ROBERT J. MISLEVY

EDUCATION

University of Chicago	1981	Ph.D.	Research Methodology
Northern Illinois University	1974	M.S. (honors)	Mathematics
Northern Illinois University	1972	B.S. (summa cum laude)	Mathematics, Business

EMPLOYMENT HISTORY

2011-	Frederic M. Lord Chair in Measurement & Statistics, Educational Testing Service, Princeton, NJ
2011-	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

Awards

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).

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).

Visiting Scholar, College Board (2001-2002).

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).

Visiting Scholar, ACT (1995).

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

Offices Held

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

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.

Defense Language Testing Advisory Board.

Idea of Testing (Spencer Foundation).

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 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.

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

National Research Council Committee on Foundations of Assessment.

National Assessment Governing Board's Design/Feasibility Team.

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)

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

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

Monographs

Computerized adaptive testing: A primer. Wainer, H., Dorans, N.J., Flaugher, R., Green, B.F., Mislevy, R.J., Steinberg, L., & Thissen, D. (1990/2000). Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Linking educational assessments: Concepts, issues, methods, and prospects (foreword by R.L. Linn) Mislevy, R.J. (1993). Princeton, NJ: Policy Information Center, Educational Testing Service. (ERIC #: ED-353-302)

Edited volumes

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

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

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

Monographs In Progress

Design and analysis in large-scale assessment. With J. Mazzeo, E. Kulick, and B. S-K Lim. Contract with Springer-Verlag.

Bayes nets in educational assessment. With R.G. Almond, D.M. Williamson, and D. Yan. Contract with Springer-Verlag.

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.

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.

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. (2002). *BILOG-MG II: Multiple group item analysis and test scoring with binary logistic models*. Mooresville, IN: Scientific Software.

Patents

System and method for assessment design. R.J. Mislevy, G.D. Haertel, L.A. Hamel, C.A. Kennedy, M. Wilson (inventors). Patent pending; USPTO Application #20050221266, October 6, 2005.

Accessibility of testing within a validity framework. E.G. Hansen, R.J. Mislevy, L.S. Steinberg (inventors). U.S. Patent # 7217134. May 15, 2007.

Portal assessment design system for educational testing. L.S. Steinberg, R.J. Mislevy, & R.G. Almond (inventors). U.S. Patent #434350000, August 4, 2005.

Research Papers in Refereed Journals

Mislevy, R.J. (in press). Evidence-centered design for simulation-based assessment. *Military Medicine* (special issue on simulation, H. O'Neil, Ed.)

Behrens, J.T., DiCerbo, K., Levy, R., & Mislevy, R.J. (in press). The interplay of design and data exploration in an evolving assessment system. *Journal of Educational Data Mining.*

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). Retrieved [date] from http://www.jtla.org.

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://escholarship.bc.edu/jtla/vol8/2

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 for evidence-centered design for educational assessment. *Educational Measurement: Issues and Practice, 25,* 6-20.

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

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., Steinberg, L.S., & Almond, R.A. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspective*s, 1, 3-67. (focus article for inaugural issue)

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. (conditionally accepted). Evidentiary relationships among data-gathering methods and reporting scales in surveys of educational achievement. *Journal of Educational Measurement*.

Mislevy, R.J., & Chang, H.H. (2000). Does adaptive testing violate local independence? *Psychometrika*.

Cameron, C.A., Beemsterboer, P.L., Johnson, L.A., Mislevy, R.J., Steinberg, L.S., & Breyer, F.J. (1999). 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. (1999). 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.

In Edited Volumes

Mislevy, R.J. Behrens, J.T., DiCerbo, K.E., Frezzo, D.C., & West, P. (in press). 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.* New York: Springer.

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

Behrens, J.T. Mislevy, R,J., DiCerbo, K. E., & Levy, R. (in press). Evidence-centered design for learning and assessment in the digital world. In D. Robinson, J. Clarke-Midura, & M. Mayrath (Eds.), *Technology-based assessments for 21st century skills: Theoretical and practical implications from modern research.* New York, NY: Springer-Verlag.

Mislevy, R.J. (in press). Modeling language for assessment. In C. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics.* Hoboken, NJ: Wiley-Blackwell.

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

Levy, R., Mislevy, R. J., & Behrens, J. T. (in press). 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.* London: Chapman & Hall/CRC Press.

Mislevy, R.J., & Wei, H. (in press). Cognitive diagnosis as evidentiary argument. In R. Roberts & S. Sinharay (Ed.), *Diagnostics for Education: Theory, Measurement, and Applications.* Erlbaum.

Mislevy, R.J., & Sabatini, J. (in press). 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.* Lanham, Maryland: R & L Education

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). New York: 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.* New York: Springer.

Mislevy, R.J.& Levy, R. (2006). Bayesian psychometric modeling from an evidence-centered design perspective. In C. R. Rao and S. Sinharay (Eds.), *Handbook of statistics, Volume 17.* North-Holland: Elsevier.

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.

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). New York: 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 a*nd 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). New York: 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

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.

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). 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., Deniz 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.

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.*

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 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

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. Freidrich 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.
2000-2001	*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.
1987-1988	*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.
1985-1987	*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

1985-2000	*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.

ROBERT J. MISLEVY

Dr. Mislevy was appointed to the Frederic M. Lord Chair in Measurement and Statistics at Educational Testing Service in 2011. He is Professor Emeritus of Measurement, Statistics, and Evaluation at the University of Maryland, where he took the position of Professor in January 2001, as well as Affiliated Professor of Second Language Acquisition and of Survey Methods. Previously, Dr. Mislevy was a Distinguished Research Scientist at ETS, where he had worked since 1984. He earned his Ph.D. in Methodology of Behavioral Research at the University of Chicago in 1981.

Dr. Mislevy's research applies developments in statistics, technology, and cognitive science to practical problems in educational assessment. His work has included a multiple-imputation approach for integrating sampling and test-theoretic models in the National Assessment of Educational Progress (NAEP), a Bayesian inference network for updating the student model in an intelligent tutoring system, and an "evidence centered" framework for assessment design.

In 1988, the American Educational Research Association presented Dr. Mislevy with the Raymond B. Cattell Early Career Award for Programmatic Research. He has won the National Council of Measurement in Education's Triennial Award for Technical Contributions to Educational Measurement three times. In 1992, he was elected president of the Psychometric Society. In 1993, he was awarded the ETS Senior Research Scientist Award. In 2002, he received the National Council of Measurement's Award for Career Contributions. In 2007, he received the Messick Lecture Award from the International Language Testing Association and the E.F. Lindquist Award from the American Educational Research Association for contributions to educational assessment. He was elected to the National Academy of Education in 2007. He has served as a member on several committees of the National Academy of Sciences and the Spencer Foundation concerning assessment, instruction, and cognitive psychology, and was a primary author of final report of the National Assessment Governing Board's Design Feasibility Team.

Selected publications:

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://escholarship.bc.edu/jtla/vol8/2

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. (2006). Cognitive psychology and educational assessment. In R.L. Brennan (Ed.), *Educational Measurement* (Fourth Edition) (pp. 257-306). Phoenix, AZ: Greenwood.

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

Mislevy, R.J., Steinberg, L.S., & Almond, R.A. (2002). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspective*s, 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., & 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. (1994). Evidence and inference in educational assessment. *Psychometrika, 59,* 439-483.

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

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