeling item responses when different examinees follow different solution strategies.* Program Research Planning Committee of Educational Testing Service.

1988-1989 *A procedure for calibrating "seeded" test items—Modification order.* U.S. Army.

1987-1988 *Modeling item responses when different examinees follow different solution strategies.* Program Research Planning Committee of Educational Testing Service.

A procedure for calibrating "seeded" test items. U.S. Army.

1987-1988 *Dealing with uncertainty in item response theory.* Office of Naval Research.

1986-1987 *Item response theory for multidimensional tests.* Program Research Planning Committee of Educational Testing Service.

Bayesian estimation in item response models. The Spencer Foundation.

1985-1987 *Exploiting collateral information in the estimation of item parameters.* Office of Naval Research.

Key Staff Identification

2012- *Games and assessment innovation laboratory.* MacArthur and Gates Foundations. Senior consultant on assessment.

1985- *The National Assessment of Educational Progress (NAEP).* ETS has been the main contractor for the design and analysis of the National Assessment of Educational Progress (NAEP) since 1983, funded by the US Department of Education, at amounts averaging $4M per year. Dr. Mislevy was identified as a key staff contributor in each contract competition 1985-2000, then again beginning in 2011 when he returned to ETS.

HONORING

Robert J. Mislevy

A collection of remembrances for a life that shaped so many.

June 28, 1950 – May 22, 2025

 https://doi.org/10.37514/JWA-J.2025.8.1.06

A Note of Welcome and Appreciation

It is with deep gratitude that we gather to celebrate the extraordinary life of our friend, colleague, and mentor, Dr. Robert "Bob" Mislevy. Bob's life and loss are profoundly across his family, friends, and the professional community he shaped.

This short booklet of Remembrances is a labor of love, compiled as a joint effort by colleagues from the Study Group and *The Journal of Writing Analytics* and the Study Group. It is designed to be a companion for this gathering at the University of Maryland capturing the essence of Bob's enduring legacy. It draws from two main sources:

- **Written Reminiscences:** Collected by editors of *The Journal of Writing Analytics.*
- **Video Reflections:** Submitted by friends and colleagues for the Study Group tribute video.

The words collected here consistently paint a picture of Bob as a towering scholar whose contributions fundamentally transformed a number of fields related to educational measurement, design, and research. Yet, his intellectual brilliance was matched, and perhaps surpassed, by his boundless humanity. He is consistently remembered for his warmth, humility, generosity, kindness, and patience as a friend and an exceptional mentor who could gently guide colleagues toward profound insights and contributions. He invited everyone into the dialogue, regardless of their background or status.

In preparing this document, the pieces within have been edited for wording, excerpted, trimmed, and often restructured for length to be shared on this occasion. Our hope is that these selected words inspire us all as we reflect on the difference Bob made in our lives. The full, unabridged version of many reminiscences will be published in Volume 8 (2025) of *The Journal of Writing Analytics.*

Thank you for joining us as we honor Bob's memory. His legacy of curiosity, rigor, and grace remains a guiding light for us all.

Sincerely,
Eric Tucker, Sheryl Gómez, Maria Elena Oliveri,
David Slomp, Norbert Elliot

Jessica L. Mislevy, *SRI International*

Any remarks about my dad as a family man would be woefully incomplete without mention of his most important collaboration in life–the one he had with our mother, Robbie, his beloved wife of 50 years. Meredith and I learned so much by the example they set. They provided us with such a strong, positive model for what marriage can be, a relationship to aspire to. They were truly best friends and partners in life, treating each other with love, kindness, and respect. Having married young, my parents grew up with each other and into new life stages. By championing each other's passions and aspirations, they grew stronger together instead of growing apart. They loved to explore new places and hobbies, be it playing volleyball in their 30's or traveling to photograph national parks and wildlife in their 60's. They never stopped wanting to learn from each other, nor laughing together. To see them with one another, you knew you were in the presence of something exceptional. So let me close by saying that we love you, Dad, and can never thank you enough. We celebrate your life and strive to live ours by your example. Today and always.

Russell Almond, *Florida State University*

Bob's impact was from more than his skill as a philosopher and statistician. It came from his willingness to listen carefully and actively to people with different points of view. When Bob didn't understand something, his instinct was to ask questions, and to try and see the point of view of the other person. He was always respectful of his conversation partners, actively listening and never wanting to imply that his viewpoint was more important than theirs.

Eva Baker, *UCLA CRESST*

Bob became valued by our team as a person as much as an expert because of his generosity and kindness. Beyond our work together, I most admired him as a scholar, who with unusual range moved fluidly from theory to framework design, to applications, and back again, He early understood that his facility and expertise across types of programs, settings, different groups of respondents and social contexts contributed enormously to the importance, quality, credibility and utility of his work. All I can add is how grateful I am to have known him.

John T. Behrens, *University of Notre Dame*
Bob was a blessing to anyone who met him. He opened doors to getting things done by solving very difficult problems, either mathematically, or conceptually or operationally. However, not only did he open doors to serve children and adults around the world through assessment experiences, he plumbed the depth of fundamental questions about what it means to think, to reason, to quantify belief, to move from observation to a socially negotiated claim. I never heard him say anything negative about anyone, ever. Bob was gracious to everyone. Decades later he would ask me how many of the people were doing that we worked with: How was the high school teacher that wrote the activities? How was the engineer that conducted the training? How was the software developer who wrote the code? Bob was a great gift to us all who knew him and to many millions of learners whose lives have been improved by the more appropriate knowledge he taught us to accumulate and critically examine.

Randy Bennett, *ETS*
As outstanding and groundbreaking as his many accomplishments were, they were more than matched in my eyes by his humility, kindness, generosity, collegiality and decency. He was a scholar and gentleman in every sense of those words.

Kristen DiCerbo, *Khan Academy*
It isn't those kinds of tangible things that I value most when I look back on our collaboration. It is just his humanity and all of the things that he taught us with his patience and his ability to just continually work through problems until we all got to solutions together. I am so grateful for my time working with Bob and thankful for all that he's taught all of us.

David Dorsey, *Human Resources Research Organization (HumRRO)*
The clarity and originality with which Bob connected ideas was remarkable. He was not trying to be persuasive or forceful; instead, he would make a small observation, pose a thoughtful question, or draw a link between lines of reasoning. It was as if the clouds briefly parted on a mountain top, and wisdom came down in a way only Bob could impart. Bob combined a razor-sharp intellect with a gentle and winsome presence. He was affable, approachable, and deeply willing to listen. You felt invited into a dialogue, one in which the quality of your own ideas rose simply by being in his company. That combination of intellectual rigor and genuine kindness is rare. It is easy to admire someone for their intellect; it is far rarer to be changed by the way their presence frames that intellect. I also remember the way he carried himself: the patience, the openness, and the humility of a person who could have commanded attention yet chose instead to offer encouragement. I will always place him at the top of my list of kind and brilliant souls whose impact will echo for generations to come.

Richard P. Durán, *University of California, Santa Barbara*
Bob shows immense and pervasive sensitivity to the complexity of human embodied learning and pattern ways of knowing and doing and their instantiation in everyday practices lived out in human projections of identities and settings. The importance of this contribution to assessment theory and practice cannot be understated. Rather than narrow the limits of assessment to a quest for finding the ideal or correct theories and models for assessments, it opened up the field of assessment so as to make it an explicit, ongoing dialogue among stakeholders with diverse points of view that could be contested by the scientific method, yet showing openness to the legitimate consideration of ethical and social justice concerns that cannot inherently be reduced to the scientific science alone.

Norbert Elliot, *New Jersey Institute of Technology*
Anyone who has ever worked with Bob will know the pattern: the lightning flashes of insight, the breathtaking writing ability, the time taken to make sure the work has a single voice, no matter how many contributors. Working with Bob was a gift. But, so, too, were the photos and notes on his three grandsons Luke and the twins, Alex and Elliot, with his daughters Jessica and Meredith. There were stories of Christmas Dinners in DC, Luke beaming in a highchair, photos of Elliot with his glasses—looking for all the world like his grandpa on the first day of second grade—and Alex at the piano. "Grandkids are lovely," he wrote to me in the fall of 2024, "and such a joy to behold!"

Howard Everson, *CUNY Graduate Center*
"Blessed is he who plants trees under whose shade he will never sit."
Robert Mislevy left us all many trees to sit under. My students and I will take every opportunity to enjoy the shade.

Michael J. Feuer, *The George Washington University*
Indeed, Bob was one of the most *soulful* psychometricians I've ever known, someone whose generosity of spirit and agility of mind made him a role model for rational discourse even among people who modestly thought they had all the answers (not to mention all the questions). I remember once watching Bob at a National Academy of Sciences meeting and thinking, wow, here was living proof that knowledge was not always inversely correlated with humility.

James Paul Gee, *Arizona State University*

Bob valued honesty, caring, and rigor in his work on assessment. He demanded that we be honest about the fact that our judgments of others are grounded in inferences made under uncertainty. Yet he showed us how to deal rigorously and transparently with that uncertainty. He argued that assessments must be made with sensitivity to the context in which people live and act. We must rigorously and continually interrogate those contexts with understanding, respect and care. At a time when honesty, caring, and intellectual rigor are an ever shorter supply, Bob's work is a beacon toward a better future. Today, it is an ethical imperative that we keep his ideas alive.

Drew Gitomer, *Rutgers University*

First, Bob asked questions that fundamentally altered the trajectory of the field and of researchers all over the world. Second, Bob was as humble a person of greatness as one could ever envision. Not only did that make everyone around him feel valuable, but it also made his own work much stronger. And that was because he took people's ideas and problems seriously, no matter their professional status or training. Whether it was a graduate student, junior researcher, data analyst, test developer, teacher, technician or dental hygienist, Bob would ask questions, value their perspectives, and elevate their contributions for others to see. Finally, Bob was the consummate teacher, one who could shift codes depending on who he was working with.

Edmund W. Gordon, *Teachers College, Columbia University*

Bob was always kind, and he was a very decent human being! On my scale of human decency, no one stands higher than he. Bob stands tall among these giants with whom I was privileged to study and work! Rest well, my friend Bob, you deserve your seat in heaven, at the right hand of the Gods. You have earned this reward! Amen!

Jiangang Hao, *ETS*

What made Bob such an extraordinary mentor was not only his knowledge but also his way of explaining. He did not lecture or impose. Instead, he listened carefully, took my questions seriously, and guided me toward insights that felt like my own discoveries.

John Hattie, *University of Melbourne*
When Fred Lord retired from ETS, this lead left a huge hole. Since 1950, we have been writing footnotes to Lord's work. But along came Bob Mislevy, who took the mantle and raised the bar to new heights, provided fresh directions to measurement, and delivered a scholarly and creative spin to our world. And we have been writing footnotes to Bob's work since. Thanks Bob. You have left another huge hole, but a whole huge contribution.

Saad M. Khan, *ETS*
Bob's jovial demeanor, warmth, and generosity were unforgettable. He readily shared knowledge, offered support, and celebrated the successes of others. This unique blend of intellect, generosity, and cheer made collaborating with him truly enriching and memorable, leaving a lasting legacy of impactful partnership and mentorship.

Neal Kingston, *University of Kansas*
Bob became one of the small handful of deeply knowledgeable psychometricians who was equally impactful with regard to social issues, viewing the test development process as well as the statistical analysis of the results. Bob, you'll be missed. Thank you for having been a part of my life.

Kurt Landgraf, *ETS*
During my Presidency of ETS I always felt that Bob represented the very best of ETS Intellectually and personally. He was a highly accomplished professional but equally as important he was a very good human being.

Ida Lawrence, *ETS*
I also want to touch on his extraordinary openness and mentoring, that I witnessed everyday in the halls of our buildings and in meetings and discussions and seminars and reviews of manuscripts and memos. Bob was a model for how to make colleagues feel comfortable when asking questions and raising issues and concerns. Bob always took the time to explain his thinking and make people feel they could pursue conversations to learn more and do a better job in their work. But what stood out most beyond these attributes was his humility. ETS was immensely fortunate because Bob inspired everyone around him. He did this not through self-promotion, but through quiet excellence, kindness, and generosity of spirit. I am grateful to Bob for the people he encouraged, and the example he set. My thoughts and prayers go out to Bob's family and friends, who I hope find comfort in hearing from me and others today about how much Bob meant to the field and to ETS and to me personally.

Susan Lyons, *Lyons Assessment Consulting*
I am so lucky to have gotten to know you briefly through your work with colleagues developing an evidentiary reasoning lens for socio-culturally responsive assessment. I will continue to be your student and work to honor your commitment to evidence and more importantly, your humility and good humor. You will be missed.

Walt MacDonald, *ETS*
ETS was immensely fortunate to have Bob Mislevy as a leading expert on assessment. His decades-long contributions to the field of educational measurement were many and profound. As the founder of Evidence-Centered Design, he had a sweeping impact on assessment development. He was a brilliant and kind gentleman – he was one of ETS's FINEST! We will dearly miss him.

Maria Elena Oliveri, *Purdue University*
What made working with Bob so meaningful was not just the depth of his ideas, but the spirit with which he shared them. He approached every conversation with humility and care, listening deeply and asking questions that gently but profoundly shifted your thinking. His mentorship was never prescriptive; it was invitational. He encouraged you to ask better questions, to write more clearly, and to think more responsibly about the impact of your work. Our work was grounded in the idea that assessment should serve learning, not constrain it, and that validity is not just a technical property but an ethical responsibility. With Bob, the work was never abstract. It was always connected to people's lives, to real consequences, and to the hope that our tools could contribute, even in small ways, to more just and responsive educational systems. His presence made the work richer, more honest, and more human. Bob changed the way I think, not just about assessment design, but about what it means to do this work with integrity. He modeled a kind of intellectual leadership that was rigorous yet never rigid, always open to new ideas, and grounded in deep respect for learners and their diverse experiences. His insights continue to shape how I build systems, how I frame research questions, and how I mentor others. I carry his influence with me every day.

Bill Penuel, *University of Colorado Boulder*

I came to see Bob, as I know many of his colleagues did, as a kind and generous person and as an intellectual powerhouse capable of putting concrete work and assessment within a larger generative frame. I think Evidence Centered Design has become such a force and framework for the field, in part because Bob was good at articulating its principles and pathways and providing powerful examples of what it can do. It's not too much to say that through his argument based approach, which is deeply rooted in concepts from the philosophy of social science, yet has transformed the field. We are indebted to you who follow imperfectly in your footsteps and have much to do yet to create and build within the landscape you've painted for us. Your memory is already a blessing to us and to your family, and may it be so for generations.

Jeremy Roschelle, *Digital Promise*

He made it so exciting and it was like a grand intellectual game to be part of. And he was so inviting and collaborative and made me feel valued and really part of what he was doing, even though I was not really a big part of it and just made me want to be an assessment person. He really just shaped how I think about so many things from his view of Evidence Centered Design. It's a framework I'll always carry with me.

André A. Rupp, *Center for Assessment*

Bob was deeply intellectual, thoughtful, patient, generous, and compassionate. He profoundly shaped how people across disciplines think about and design assessments, most recently in the context of evidence-centered design and sociocognitive assessment. When I came to the US as a new immigrant, Bob spent a whole Saturday with me to help me find a used car because I needed it urgently for where I lived. Since I did not have any credit history in the US, he offered to co-finance the car with me - even though he did not really know me that deeply. My wife and I still drive this car today. After almost 350,000 miles the "Bobmobile" is still going and I think of him everyday when I am in it.

Valerie J. Shute, *Florida State University*

Whenever someone asks who my heroes are, Bob's name is always one of the first I say. We met in the mid-to-late 1980s — two different minds chasing the same truths, working on intelligent tutoring systems in labs and classrooms, and crossing paths at conferences around the world. I was in awe of him then, and even more so today. Over the decades, we worked at some of the same places — though oddly, never at the same time. For example, I arrived at ETS about a month after Bob left for the University of Maryland (which I had attended back in the early

1970s). Then, when I later left ETS for FSU, Bob returned to ETS. My hope is that these job-hops were merely correlational, not causal! The stealth assessments I embedded in games stand on the shoulders of his quiet genius. And it wasn't just his brilliant mind I admired. It was his kindness. His patience. That rare gift of making everyone around him want to rise. He never forgot that behind every test, every score, was a human being — a story worth telling, *always in context.* Now I find myself missing Bob deeply, searching for words bright enough to honor his legacy, trying not to cry. But maybe it's enough to simply say: Bob shaped not just how I think, but who I strive to be. I was so very lucky to know him. Rest well, my friend. My hero. My colleague. Your light lives on — in every mind you touched, and every heart you inspired. The photo below is from my online retirement celebration in 2021, where Bob said some kind things about me. That meant the world.

David Slomp, *University of Lethbridge*

I took every conversation with Bob, every e-mail I received from him as an opportunity to learn. I know Bob was sort of excited to see the application of these ideas in real case studies and I can imagine that work will be a major focus for me and my colleagues moving forward.

Eric Tucker, *The Study Group*

Bob frequently taught by parable, illustrating deeper insights by holding up teams of game designers, mechanics, and software engineers to reveal an architecture for how we communicate with evidence. He consistently shifted credit to his colleagues, believing the quality of human interaction is inseparable from the quality of the intellectual product. His greatest model was his own life: a demonstration that brilliance and humanity are not conflicting virtues, but mutually reinforcing forces. In a world that rewards quick answers and loud assertions, Bob was patient, curious, and humble. He loved telling human stories because he believed the value of the work lay with the people who did it and the learners it served. He built us up through simple parables, demonstrating that the patient framing of a problem and the generous crediting of collaborators are amongst the most powerful models we have. His career was a quiet refutation of the myth that brilliance must have sharp edges, proving that the most generative intellectual force is the patient cultivation of collective intelligence. Generosity was the operating system for his scholarly and design work. Bob's legacy is not a monument to be admired, but a practice to be inhabited: a challenge to ask the clarifying question, share the credit, and solve the next tough problem together, with a rigorously applied kindness.

Alina A. von Davier, *Duolingo*

At Educational Testing Service (ETS), he advised the Computational Psychometrics Research Center and consistently backed my most novel ideas, giving me both courage and a rigorous scaffold for pursuing them. Bob's voice nudged us back to first principles: what claim are you making about the learner, and what evidence—elicited where, when, and how—justifies that claim? The intellectual lineage from ECD to e-ECD to the DET ecosystem is, to me, the clearest testament to his lasting impact: an elegant argument structure sturdy enough to support new data types, new interactions, and new threats, without losing sight of learners themselves. More personally, Bob's mentorship—his steady encouragement, incisive questions, and unfailing kindness—made room for my own voice. I will miss him dearly, and I will keep building with the tools he taught us to use.

David M. Williamson, *College Board*

Bob served as an official external co-mentor on my dissertation and attended my defense. None of this was part of his job responsibility and was something he simply made time for in addition to his regular work. Later, as he was leaving ETS to join the University of Maryland, he cited his experiences with me as one of his motivations for becoming a professor and in so doing gifted me with one of the greatest professional compliments I ever received. Bob's incredible kindness wasn't restricted to the academic realm. When one of our colleagues had newborn twins, he went to their house to congratulate them. While there he earnestly asked if he could help in any way and ended up spending the afternoon doing several loads of their laundry for them. No one found this surprising.

Duanli Yan, *Rutgers University*

Despite his impressive achievements, he remained remarkably humble and approachable. He had a way of patiently guiding you toward finding answers for yourself, building your confidence step by step. His guidance, encouragement, and vision have shaped my career and my life in ways I could never have imagined. I would not be where I am today without him, and I am eternally grateful for his mentorship and friendship.

Diego Zapata-Rivera, *ETS*

Bob had a remarkable ability to inspire confidence and curiosity in those around him. He listened patiently and asked great questions. He offered excellent suggestions and advice. Such discussions frequently resulted in research projects that have influenced the trajectory of my professional career. Bob will remain as a guiding light and source of inspiration for many of us.

Robert Joseph "Bob" Mislevy, 74, formerly of Lawrenceville, New Jersey, and Wildwood, Illinois, passed away on May 22, 2025, in Severna Park, Maryland. Bob was born June 28, 1950, to Joseph and Ruth Mislevy (née Neisner).

Bob graduated as valedictorian from Warren Township High School in 1968. He obtained a bachelor's degree in mathematics in 1972 from Northern Illinois University, where he met and married his beloved wife, Roberta "Robbie" Mislevy (née Bronson) in 1973. Bob went on to earn a master's degree in mathematics from Northern Illinois University in 1974 and doctoral degree in research methodology from the University of Chicago in 1981.

In 1984, Bob accepted a job with Educational Testing Service and relocated with his family to Lawrenceville, New Jersey. During this time, Bob and Robbie raised two beautiful daughters, Jessica and Meredith, and enjoyed playing in volleyball leagues together.

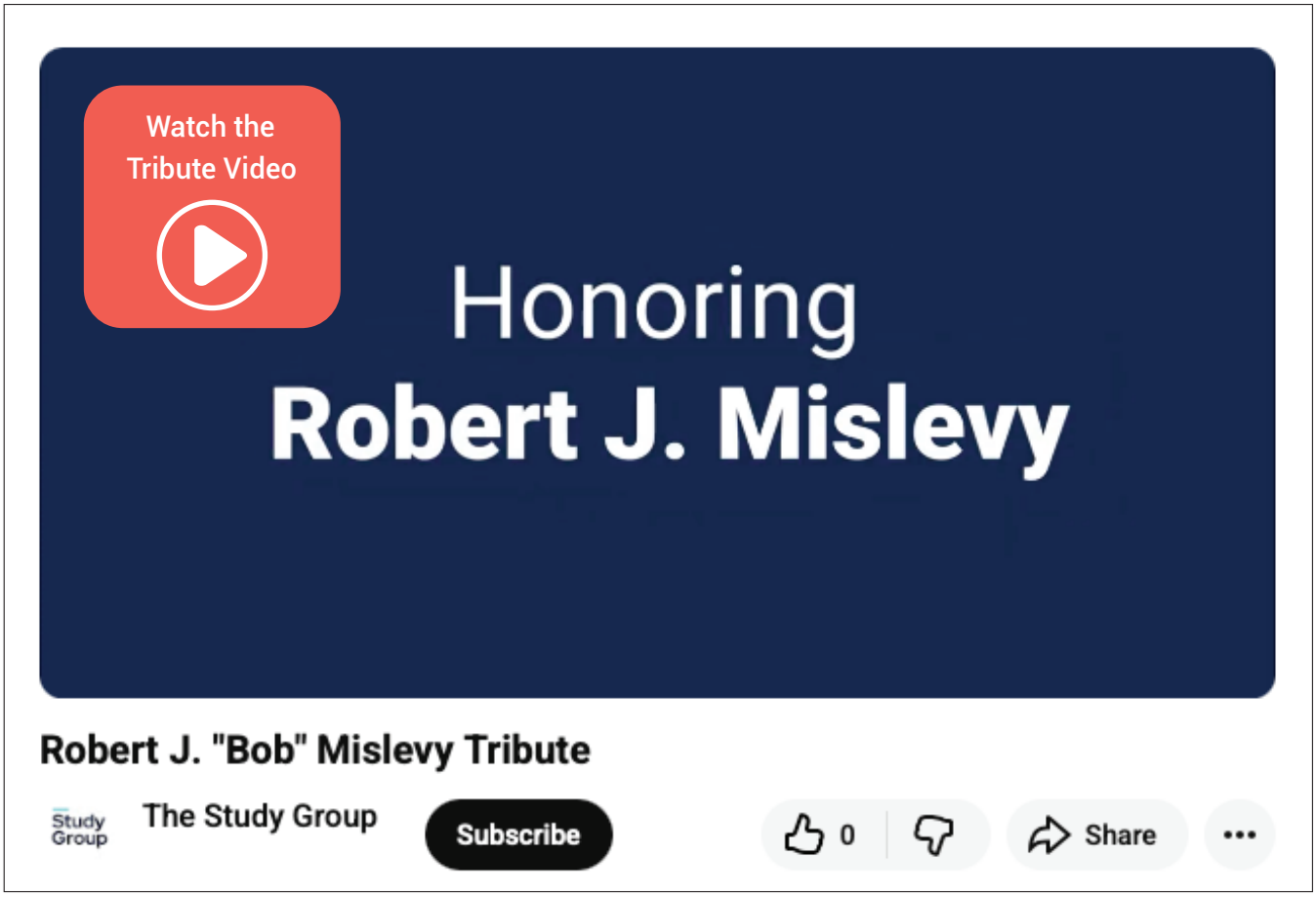
In 2001, Bob joined the faculty at the University of Maryland in the Department of Measurement, Statistics, and Evaluation. There he proudly hooded his daughters when they earned their graduate degrees from the same department. In 2011, he returned to Educational Testing Service as the Frederic M. Lord Chair in Measurement and Statistics, serving in this distinguished role until his retirement in 2021.

Bob's accomplished career in educational measurement included serving as president of the Psychometric Society, serving on several National Academy of Sciences committees on improving learning and assessment, receiving career contribution awards from the AERA and NCME, and earning four times the NCME Award for Technical Contributions to Measurement.

In his later years Bob enjoyed traveling in Europe and to national parks across the United States with wife Robbie and spending much cherished time with his grandsons, while continuing to publish and advise on assessment projects. Throughout his life he was an avid reader, classic film enthusiast, drawing hobbyist, fan of classic rock music, and devoted family man.

Bob is survived by his children Jessica (Rémi) Mislevy and Meredith (Zachary) Hughes and three grandsons Alex and Elliot Hughes and Luke Gottheil; brother Richard (Cynthia) Mislevy and nieces Lea Ann Mislevy (Arjun) and Lorin Mislevy; sister-in-law Debra "Debbie" Bronson; and niece Cassara Bronson and grandnephews Rhyan, Reid, and Remi. Bob was preceded in death by his wife Robbie, his parents, and brother-in-law Robert Jay Bronson.

Honoring Robert J. “Bob” Mislevy Tribute Video



https://www.youtube.com/watch?v=7m1QIH0JiQ0

Study Group