

Further Readings about the ECD Project

(prepared in 2003)

The following is an annotated list of publications that have been produced in the ECD research program. They are classified into three groups: publications about the ECD framework itself, applications of the ideas, and particular aspects of assessment design and analysis from the perspective of evidentiary reasoning.

The ECD Framework

- 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> Also available as *CSE Technical Report 543*. Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA. <http://www.cse.ucla.edu/CRESST/Reports/TECH543.pdf>. [Extended discussion of the four-process delivery system architecture, including explanation of relationships between the design objects of the conceptual assessment framework and the processes and messages in an assessment delivery system.]
- 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. 28-288). Tokyo: Springer. [Shorter description of the four-process delivery system, with descriptions of what the four processes do and how they interact in assessments designed to achieve different purposes.]
- 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. [This article provides an overview of developments in the use of technology in assessment. One of these is a section on the evidence-centered design system.]
- Mislevy, R.J., Almond, R.G., & Lukas, J. (in press). A brief introduction to evidence-centered design. *CSE Technical Report*. Los Angeles: The National Center for

Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA. [Readable introduction to the basic ideas and structures of evidence-centered design. Contains an annotated bibliography.]

Mislevy, R.J., Steinberg, L.S., & Almond, R.G. (2002). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, 1, 3-67. Forthcoming as CSE Research Report. [Currently the most comprehensive overview available of evidence centered design, spanning assessment arguments, to design elements, to delivery system architecture, and the connections within and across these levels.]

Applications

Bauer, M., Williamson, D.M., Steinberg, L.S., Mislevy, R.J., & Behrens, J.T. (April, 2001). *How to create complex measurement models: A case study of principled assessment design*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA. [ECD design rationale for a simulation-based assessment of troubleshooting and design of computer networks. Foundational analysis for the NetPASS on-line assessment of networking skill, by the Cisco Learning Institute, Educational Testing Service, and the University of Maryland. Includes expert-novice analysis of problem-solving.]

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. [Expert-Novice study of expertise in problem-solving in dental hygiene, with implications for assessment design.]

Levy, R., & Mislevy, R.J. (2003). 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. <http://www.cse.ucla.edu/reports/R619.pdf> [Focus on estimation of conditional probability models in the Bayes net psychometric model in the Cisco Learning Institute's NetPASS prototype assessment. A fairly technical psychometric paper.]

Mislevy, R.J., Almond, R.G., Dibello, L.V., 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. <http://www.cse.ucla.edu/CRESST/Reports/TR580.pdf>. [Focus on

estimation of conditional probability models in the Bayes net psychometric model in the Biomass prototype assessment. A fairly technical psychometric paper.]

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. Also available as *CSE Technical Report 413*. Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA.

<http://www.cse.ucla.edu/CRESST/Reports/TECH413.PDF> [Good foundational explanation of the use of Bayesian inference in complex assessments, illustrated with the HYDRIVE intelligent tutoring system for troubleshooting aircraft hydraulics.]

Mislevy, R.J., Steinberg, L.S., & Almond, R.A. (2002). Design and analysis in task-based language assessment. *Language Assessment*, 19, 477-496. Also available as *CSE Technical Report 579*. Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA. <http://www.cse.ucla.edu/CRESST/Reports/TR579.pdf>. [ECD perspective on designing task-based language assessments. Includes examples of Bayes nets for tasks that tap multiple aspects of knowledge and skill.]

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. Also available as *CSE Technical Report 487*. Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA. <http://www.cse.ucla.edu/CRESST/Reports/TR487.pdf>. [Design and conduct of a cognitive task analysis of expertise in dental hygiene, from the perspective of informing the construction of the models in the ECD conceptual assessment framework.]

Mislevy, R.J., Steinberg, L.S., Breyer, F.J., Almond, R.G., & Johnson, L. (2002). Making sense of data from complex assessment. *Applied Measurement in Education*, 15, 363-378. Also available as *CSE Technical Report 538*. Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA.

<http://www.cse.ucla.edu/CRESST/Reports/RML%20TR%20538.pdf>. [Argument that the way to design and analyze complex assessments, such as computer-based simulations, is from the perspective of the evidentiary argument--not from the perspective of technology. Ideas are illustrated in some detail with the DISC prototype assessment of problem-solving in dental hygiene.]

Steinberg, L.S., & Gitomer, D.G. (1996). Intelligent tutoring and assessment built on an understanding of a technical problem-solving task. *Instructional Science*, 24, 223-258. [Concerns the interplay among cognitive analysis, instructional strategy, and assessment design, in the context of the HYDRIVE intelligent tutoring system for troubleshooting aircraft hydraulics.]

Steinberg, L.S., Mislevy, R.J., Almond, R.G., Baird, A.B., Cahallan, C., DiBello, L.V., Senturk, D., Yan, D., Chernick, H., & Kindfield, A.C.H. (2003). *Introduction to the Biomass project: An illustration of evidence-centered assessment design and delivery capability*. CSE Technical Report #609. Los Angeles, CA: UCLA Center for the Study of Evaluation. <http://www.cse.ucla.edu/reports/R609.pdf> [Design rationale for a standards-based, web-delivered assessment of science inquiry, in the areas of transmission genetics and microevolution. Much discussion of working with experts and National Science Education Standards, to carry out the ECD design work and then implement a prototype assessment at the level of secondary science.]

Williamson, D. M., Bauer, M., Mislevy, R. J., Behrens, J. T. (April, 2003). *An ECD Approach to Designing for Reusability in Innovative Assessment*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL

Williamson, D. M., Bauer, M., Steinberg, L. S., Mislevy, R. J., Behrens, J. T. (April, 2003). *Creating a Complex Measurement Model Using Evidence Centered Design*. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL

Aspects of Assessment Design and Analysis

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. Also available as *CSE Technical Report 480*. Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA.

<http://www.cse.ucla.edu/CRESST/Reports/TECH480.pdf>. [Concerns the technical issue of maintaining student-model and measurement-model fragments of Bayes nets, to be assembled dynamically as is required in adaptive assessments.]

Almond, R.G., & Mislevy, R.J. (1999). Graphical models and computerized adaptive testing. *Applied Psychological Measurement*, 23, 223-237. Also available as *CSE Technical Report 434*. Los Angeles: The National Center for Research on

Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA. <http://www.cse.ucla.edu/CRESST/Reports/TECH434.PDF>. [Early discussion of the kinds of variables that arise in language assessment, the roles they play in the assessment argument, and where they fit in with Bayes net modeling of performance.]

Gitomer, D.H., & Steinberg, L.S. (1999). Representational issues in assessment design. In I.E. Sigel (Ed.), *Development of mental representation* (pp. 351-370). Hillsdale, NJ: Erlbaum. [Discussion of the key role of representational forms in assessment. Addresses both the use of representational forms to provide information and elicit responses from examinees, and the role of assessments as representations themselves as to what is important in a domain and how it is evaluated.]

Mislevy, R.J. (2003). Substance and structure in assessment arguments. *Law, Probability, and Risk*, 2, 237-258. [Analysis of assessment arguments through the lens of Toulmin's schema, comparing assessment cast under the trait, behavioral, information-processing, and sociocultural psychological perspectives.]

Mislevy, R.J. (1994). Evidence and inference in educational assessment. *Psychometrika*, 59, 439-483. Also available as *CSE Technical Report 414*. Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA. <http://www.cse.ucla.edu/CRESST/Reports/TECH414.PDF>. [Foundational, not overly technical, discussion of the role that probability-based reasoning plays in assessment and assessment design.]

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. Also available as *CSE Technical Report 518*. Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA. <http://www.cse.ucla.edu/CRESST/Reports/TECH518.pdf>. [Discussion of Markov Chain Monte Carlo estimation in a binary skills multivariate latent class model for cognitive diagnosis. Illustrated with analysis of data from Kikumi Tatsuoka's studies of mixed number subtraction.]

Mislevy, R.J., & Patz, R.J. (1995). On the consequences of ignoring certain conditional dependencies in cognitive diagnosis. *Proceedings of the Section on Bayesian Statistical Science: Papers presented at the Annual Meeting of the American Statistical Association*, Orlando, FL, August 13-17, 1995 (pp. 157-162). [Technical

paper on the implications of simplifications of Bayes net structures in assessment for computing advantage. Conclusion: Ignoring dependencies among student-model variables is generally conservative, but ignoring conditional dependencies among observations can lead to over-counting evidence.]

Mislevy, R.J., Steinberg, L.S., & Almond, R.G. (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. Also available as *CSE Technical Report 500*. Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA. <http://www.cse.ucla.edu/CRESST/Reports/TECH500.pdf>.

Mislevy, R.J., Steinberg, L.S., Almond, R.G., Haertel, G., & Penuel, W. (in press). Leverage points for improving educational assessment. In B. Means & G. Haertel (Eds.), *Evaluating the effects of technology in education*. New York: Teachers College Press. Also available as *CSE Technical Report 534*. Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA. <http://www.cse.ucla.edu/CRESST/Reports/newTR534.pdf>. [Looking from the perspective of ECD at ways that assessment can be improved by developments in statistics, technology, and cognitive psychology.]

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. Forthcoming as a CSE Technical Report. [Exploration of validity, reliability, comparability, and fairness, as viewed from the perspective of evidentiary arguments.]

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. [An initial investigation into model-fit indices for the use of Bayes nets in educational assessments.]