

## 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

2001-	Professor, Department of Measurement, Statistics, and Evaluation; Affiliate Professor, Joint Program in Survey Methodology (2007- ), Affiliate Professor, Second Language Acquisition (2004- ), 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 (Psychometric methods for sample surveys)
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

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 (ACT) (2007)

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

*E-Learn 2003 Outstanding Paper Award* 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)

## **Offices Held**

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**

Technical Panel on 12th Grade Preparedness Research, National Assessment  
Governing Board  
Task Force on CAT 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  
Measurement: Interdisciplinary Research and Perspectives  
Journal of Learning, Technology, and Assessment  
Psychometrika, Associate Editor (1995), Guest Associate Editor (2004)

## **Affiliations**

American Educational Research Association (program chair for Division D;  
Lindquist Award Committee member)  
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**

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**

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.*

## **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

- Frezzo, D.C., Behrens, J.T., & Mislevy, R.J. (in press). 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*.
- Levy, R., Mislevy, R.J., & Sinharay, S. (in press). Posterior predictive model checking for multidimensionality in item response theory. *Applied Psychological Measurement*.
- Mislevy, R.J. (in press). How cognitive science challenges the educational measurement tradition. *Measurement: Interdisciplinary Research and Perspectives*.
- 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 Measurement*, 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 Measurement*, 4, 303-332.
- Levy, R., & Mislevy, R.J. (2004). Specifying and refining a measurement model for a simulation-based assessment. *International Journal of Measurement*, 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 Perspectives*, 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 Assessment*, 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., Gee, J.P., & Moss, P.A. (in press). On qualitative and quantitative reasoning about assessment validity. In K. Ercikan & W.-M. Roth (Eds.), *Generalizing from educational research: Beyond the quantitative–qualitative opposition*. Mahwah, NJ: Erlbaum.
- Mislevy, R.J., Bejar, I.I., Bennett, R.E., Haertel, G.D., & Winters, F.I. (in press). Technology supports for assessment design. In B. McGaw, E. Baker, & P. Peterson, (Eds.), *International Encyclopedia of Education*, 3rd Edition. Amsterdam: Elsevier.
- Mislevy, R.J. (in press). 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.), *Opportunity to learn*. New York: Cambridge University Press.
- Mislevy, R.J., Levy, R., Kroopnick, M., & Rutstein, D. (in press). Evidentiary foundations of mixture item response theory models. In G.R. Hancock & K.M. Samuelsen (Eds.), *Advances in Latent Variable Mixture Models*. Information Age Publishing.
- 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.
- Bao, H., Gotwals, A.W., Songer, N.B., & Mislevy, R.J. (in press). 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.
- Behrens, J.T., Frezzo, D., Mislevy, R.J., Kroopnick, M., & Wise, D. (in press). Structural, functional and semiotic symmetries in simulation-based games and assessments. In H. O'Neill (Ed.), *Assessment of problem solving using simulations*. Mahwah, NJ: Lawrence Erlbaum Associates.
- 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. *Educational Measurement* (Fourth Edition). Phoenix, AZ: Greenwood.
- 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 and Statistics 2001* (137–143). San Francisco: Morgan Kaufmann.
- Williamson, D., Mislevy, R.J., & Almond, R.G. (2000). Model criticism of Bayesian networks with latent variables. In C. Boutilier & M. Goldszmidt (Eds.), *Uncertainty in artificial intelligence 16*, pp. 634-643. San Francisco: Morgan Kaufmann.
- Almond, R.G., Herskovits, E., Mislevy, R.J., and Steinberg, L.S. (1999). Transfer of information between system and evidence models. In D. Heckerman & J. Whittaker (Eds.), *Artificial Intelligence and Statistics 99* (pp. 181-186). San Francisco: Morgan Kaufmann.

- Mislevy, R.J., Almond, R.G., Yan, D., & Steinberg, L.S. (1999). Bayes nets in educational assessment: Where do the numbers come from? In K.B. Laskey & H. Prade (Eds.), *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence* (437-446). San Francisco: Morgan Kaufmann.
- Mislevy, R.J. (1997). Postmodern test theory. In A. Lesgold, M. J. Feuer, & A. M. Black (Eds.), *Transition in work and learning: Implications for assessment*, pp. 180-199. Berkeley, CA: McCutchan.
- Mislevy, R.J. (1997). Assessing student learning. In H. J. Walberg & G. D. Haertel (Eds.), *Educational Psychology: Effective Practices and Policies*, pp. 176-195. Berkeley, CA: McCutchan.
- Gitomer, D.H., Steinberg, L.S., & Mislevy, R.J. (1995). Diagnostic assessment of trouble-shooting skill in an intelligent tutoring system. In P. Nichols, S. Chipman, & R. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 73-101). Hillsdale, NJ: Erlbaum.
- Mislevy, R.J. (1997). Probability-based inference in cognitive diagnosis. In P. Nichols, S. Chipman, & R. Brennan (Eds.), *Cognitively diagnostic assessment*, pp. 43-71. Hillsdale, NJ: Erlbaum.
- Mislevy, R.J. (1993). A framework for studying differences between multiple-choice and free-response test items. In R.E. Bennett & W.C. Ward (Eds.), *Construction versus choice in cognitive measurement* (pp. 75-106). Hillsdale, NJ: Erlbaum.
- Mislevy, R.J. (1993). Foundations of a new test theory. In N. Frederiksen, R. J. Mislevy, & I. I. Bejar (Eds.), *Test theory for a new generation of tests*. Hillsdale, NJ: Erlbaum, 1993.
- Masters, G.N., & Mislevy, R.J. (1993). New views of student learning: Implications for educational measurement. In N. Frederiksen, R.J. Mislevy, & I. Bejar (Eds.), *Test theory for a new generation of tests* (pp. 219-242). Hillsdale, NJ: Erlbaum.
- Mislevy, R.J., Yamamoto, K., & Anacker, S. (1992). Toward a test theory for assessing student understanding. In R.A. Lesh & S. Lamon (Eds.), *Assessments of authentic performance in school mathematics* (pp. 293-318). Washington, DC: American Association for the Advancement of Science.
- Mislevy, R.J., & Bock, R.D. (1988). A hierarchical item response model for educational assessment. In R.D. Bock (Ed.), *Multilevel Analysis of Educational Data* (pp. 57-74). 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**

- 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.
- 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. (2006). 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 S.E. Embretson & S. Reise's *Item response theory for psychologists*. *Contemporary Psychology*, in press.
- 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

- A taxonomy of adaptive testing.* 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.
- 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.
- Cognitive diagnosis as evidentiary argument.* Invited address at Pearson Educational Measurement, Iowa City, IA. September 20, 2007.
- Toward a test theory for the interactionist 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

- 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.
- 1986-1987 *Item response theory for multidimensional tests.* Program Research Planning Committee of Educational Testing Service.
- 1987-1988 *A procedure for calibrating “seeded” test items.* 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 *Dealing with uncertainty in item response theory.* Office of Naval Research.
- 1988-1989 *A procedure for calibrating “seeded” test items—Modification order.* U.S. Army.
- 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.
- 1990-1991 *Equating with little or no data.* Educational Testing Service.
- 1992-1993 *Linking educational tests.* Educational Testing Service.
- 1991-1993 *Foundations of a new test theory.* National Opinion Research Center.
- 1993-1994 *Test theory reconceived.* Educational Testing Service.
- 1993-1994 *Preparation of presidential address for the Psychometric Society.* Educational Testing Service.
- 1993-1994 *Primary school children’s attitudes toward science.* Educational Testing Service.
- 1991-1994 *Diagnosis of cognitive skills and expertise.* Office of Naval Research.
- 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.

- 1993-1995 *Two approaches to inference involving choice in assessment.* Program Research Planning Committee of Educational Testing Service.
- 1996-1997 *Foundations of a new test theory—Continuation Grant.* U.S. Department of Education.
- 1996-1997 *Methodological foundations for assessing communicative competence.* Test of English as a Foreign Language.
- 1996-1998 *New technology in adaptive testing using collateral information about test items.* Graduate Records Examinations Board.
- 1998-1999 *Markov chain Monte Carlo estimation.* Educational Testing Service.
- 1999-2000 *Foundations of a new test theory—Continuation Award.* U.S. Department of Education.
- 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.
- 2001-2005 *Valid Assessment for English Language Learners.* 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- *Evidence-centered assessment design.* Cisco Learning Institute.
- 2006- *Principled assessment designs for special education.* U.S. Department of Education.
- 2007- *Application of evidence-centered-design to states' large-scale science assessment.* National Science Foundation.

### **Key Staff Identification**

- 1985- *The National Assessment of Educational Progress (NAEP).* ETS has been the main contractor for the design and analysis of the National Assessment of Educational Progress (NAEP) since 1983, funded by the US Department of Education, at amounts averaging \$4M per year. Dr. Mislevy has been identified as a key staff contributor in each contract competition since 1986.

## ROBERT J. MISLEVY

Dr. Mislevy joined the Department of Measurement, Statistics, and Evaluation at the University of Maryland as a Professor in January 2001. In 2004, he also became an Affiliated Professor of Second Language Acquisition. Previously, Dr. Mislevy was a Distinguished Research Scientist at Educational Testing Service, 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 interests are applying developments in statistic, technology, and cognition to practical problems in educational and psychological measurement. 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. (2006). Cognitive psychology and educational assessment. In R.L. Brennan (Ed.), *Educational Measurement* (Fourth Edition) (pp. 257-306). Phoenix, AZ: Greenwood.
- 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.
- 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 Perspectives*, 1, 3-67.
- Mislevy, R.J., Steinberg, L.S., Breyer, F.J., Johnson, L., & Almond, R.A. (2002). Making sense of data from complex assessments. *Applied Measurement in Education*, 15, 363-378.
- Mislevy, R.J., & 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. (1993). *Linking educational assessments: Concepts, issues, methods, and prospects*. Princeton, NJ: Policy Information Center, Educational Testing Service.
- 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.