Psychometrika (2025), **90**, 1588–1593
doi:10.1017/psy.2025.10049

PSYCHOMETRIC SOCIETY

THEORY AND METHODS

Obituary: Robert J. Mislevy (1950–2025)

Roy Levy[1] and Russell G. Almond[2]

[1]Arizona State University, USA; [2]Florida State University, USA

Corresponding author: Roy Levy; Email: Roy.Levy@asu.edu

(Received 19 August 2025; revised 28 August 2025; accepted 1 September 2025)

Respecting others in dispute
Opening ears and heart and mind
Because through listening would take root
Extra wisdom he would find

Respect that others gave in turn
To theories drawn from novel routes
Judging that they too could learn
More by listening than disputes

Infer from what they do and say
Student's hidden states of mind
Leading learners so they may
Each their best potential find

Validity from evidence
Yields assessment excellence

—Russell Almond

Robert Joseph Mislevy (known to friends and colleagues as Bob) was born June 28, 1950, and passed away at the age of 74 in Severna Park, Maryland, USA on May 22, 2025, with his daughters by his side.

In his career, Bob had tremendous impact on research, practice, and philosophy. 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.

© The Author(s), 2025. Published by Cambridge University Press on behalf of Psychometric Society.
This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial licence (https://creativecommons.org/licenses/by-nc/4.0), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original article is properly cited. The written permission of Cambridge University Press must be obtained prior to any commercial use.

https://doi.org/10.1017/psy.2025.10049 Published online by Cambridge University Press

Bob had a commensurate impact on those who interacted with him. Whether in conversation with a colleague, working with a student, or deep in discussions with those who thought differently, Bob had a gentle manner and soft-spoken tone, generosity of attention, and above all was unfailingly kind. Bob always saw the best in others and their ideas. If he disagreed with something you said, Bob's assumption was that there was something that you understood differently than he did, which presented an opportunity for doing the kind of work he valued: that which deepened his understanding of things, mapped the border between what was known and unknown, and illuminated paths forward to forge new ground, both for his understanding and for the field at large. His open and inquisitive nature aligned with his view that psychometrics and assessment could gain by learning about and leveraging ideas, developments, and criticisms from other disciplines.

In 1968, Bob graduated as valedictorian from Warren Township High School and attended Northern Illinois University, graduating with a bachelor's degree (1972). It was there that he met his beloved Roberta (Robbie), who he married in 1973. Together, they would go on to raise two daughters, Jessica and Meredith, and later welcome three grandsons.

Bob completed a Master's degree (1974), also in Mathematics, at Northern Illinois University. He took a position at the Institute for Educational Research before attending the University of Chicago, where he completed his PhD in Research Methodology under the supervision of Darell Bock in 1981. After 2 years at the National Opinion Research Center, he joined Educational Testing Service (ETS) in 1984, where he would emerge as a leading scholar.

His early work on parameter estimation included the development of the use of EAP estimators for item response theory (IRT) models (Bock & Mislevy, 1982), and in the production of the BILOG software program (Mislevy & Bock, 1983; Mislevy & Stocking, 1989) for implementing marginal maximum likelihood and Bayes modal estimation (Mislevy, 1986) in IRT. These became widely used techniques and tools in fitting IRT models.

Bob's work helped to popularize Bayesian approaches to facilitate or improve myriad aspects of psychometrics, including parameter estimation, adaptive testing (Bock & Mislevy, 1982), equating (Mislevy et al., 1993), and working at various levels of hierarchically constructed models (Mislevy, 1984, 1986). This hierarchical framing, combined with an understanding of Don Rubin's ideas about multiple imputation as Bayesian in nature, became the key ideas behind the development of plausible values methodologies (Mislevy, Beaton, et al., 1992; Mislevy, Johnson, et al., 1992). Developed for the National Assessment of Educational Progress, these methods would go on to be adopted by other survey assessments (Martin & Mullis, 2019). Such advances reflect what may be characterized as Bob having had an engineering mindset (Wijsen & Borsboom, 2021), where the activities we undertake (e.g., parameter estimation; assessment design, scoring, and reporting) can be framed in terms of our purposes, resources, and constraints. Accordingly, a considerable amount of his work involved the development of conceptual and computational tools for doing this sort of work.

Bob was in many ways a philosopher, eager to discuss all manner of topics: definitions of and distinctions among terms; notions of objectivity, subjectivity, and their implications; (il)legal and (un)ethical (mis)uses of psychometrics and assessment; pathologies in our discipline; the science of managing differential access to information and resolving competing constraints; hermeneutic circles; and the like. Bob valued this not because such topics represented challenges to our field that needed to be rebuffed, but because the field could gain by recognizing the roles these kinds of elements play in shaping the foundations and frontiers of the field. Along the same lines, Bob understood the limitations of assessments, and both the good and the ill that they could be, and have been, used for in society. He was quick to support various points that critics of assessment would make. Nevertheless, he pushed back against criticisms of testing, assessment, or psychometrics (which he was keen to distinguish among) that he found less than persuasive. Importantly, he always did so in a way that sought to illuminate and educate, rather than denigrate.

His philosophical bent and inclination to consider how other disciplines could shape assessment and psychometrics were instrumental in how he came to view assessment as an instance of evidentiary reasoning, essentially always under uncertainty. This marked a confluence of certain influences that

https://doi.org/10.1017/psy.2025.10049 Published online by Cambridge University Press

Bob would point to for the rest of his career as having heavily shaped his thinking. One theme, to a large extent already present if not emphasized as such in his earlier work, was the centrality of Bayesian reasoning to certain inferential problems in assessment. In particular, Bob considered the central inference of assessment to be: reasoning from how people behave or what they say or do to make an inference about their capabilities or dispositions more broadly construed. Frequentist inference could support inference of the kind where we reason that if some parameter θ (e.g., a person parameter in an IRT model) had a particular value, then there is a distribution we would see for observable data x (e.g., scored item responses). But the desired inference in most applied problems runs the other direction. What we wish to be able to do is to say: now, having observed this x, we will act as if this tells us what to think the value of θ might be in the real world and use it as the basis of action. Bob's reading of the probabilist Bruno de Finetti's work on the construction of probability models and Bayesian inference (de Finetti, 1937/1964, 1974) had convinced him that this central inference in assessment was inherently, unavoidably, Bayesian in nature. In Bayesian terms, our model is constructed as $p(x \mid \theta)$. For a given x, inference then amounted to arriving at $p(\theta \mid x)$. This directional reversal was facilitated by Bayesian inference. The underlying Bayesian thinking was present, even if the analyst was not actually using formal Bayesian methods. In this view, even if one was computing a frequentist estimate and standard error, they were interpreted as summaries of posterior belief under a Bayesian perspective.

Bob was convinced that Bayesian inference was the proper mechanism to execute such inferences within probability models, including for models of far greater complexity than familiar unidimensional IRT models. However, he was also convinced by the work of evidentiary scholar David Schum (1987, 1994) that the Bayesian framework is not sufficient for all the reasoning under uncertainty that humans have to do. Our models, probability statements, judgments, and the like are necessarily based on what we know, expect, or theorize. But there are other things that we do not know about. Models encode what things we are willing to entertain as possible, perhaps layered with probabilities expressing beliefs about the relative plausibility of those possibilities. But there is always the possibility that there are relevant things that lie beyond our current understanding of the situation. Different reasoning agents with differential access to relevant information would have different models and inferences, all of which could be framed in terms of evidentiary arguments (Toulmin, 1958).

Early versions of these themes are present in Bob's presidential address to the Psychometric Society (Mislevy, 1994; the title of which was an homage to Schum's (1987) work in the area of military intelligence). While he would continue to work on solving technical problems in psychometrics for the rest of his career, his work took on a new flavor. From this time forward, Bob would emphasize the nature of assessment as being an instance of evidentiary reasoning, and how the work we do in assessment related to, could draw from, or inform other areas that also fell under the larger umbrella of evidentiary reasoning.

In 1995, Bob Mislevy, Linda Steinberg, and Russell Almond started a series of conversations in Bob's office at ETS about a language for describing assessments which could scale from familiar examples like the GRE to novel simulation based assessments such as HyDrive (Mislevy & Gitomer, 1996). It was clear that increases in computing availability was going to allow capturing more information during assessments. Meanwhile the trends in education, which would eventually lead to things like the multidimensional Next Generation Science Standards and the new National Council of Teaching of Mathematics standards, would require assessments that target more than one aspect of proficiency. The goal was to produce a way for describing test design which would allow assessment designers to understand when they needed new methods to meet new challenges and when they could reuse well-studied methods.

Bob's prior work about evidentiary reasoning in psychometrics took on a central role in the new framework, which eventually was called evidence-centered assessment design (ECD; Mislevy et al., 2003). The guiding principle was that assessment designers should first think about what observations about the subjects would provide evidence for or against the presence of the constructs of interest, and then how to structure tasks which would provide opportunities to make those observations (Mislevy et al., 2003). A hallmark of ECD was its call for all aspects of the assessment, including any scoring

https://doi.org/10.1017/psy.2025.10049 Published online by Cambridge University Press

and psychometric models, to be developed in concert with other aspects, including those pertaining to content, form, and physical or digital materials. These all need to cohere, and be aimed at facilitating the desired inferences following the underlying evidentiary argument.

ECD provided a common framework and language for conceiving of assessment activities as parts of evidentiary arguments. Doing so had the benefit of providing grounding for the kinds of inferences supported by conventional assessments and well-developed psychometric models and practices. Importantly, it also supported more sophisticated assessments that called for more complicated psychometric models, often in the form of complex networks, with Bayesian inference as the machinery for propagating information throughout such networks (Almond et al., 2015; Mislevy, 1994, 1995; Mislevy et al., 2003). Bob would go on to apply and extend the ideas of ECD to a whole host of innovative assessment environments, including simulation- and game-based assessments as well as assessments integrated with tutoring systems (Mislevy, 2011, 2013; Mislevy et al., 2014; Mislevy & Gitomer, 1996; Rupp et al., 2010). Relative to conventional assessments, these tended to contain more involved tasks, yielded data of different types and dependencies, and targeted inferences regarding a constellation of constructs. It was this forward looking aspect, and the drawing of connections among fields of probability theory, graphical models, psychology of cognition and performance, and evidentiary argument that led to a joke told in some circles in the early 2000s—that there were four generations of test theory: (1) classical, (2) modern, or IRT, (3) Bayesian, and (4) whatever the hell Mislevy was working on.

Much of this work was done after Bob joined the University of Maryland in 2001 as a Professor in the Department of Measurement, Statistics and Evaluation. There, he taught and mentored students on technical matters of psychometrics, and in his burgeoning interests in the connections among psychometrics, assessment design, evidentiary reasoning, and theories of learning, expertise, instruction, and psychology. He always made time for students, including ones from other schools and disciplines. Bob's students came to learn that his feedback needed to be interpreted in light of his gentle manner. If Bob opened his feedback to a dissertation proposal with "That's a good start," it could more be interpreted as meaning the proposal was actually woefully incomplete. Even an incontrovertibly wrong statement (e.g., 2 + 2 = 5) would elicit from Bob nothing harsher than "That's one way to look at it. Another way to look at it is. . . ."

Bob's work at this time also turned to more explicitly consider the connections and frictions between assessment practice and different psychological theories that could underpin such practices (Mislevy, 2006, 2008). This would be a focal element of much of Bob's scholarship for the rest of his career, including when he returned to ETS in 2011 as the Frederic M. Lord Chair in Measurement & Statistics. His book, *Sociocognitive Foundations of Educational Measurement* (Mislevy, 2018), fleshes out how existing psychometric and assessment concepts and practices, most of which were forged in trait or behaviorist psychology, could be gainfully recast in sociocognitive terms. Like so much of his work, this sought to deepen our understanding of what already is done in the field, and provide a lens for more nuanced assessment practices moving forward. Bob's underlying interest in the intersection between assessment and psychological theories of human learning and behavior shaped much of the rest of his work, including in technical work and in reflections both on our field's history and its future (Von Davier et al., 2021). He continued to publish and advise colleagues on projects, even following his retirement from ETS in 2021. Much of this scholarship emphasized that assessment could be improved by explicitly recognizing the varying social and cultural milieu in which its stakeholders (examinees, designers, users) operate. Furthermore, assessment is not somehow divorced from that milieu, but indeed reflects and contributes to it. Assessment therefore has the opportunity and responsibility to adopt such a stance, such as through socioculturally responsive assessment (Mislevy, 2025; Mislevy et al., 2025). Doing so strengthens assessment not only in terms of evidentiary arguments and conventional psychometric concerns (e.g., validity, reliability, score comparability, Mislevy, 2018), but also in its role as an activity embodying values, morals, and ethics (Mislevy, 2025; Mislevy et al., 2025; Mislevy & Elliot, 2020).

There is hardly a topic in the field that Bob's work did not speak to in some capacity. Throughout his career, he developed novel techniques for doing psychometric work, advanced innovative models, and

https://doi.org/10.1017/psy.2025.10049 Published online by Cambridge University Press

worked across varied settings and contexts. His later scholarship further emphasized the foundations underlying various psychometric and assessment practices, the presumptions and limitations of those foundations, and how they could be built upon moving forward. In all that he did, he was solution oriented and sought to improve practice, and recognized that principled approaches were most fruitful for forging the ways forward. In his interactions with colleagues, students, and friends in and out of the workplace, Bob brought a gentle affect, wide-ranging curiosity, and belief that there was much to be gained by hearing from those with different perspectives. His personality further infused his approach to scholarship. Ever humble about what he and the field had accomplished, he sought to understand what other disciplines had to say, and how those insights could improve assessment and its potential to positively contribute to people's lives. In doing so, he expanded what the field could and should consider as part of its purview, and how it could draw from and contribute to other fields.

Bob left a legacy of scholarship and contributions that will long be remembered and celebrated by those in the field. Those of us fortunate enough to count ourselves as his friends, colleagues, and students will remember and celebrate a legacy no less treasured—one of kindness, generosity, and humility.

References

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

Bock, R. D., & Mislevy, R. J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement*, *6*(4), 431–444. https://doi.org/10.1177/014662168200600405

de Finetti, B. (1937). La prévision: Ses lois logiques, ses sources subjectives. In *Annales de l'Institut Henri Poincaré 7* (pp. 1–68). Translated by Kyburg and Smokler, eds. (1964), Studies in Subjective Probability. New York: Wiley, pp. 93–158.

de Finetti, B. (1974). *Theory of probability* (Vol. *1*). John Wiley & Sons.

Martin, M. O., & Mullis, I. V. S. (2019). TIMSS 2015: Illustrating Advancements in Large-Scale International Assessments. *Journal of Educational and Behavioral Statistics*, *44*(6), 752–781. https://doi.org/10.3102/1076998619882030

Mislevy, R. J. (1984). Estimating latent distributions. *Psychometrika*, *49*(3), 359–381.

Mislevy, R. J. (1986). Bayes modal estimation in item response models. *Psychometrika*, *51*(2), 177–195. https://doi.org/10.1007/BF02293979

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

Mislevy, R. J. (1995). Probability-based inference in cognitive diagnosis. In P. Nichols & R. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 43–71). Erlbaum.

Mislevy, R. J. (2006). Cognitive psychology and educational assessment. In R. Brennan (Ed.), *Educational Measurement* (4th ed., pp. 257–305). Praeger Publishers.

Mislevy, R. J. (2008). How cognitive science challenges the educational measurement tradition. *Measurement: Interdisciplinary Research & Perspective*, *6*, 124. https://doi.org/10.1080/15366360802131635

Mislevy, R. J. (2011). *Evidence-centered design for simulation-based assessment. CRESST Report 800.* (CRESST REPORT 800). http://eric.ed.gov/?id=ED522835

Mislevy, R. J. (2013). Evidence-centered design for simulation-based assessment. *Military Medicine*, *105*, 107–114. https://doi.org/10.7205/MILMED-D-13-00213

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

Mislevy, R. J. (2025). An evidentiary-reasoning perspective on culturally responsive assessment—Commentary on section 2. In C. M. Evans & C. S. Taylor, *Culturally responsive assessment in classrooms and large-scale contexts* (1st ed., pp. 228–242). Routledge. https://doi.org/10.4324/9781003392217-16

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*(2), 133–161. https://doi.org/10.1111/j.1745-3984.1992.tb00371.x

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

Mislevy, R. J., & Elliot, N. (2020). Ethics, psychometrics, and writing assessment: A conceptual model. In J. Duffy & L. Agnew (Eds.), *After Plato: Rhetoric, ethics, and the teaching of writing* (pp. 143–162). Utah State University Press. https://doi.org/10.7330/9781607329978.c008

Mislevy, R. J., & Gitomer, D. H. (1996). The role of probability-based inference in an intelligent tutoring system. *User Modeling and User-Adapted Instruction*, *5*, 253–282. https://doi.org/10.1007/BF01126112

Mislevy, R. J., Johnson, E. G., & Muraki, E. (1992). Scaling procedures in NAEP. *Journal of Educational Statistics*, *17*(2), 131–154. https://doi.org/10.2307/1165166

https://doi.org/10.1017/psy.2025.10049 Published online by Cambridge University Press

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, *Socioculturally responsive assessment* (1st ed., pp. 199–241). Routledge. https://doi.org/10.4324/9781003435105-13

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

Mislevy, R. J., Sheehan, K. M., & Wingersky, M. (1993). How to equate tests with little or no data. *Journal of Educational Measurement*, *30*(1), 55–78.

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., & Stocking, M. L. (1989). A consumer's guide to LOGIST and BILOG. *Applied Psychological Measurement*, *13*(1), 57–75. https://doi.org/10.1177/014662168901300106

Rupp, A. 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), 1–47. http://www.jtla.org.

Schum, D. A. (1987). *Evidence and inference for the intelligence analyst* (2nd ed.). University Press of America.

Schum, D. A. (1994). *The evidential foundations of probabilistic reasoning*. Wiley & Sons.

Toulmin, S. E. (1958). *The uses of argument*. Cambridge University Press.

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

Wijsen, L. D., & Borsboom, D. (2021). Perspectives on psychometrics interviews with 20 past psychometric society presidents. *Psychometrika*, *86*(1), 327–343. https://doi.org/10.1007/s11336-021-09752-7

Cite this article: Levy, R. and Almond, R. G. (2025). Obituary: Robert J. Mislevy (1950–2025). *Psychometrika*, 1588–1593. https://doi.org/10.1017/psy.2025.10049

https://doi.org/10.1017/psy.2025.10049 Published online by Cambridge University Press