

References

- Alexander, C., Ishikawa, S., & Silverstein, M. (1977). *A pattern language: Towns, buildings, construction*. New York, NY: Oxford University Press.
- Allen, D., Kling, G., & van der Pluijm, B. (2005). Global change curriculum, Unit 1a: Introduction to systems dynamic modeling with STELLA. Ann Arbor, MI: Global change Program, University of Michigan. Retrieved August 2, 2007 from http://www.globalchange.umich.edu/globalchange1/current/labs/Lab2/Intro_Stella.htm
- 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>
- Azevedo, R., & Cromley, J. G. (2004). Does training on self-regulated learning facilitate students' learning with hypermedia? *Journal of Educational Psychology*, 96, 523–535.
- Baxter, G., Elder, A., & Glaser, R. (1996). Knowledge-based cognition and performance assessment in the science classroom. *Educational Psychologist*, 31(2), 133–140.
- Baxter, G., & Mislevy, R. (2005). *The case for an integrated design framework for assessing science inquiry (PADI Technical Report 5)*. Menlo Park, CA: SRI International.
- Bejar, I. I., Mislevy, R. J., Rupp, A. A., & Zhang, M. (2016). Automated scoring with validity in mind. In A. Rupp & J. Leighton (Eds.), *Handbook of cognition and assessment*. Hoboken, NJ: Wiley-Blackwell.
- Belsley, D. A., Kuh, E., & Welch, R. E. (1980). *Regression diagnostics: Identifying influential data and source of collinearity*. New York: John Wiley.
- Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education*, 24, 61–100.
- Buckley, B. C. (2008). Model-based teaching. In N. M. Seel (Ed.), *Encyclopedia of the Sciences of Learning* (pp. 2312–2315). New York: Springer.
- Cartier, J. (2000). *Assessment of explanatory models in genetics: Insights into students' conceptions of scientific models (Research Report No. 98-1)*. Madison, WI: National Center for Improving Student Learning and Achievement in Mathematics and Science.
- Chi, M. T. H. (2005). Common sense conceptions of emergent processes: Why some misconceptions are robust. *Journal of the Learning Sciences*, 14, 161–199.
- Chi, M. T. H., Feltovich, P., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5, 121–152.
- Clarke-Midura, J., Code, J., Zap, N., & Dede, C. (2012). Assessing science inquiry in the classroom: A case study of the virtual assessment project. In L. Lennex & K. Nettleton (Eds.), *Cases on inquiry through instructional technology in math and science: Systemic approaches* (pp. 138–164). New York, NY: IGI.
- Clarke-Midura, J., & Dede, C. (2010). Assessment, technology, and change. *Journal of Research on Technology in Education*, 42(3), 309–328.

- Clement, J. (1989). Learning via model construction and criticism: Protocol evidence on sources of creativity in science. In J. A. Glover, R. R. Ronning, & C. R. Reynolds (Eds.), *Handbook of creativity: Assessment, theory and research* (pp. 341–381). New York: Plenum Press.
- Clement, J. (2000). Model based learning as a key research area for science education. *International Journal of Science Education*, 22, 1041–1053.
- Collins, A., & Ferguson, W. (1993). Epistemic forms and epistemic games: Structures and strategies to guide inquiry. *Educational Psychologist*, 28, 25–42.
- Corcoran, T., Mosher, F. A., Rogat, A. (2009). *Learning progressions in science: An evidence-based approach to reform*. Teachers College-Columbia University: Consortium for Policy Research in Education.
- DeBarger, A. H., Krajcik, J. S., Harris, C. J., & Penuel, W. R. (2013). *Designing NGSS assessment to evaluate the efficacy of curriculum interventions*. Princeton, N.J.: Educational Testing Service K-12 Center.
- DiCerbo, K., Bertling, M., Stephenson, S., Jia, Y., Mislavy, R. J., Bauer, M., et al. (2015). The role of exploratory data analysis in the development of game-based assessments. In C. S. Loh, Y. Sheng, & D. Ifenthaler (Eds.), *Serious games analytics: Methodologies for performance measurement, assessment, and improvement* (pp. 319–342). New York: Springer.
- diSessa, A. A. (1983). Phenomenology and the evolution of intuition. In D. Gentner & A. Stevens (Eds.), *Mental models* (pp. 15–33). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- diSessa, A. A. (1993). Towards an epistemology of physics. *Cognition and Instruction*, 10, 105–225.
- Duncan, R. G. (2006). The role of domain-specific knowledge in promoting generative reasoning. In *Proceedings of the 7th international conference on learning sciences* (pp. 147–153). Bloomington, Indiana: International Society of the Learning Sciences.
- Fauconnier, G., & Turner, M. (2002). *The way we think*. New York: Basic Books.
- Frederiksen, J. R., & White, B. Y. (1988). Implicit testing within an intelligent tutoring system. *Machine-Mediated Learning*, 2, 351–372.
- Frezzo, D. C., Behrens, J. T., & Mislavy, 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*. Retrieved January 27, 2010, from <http://www.springerlink.com/content/566p6g4307405346/>
- Frigg, R. & Hartmann, S. (2006). Models in science. In N. Z. Edward (Ed.). *The Stanford Encyclopedia of Philosophy (Spring 2006 Edition)*. <http://plato.stanford.edu/archives/spr2006/entries/models-science/>
- Gamma, E., Helm, R., Johnson, R., & Vlissides, J. (1994). *Design patterns*. Reading, MA: Addison-Wesley.
- Gentner, D., & Stevens, A. L. (Eds.). (1983). *Mental models*. Hillsdale, NJ: Erlbaum.
- Gibson, D., & Clarke-Midura, J. (2015). Some psychometric and design implications of game-based learning analytics. In P. Isaías, J. M. Spector, D. Ifenthaler, & D. G. Sampson (Eds.), *E-learning systems, environments and approaches* (pp. 247–261). Cham, Switzerland: Springer International Publishing.
- Giere, R. N. (1987). The cognitive study of science. In N. J. Neressian (Ed.), *The process of science* (pp. 139–159). Dordrecht: Martinus Nijhoff.
- Giere, R. N. (2004). How models are used to represent reality. *Philosophy of Science*, 71, 742–752.
- Gilbert, J. K., & Justi, R. (2016). *Modelling-based teaching in science education*. Switzerland: Springer International Publishing.
- Glaser, R., Chi, M. T., & Farr, M. J. (Eds.). (1988). *The nature of expertise*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Gobert, J., & Buckley, B. (2000). Special issue editorial: Introduction to model-based teaching and learning. *International Journal of Science Education*, 22, 891–894.

- Gobert, J. D., Sao Pedro, M., Baker, R. S. J. D., Toto, E., & Montalvo, O. (2012). Leveraging educational data mining for real time performance assessment of scientific inquiry skills within microworlds. *Journal of Educational Data Mining*, 5, 153–185.
- Gorin, J. S., & Mislavy, R. J. (2013). Inherent measurement challenges in the Next Generation Science Standards for both formative and summative assessment. Princeton, NJ: K-12 Center at ETS.
- Gotwals, A. W., & Songer, N. B. (2010). Reasoning up and down a food chain: Using an assessment framework to investigate students' middle knowledge. *Science Education*, 94, 260–281.
- Greeno, J. G. (1989). Situations, mental models, and generative knowledge. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing* (pp. 285–318). Hillsdale, NJ: Lawrence Erlbaum.
- Grosslight, L., Unger, C., Jay, E., & Smith, C. L. (1991). Understanding models and their use in science: Conceptions of middle and high school students and experts. *Journal of Research in Science Teaching*, 28, 799–822.
- Haertel, G. D., Haydel DeBarger, A., Villalba, S., Hamel, L., & Mitman Colker, A. (2010). *Integration of evidence-centered design and universal design principles using PADI, an online assessment design system* (Assessment for Students with Disabilities Technical Report 3). Menlo Park, CA: SRI International. http://padi-se.sri.com/downloads/TR3_IntegratingEDCandUDL.pdf
- Hammer, D., Elby, A., Scherr, R. E., & Redish, E. F. (2005). Resources, framing, and transfer. In J. P. Mestre (Ed.), *Transfer of learning from a modern multidisciplinary perspective* (pp. 89–120). Greenwich, CT: Information Age Publishing.
- Hansen, E. G., Mislavy, 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.
- Harris, C. J., Krajcik, J. S., Pellegrino, J. W., & McElhaney, K. W. (2016). *Constructing assessment tasks that blend disciplinary core Ideas, crosscutting concepts, and science practices for classroom formative applications*. Menlo Park, CA: SRI International.
- Harrison, A. G., & Treagust, D. F. (2000). A typology of school science models. *International Journal of Science Education*, 22, 1011–1026.
- Heller, P., & Heller, K. (2001). *Cooperative group problem solving in physics*. Pacific Grove, CA: Thomson Brooks/Cole.
- Hesse, F., Care, E., Buder, J., Sassenberg, K., & Griffin, P. (2015). A framework for teachable collaborative problem solving skills. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 37–56). Dordrecht, The Netherlands: Springer.
- Hestenes, D. (1987). Toward a modeling theory of physics instruction. *American Journal of Physics*, 55, 440–454.
- Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher*, 30, 141–151.
- Hunt, E., & Minstrell, J. (1994). A cognitive approach to the teaching of physics. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice* (pp. 51–74). Cambridge, MA: MIT Press.
- Ingham, A. M., & Gilbert, J. K. (1991). The use of analogue models by students of chemistry at higher education level. *International Journal of Science Education*, 13, 193–202.
- Johnson, S. K., & Stewart, J. (2002). Revising and assessing explanatory models in a high school class: A comparison of unsuccessful and successful performance. *Science Education*, 86, 463–480.
- Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard University Press.
- Jungck, J. R., & Calley, J. (1985). Strategic simulations and post-Socratic pedagogy: Constructing computer software to develop long-term inference through experimental inquiry. *American Biology Teacher*, 47, 11–15.
- Kahnemann, D. (2011). *Thinking, fast and slow*. NY: Farrar, Straus & Giroux.

- Kalyuga, S. (2006). Rapid cognitive of learners' knowledge structures. *Learning and Instruction*, 16, 1–11.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York: Cambridge University Press.
- Kintsch, W., & Greeno, J. G. (1985). Understanding and solving word arithmetic problems. *Psychological Review*, 92, 109–129.
- Larkin, J. (1983). The role of problem representation in physics. In D. Gentner & Stevens, A. L. (Eds.), *Mental models* (pp. 75–98). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Latour, B. (1987). *Science in Action*. Cambridge, MA: Harvard University Press.
- Lehrer, R., & Schauble, L. (2006). Cultivating in science education. In K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (pp. 371–388). New York: Cambridge University Press.
- Linn, R. L. (2000). Assessments and accountability. *Educational Researcher*, 29(2), 4–16.
- Liu, M., & Haertel, G. (2011). *Design patterns: A tool to support assessment task authoring (Large-scale Assessment Technical Report 11)*. Menlo Park: SRI International.
- Margolis, M. J., & Clauser, B. E. (2006). A regression-based procedure for automated scoring of a complex medical performance assessment. In D. M. Williamson, R. J. Mislevy, & I. I. Bejar (Eds.), *Automated scoring of complex in computer based testing* (pp. 132–167). Mahwah, NJ: Erlbaum.
- Markman, A. B. (1999). *Knowledge representation*. Mahwah, NJ: Erlbaum.
- Marshall, S. P. (1993). Assessing schema knowledge. In N. Frederiksen, R. J. Mislevy, & I. I. Bejar (Eds.), *Test theory for a new generation of tests* (pp. 155–180). Hillsdale, New Jersey: Lawrence Erlbaum.
- Marshall, S. P. (1995). *Schemas in problem solving*. Cambridge University Press, New York.
- Martin, J. D., & VanLehn, K. (1995). A Bayesian approach to cognitive. In P. Nichols, S. Chipman, & R. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 141–165). Hillsdale, NJ: Erlbaum.
- Means, B. & Haertel, G. (2002). Technology supports for assessing science inquiry. In *Technology and assessment: Thinking ahead* (pp. 12–25). National Research Council. Washington, DC: National Academy Press.
- Messick, S. J. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York: Macmillan.
- Messick, S. J. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23(2), 13–23.
- Mislevy, R. J. (2003). Substance and structure in assessment arguments. *Law, Probability, and Risk*, 2, 237–258.
- Mislevy, R. J. (2006). Cognitive psychology and educational assessment. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 257–305). Westport, CT: American Council on Education/Praeger.
- Mislevy, R. J. (2017). Resolving the paradox of rich performance tasks. In H. Jiao & R. W. Lissitz (Eds.), *Test fairness in the new generation of large-scale assessment* (pp. 1–46). Charlotte, NC: Information Age Publishing.
- Mislevy, R. J., Corrigan, S., Oranje, A., DiCerbo, K., John, M., Bauer, M. I., et al. (2014). *Psychometric considerations in game-based*. New York: Institute of Play.
- Mislevy, R. J., Haertel, G., Cheng, B. H., Ructtinger, L., DeBarger, A., Murray, E., et al. (2013). A “conditional” sense of fairness in assessment. *Educational Research and Evaluation*, 19, 121–140.
- Mislevy, R.J., Hamel, L., Fried, R., Gaffney, G., Haertel, T., Hafter G., et al. (2003). *Design patterns for assessing science inquiry (PADI Technical Report 1)*. Menlo Park: SRI International.
- 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.

- Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives, 1*, 3–67.
- Mislevy, R. J., Steinberg, L. S., Breyer, F. J., Johnson, L., & Almond, R. A. (2002). Making sense of data from complex assessments. *Applied Measurement in Education, 15*, 363–378.
- Mosteller, F., & Tukey, J. W. (1977). *Data analysis and regression: A second course in statistics*. Reading, MA: Addison-Wesley.
- National Research Council. (1996). *National Science Education Standards*. Washington, DC: National Academy Press.
- National Research Council (2000). How people learn: Brain, mind, experience and school (2nd ed.). In J. D. Bransford, A. L. Brown, & R. Cocking (Eds.). *Committee on developments in the science of learning*. Washington, DC: National Academy Press.
- National Research Council. (2001). *Knowing what students know: The science and design of educational*. Committee on the Foundations of Assessment, J. Pellegrino, R. Glaser, & N. Chudowsky (Eds.), Washington DC: National Academy Press.
- National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Committee on a Conceptual Framework for New K-12 Science Education Standards. Board on Science Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.
- Newell, A., & Simon, H.A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.
- NGSS Lead States. (2013a). *How to read the Next Generation Science Standards*. Retrieved January 23, 2016 from <http://www.nextgenscience.org/sites/ngss/files/How%20to%20Read%20NGSS%20-%20Final%204-19-13.pdf>
- NGSS Lead States. (2013b). *Next Generation Science Standards: For States, By States*. Washington, DC: The National Academies Press.
- Norman, D. A. (1993). *Things that make us smart*. Boston: Addison-Wesley.
- Novick, S., & Nussbaum, J. (1981). Pupils' understanding of the particulate nature of matter: A cross-age study. *Science Education, 65*, 187–196.
- Pellegrino, J. W. (2013). Proficiency in science: Assessment challenges and opportunities. *Science, 340*, 320–323.
- Perkins, D. N., & Salomon, G. (1989). Are cognitive skills context-bound? *Educational Researcher, 18*, 16–25.
- Piaget, J. (1976). Piaget's theory. In B. Inhelder, H. H. Chipman, & C. Zwingmann (eds.). *Piaget and his school* (pp. 11–23). Springer: Berlin Heidelberg.
- Quellmalz, E. S., Timms, M. J., Buckley, B. C., Davenport, J., Loveland, M., & Silberglitt, M. D. (2012). 21st century dynamic assessment. In M. Mayrath, J. Clarke-Midura, & D. H. Robinson (Eds.), *Technology-based assessments for 21st century skills: Theoretical and practical implications from modern research*. Charlotte, NC: Information Age.
- Redish, E. F. (2004). A theoretical framework for physics education research: Modeling student thinking. In E. F. Redish & M. Vicentini (Eds.), *Proceedings of the Enrico Fermi Summer School Course, CLVI* (pp. 1–63). Bologna, Italy: Italian Physical Society.
- Richmond, B. (2005). *An introduction to systems thinking, featuring STELLA*. Lebanon, NH: isee systems.
- Riconscente, M. M., Mislevy, R. J., & Corrigan, S. (2015). In S. Lane, T. M. Haladyna, & M. Raymond (Eds.), *Handbook of test development* (2nd ed., pp. 40–63). Routledge: Informa, Taylor & Francis.
- Riley, M. S., Greeno, J. G., & Heller, J. I. (1983). Development of children's problem-solving ability in arithmetic. In H. P. Ginsburg (Ed.), *The development of mathematical thinking* (pp. 153–196). New York: Academic Press.
- Rose, D, Murray, E., & Gravel, J. (2012). UDL and the PADI process: The foundation (Assessment for Students with Disabilities Technical Report 4). Menlo Park, CA: SRI International. http://padi-se.sri.com/downloads/TR4_UDL_Tech_Report2012FL.pdf
- Ruiz-Primo, M. A., & Shavelson, R. J. (1996). Rhetoric and reality in science performance assessments: An update. *Journal of Research in Science Teaching, 33*, 1045–1063.

- Strauss, C., & Quinn, N. (1998). *A cognitive theory of cultural meaning*. New York: Cambridge University Press.
- Suárez, M. (2004). An inferential conception of scientific representation. *Philosophy of Science*, 71, 767–779.
- Swoyer, C. (1991). Structural representation and surrogative reasoning. *Synthese*, 87, 449–508.
- Thornton, R. K., & Sokoloff, D. R. (1998) Assessing student learning of Newton's laws: The force and motion conceptual evaluation and the evaluation of active learning laboratory and lecture curricula. *American Journal of Physics*, 66, 338–351.
- Toulmin, S. E. (1958). *The uses of argument*. Cambridge, England: Cambridge University Press.
- Wertsch, J. V. (1998). *Mind as action*. Oxford: Oxford University Press.
- West, P., Wise Rutstein, D., Mislevy, R. J., Liu, J., Levy, R., DiCerbo, K. E., et al. (2012). A Bayesian network approach to modeling learning progressions. In A. C. Alonzo & A. W. Gotwals (Eds.), *Learning progressions in science* (pp. 255–291). Rotterdam: Sense Publishers.
- White, B. Y., & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction*, 16, 3–118.
- White, B. Y., Shimoda, T. A., & Frederiksen, J. R. (1999). Enabling students to construct theories of collaborative inquiry and reflective learning: Computer support for metacognitive development. *International Journal of Artificial Intelligence in Education*, 10, 151–182.
- Wiley, D. E., & Haertel, E. H. (1996). Extended assessment tasks: Purposes, definitions, scoring, and accuracy. In M. B. Kane & R. Mitchell (Eds.), *Implementing performance assessments: Promises, problems, and challenges*. Mahwah, NJ: Erlbaum.
- 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. http://ecd.sri.com/downloads/ECD_TR10_Learning_Progressions.pdf