The John D. and Catherine T. MacArthur Foundation Reports on Digital Media and Learning

# Measuring What Matters Most

## Choice-Based Assessments for the Digital Age

Daniel L. Schwartz and Dylan Arena

MacArthur Foundation

John D. and Catherine T. MacArthur Foundation

www.macfound.org

## Measuring What Matters Most

This report was made possible by grants from the John D. and Catherine T. MacArthur Foundation in connection with its grant making initiative on Digital Media and Learning. For more information on the initiative visit http://www.macfound.org.

**The John D. and Catherine T. MacArthur Foundation Reports on Digital Media and Learning**

*Peer Participation and Software: What Mozilla Has to Teach Government,* by David R. Booth

*The Future of Learning Institutions in a Digital Age,* by Cathy N. Davidson and David Theo Goldberg with the assistance of Zoë Marie Jones

*The Future of Thinking: Learning Institutions in a Digital Age,* by Cathy N. Davidson and David Theo Goldberg with the assistance of Zoë Marie Jones

*Kids and Credibility: An Empirical Examination of Youth, Digital Media Use, and Information Credibility,* by Andrew J. Flanagin and Miriam Metzger with Ethan Hartsell, Alex Markov, Ryan Medders, Rebekah Pure, and Elisia Choi

*New Digital Media and Learning as an Emerging Area and "Worked Examples" as One Way Forward,* by James Paul Gee

*Digital Media and Technology in Afterschool Programs, Libraries, and Museums,* by Becky Herr-Stephenson, Diana Rhoten, Dan Perkel, and Christo Sims with contributions from Anne Balsamo, Maura Klosterman, and Susana Smith Bautista

*Living and Learning with New Media: Summary of Findings from the Digital Youth Project,* by Mizuko Ito, Heather Horst, Matteo Bittanti, danah boyd, Becky Herr-Stephenson, Patricia G. Lange, C. J. Pascoe, and Laura Robinson with Sonja Baumer, Rachel Cody, Dilan Mahendran, Katynka Z. Martínez, Dan Perkel, Christo Sims, and Lisa Tripp

*Young People, Ethics, and the New Digital Media: A Synthesis from the GoodPlay Project,* by Carrie James with Katie Davis, Andrea Flores, John M. Francis, Lindsay Pettingill, Margaret Rundle, and Howard Gardner

*Confronting the Challenges of Participatory Culture: Media Education for the 21st Century,* by Henry Jenkins (PI) with Ravi Purushotma, Margaret Weigel, Katie Clinton, and Alice J. Robison

*The Civic Potential of Video Games,* by Joseph Kahne, Ellen Middaugh, and Chris Evans

*Quest to Learn: Developing the School for Digital Kids,* by Katie Salen, Robert Torres, Loretta Wolozin, Rebecca Rufo-Tepper, and Arana Shapiro

*Measuring What Matters Most: Choice-Based Assessments for the Digital Age,* by Daniel L. Schwartz and Dylan Arena

*Learning at Not-School? A Review of Study, Theory, and Advocacy for Education in Non-Formal Settings,* by Julian Sefton-Green

# Measuring What Matters Most

## Choice-Based Assessments for the Digital Age

**Daniel L. Schwartz and Dylan Arena**

**The MIT Press**
**Cambridge, Massachusetts**
**London, England**

This book was set in Stone Serif and Stone Sans by the MIT Press. Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Schwartz, Daniel L.
Measuring what matters most : choice-based assessments for the digital age / Daniel L. Schwartz and Dylan Arena.
p. cm. — (The John D. and Catherine T. MacArthur Foundation reports on digital media and learning)
Includes bibliographical references.
ISBN 978-0-262-51837-6 (pbk. : alk. paper)
1. Educational tests and measurements—Data processing. 2. Decision-making—Evaluation. I. Arena, Dylan. II. Title.
LB3060.55.S39 2013
371.260285—dc23
2012029445

10 9 8 7 6 5 4 3 2 1

## Contents

## Series Foreword

The John D. and Catherine T. MacArthur Foundation Reports on Digital Media and Learning, published by the MIT Press in collaboration with the Monterey Institute for Technology and Education (MITE), present findings from current research on how young people learn, play, socialize, and participate in civic life. The reports result from research projects funded by the MacArthur Foundation as part of its fifty million dollar initiative in digital media and learning. They are published openly online (as well as in print) to support broad dissemination and stimulate further research in the field.

## Acknowledgments

We would like to thank the members of the AAA Lab at Stanford University (http://aaalab.stanford.edu) for their contributions to the ideas and research presented here. Additionally, we wish to acknowledge their infinite patience when listening to the authors endlessly trying out pithy sentences in meetings. The development of this book and its ideas was supported by a grant from the MacArthur Foundation to James Gee, and then to the first author. The material includes work supported by the National Science Foundation under Grant No 0904324. Any opinions, findings, conclusions, or recommendations are those of the authors and do not necessarily reflect the views of the granting agencies.

# I What Matters

# 1 Beliefs about Useful Learning

Educational assessment is a normative endeavor. The ideal assessment both reflects and reinforces educational goals that society deems valuable. One fundamental goal of education is to prepare students to act independently in the world—which is to say, make good choices. It follows that an ideal assessment would measure how well we are preparing students to do so. The argument of this report is that current assessments, which primarily focus on how much knowledge and skills students have accrued, are inadequate. Choice, rather than knowledge, should be the interpretative frame within which learning assessments are organized. Digital technologies make this possible, because interactive assessments can evaluate students in a context of choosing whether, what, how, and when to learn.

In education, most people see choice as a catalyst for learning. For instance, giving students choices can increase their motivation and learning (Iyengar and Lepper 1999). Choice is also important for learning, if only because students need to experience choices in the protected atmosphere of education so they can learn how to handle them before becoming independent.

The current assertion starts differently. It examines why choice should be viewed as the outcome of learning and not

solely an instructional ingredient to improve learning. We contend that choice should be the interpretative framework for understanding learning outcomes. To achieve this reorientation in how people think about learning, assessment provides a powerful lever. Assessments shape the public mind, and everything else flows from that.

Assessment is not a sexy topic. It is tolerated as a necessary nuisance. This is the dulling fog that comes from accepting the premise that what exists must exist. Do not underestimate the power of assessments or the degree to which they have shaped how you think about learning.

Formulated in 1956, Benjamin Bloom's taxonomy of educational outcomes is still arguably one of the most influential frameworks for the design of instruction. It describes a pyramid of the following order, going from bottom to top: memory (called "knowledge" back then), comprehension, application, analysis, synthesis, and evaluation. Bloom's taxonomy was designed by a committee as an *assessment* framework, not an instructional one. It is not based on learning or pedagogical theory. Yet in the way that assessments always manage to do, it has commanded the instructional enterprise. Based on the pyramid, many people believe that students must first learn from the bottom of the pyramid (memorize) before engaging in higher-order thinking near the top (evaluate). This belief is wrong. Most people would recognize this if they could reclaim their common sense from the grip of assessment. For example, comprehension improves the formation of memories (Bransford and Johnson 1972), so making memories a prerequisite for comprehension does not work well. Similarly, having students learn a new topic

in an application context is a great way to help them simultaneously learn the facts and evaluate their applications.

People have beliefs about learning that are mistaken. Current classroom and high-stakes assessments are largely responsible for this situation, because they send the wrong message about what matters. Teachers may tell students about the importance of persistence, critical thinking, interest development, and a host of other keys to a successful life. But tests provide the empirical evidence that students use to decide what is truly valued. If an assessment focuses on the retrieval and procedural application of narrow skills and facts, this is what students will think counts as useful learning. How can they not? It is the basis for promotion and approbation. By changing assessments to concentrate on choices, we should be able to improve beliefs about what constitutes useful learning.

There is a befuddling but extremely strong correlation in the Trends in International Mathematics and Science Study—an assessment taken by students around the world (http://timssandpirls.bc.edu). It is meant to help nations decide their standing. The study's actionable information is at the level of national policy rather than teachers and students. The odd finding is that the students of the nations that do the best on the test also exhibit the least "liking" of mathematics and science (e.g., Shen 2002). The better a nation scores on the math or science tests, the less interest the children there have in pursuing math or science. Nobody knows exactly why the negative correlation is so strong. There may be some statistical oddity that involves averaging individuals to compare nations (Robinson 1950). There are also more substantive possibilities. One is that

students who do the best on these tests spend a lot of time learning with testlike questions. They interpret these questions as markers of what it means to have learned in science and mathematics. They do not like the vision that from their test-based vantage, learning is primarily an act of replicating what they have been told. It makes sense that they would not like math and science, despite doing well. They have missed the generative and contributive aspects of learning. Under this interpretation, rather than helping to prepare students for future learning in science and math, current assessments are propelling students to choose not to learn these domains.

Distortions of what counts as useful learning suffuse US culture. Our greatest fear is that those fortunate enough to have the resources to guide education may also have distorted visions of learning. What could be worse than creating educational technologies that become increasingly efficient at teaching the wrong thing? Successful people have gained many implicit lessons about what it took for them to achieve their successes, often accompanied by narratives of passion and perseverance. Yet these same people are at risk of supporting learning environments that ignore those lessons, and instead teach to outcomes that seem mostly important for standardized and end-of-chapter tests. Such is the sway of assessments.

The aim of assessment should be to advance the goals of society rather than misrepresent them. With new developments in technology, it should be possible to advance goals that were beyond the reach of prior assessments. To date, this has not been the case. Howard Wainer (2010, 17) argues that "the promise of [computerized testing] has yet to be fully realized. So far,

when it has been applied, it has been used as a mechanical horse, not doing much more than could have been done with paper and pencil testing except that it is faster (a little) and more expensive (a lot)." We believe it is possible to do better, and the following is our plan.

In chapter 2, we situate our discussion in the context of new technologies that make it possible for choice to become the core of assessment (and not in the degraded sense of multiple-choice tests). We also provide an anchoring example of a computerized, choice-based assessment. In part II, we turn to theoretical matters to help unseat current beliefs about what we should be assessing. Chapter 3 maintains that choice is what most of the stakeholders in education care about, despite the fact that they often talk in terms of knowledge and skills. To make room for choice-based assessment, chapter 4 tries to clarify why knowledge-based assessments are a mismatch for the aims of education. The chapter highlights the fact that knowledge has not always been the frame of assessment and that the current emphasis on knowledge has made it difficult to connect assessments to outcomes beyond knowledge. Chapter 5 continues the argument by focusing on the static nature of knowledge assessments, and it offers an alternative model of a dynamic assessment that evaluates learning in action.

In part III, we turn to more practical matters. Chapter 6 provides several concrete cases of choice-based assessments that reveal what knowledge-based assessments cannot—for example, persistence after failure. Chapter 7 considers a related practical matter: twenty-first-century standards. The chapter supplies a pair of organizing frames that can integrate choice outcomes

into standards while avoiding laundry lists of goals, which can leave assessment designers without guiding principles.

In part IV, we turn to matters of practice. We concentrate on the practice of designing assessments. Chapter 8 provides a brief tutorial on technical aspects of assessment, including constructs, validity, and reliability. Reliability, in particular, is problematic, because it presumes a stable construct, whereas education presupposes a trajectory of change. Chapter 9 contends that assessments would be more useful if we loosen the grip of some past approaches, so that assessments can be designed to evaluate learning experiences rather than just individual student achievement. Also, new computational developments make it possible to handle much more complex views of learning, but this depends on exploratory data mining as opposed to hypothesis testing. Chapter 10 lays out a research and development agenda for creating choice-based assessments. It includes the description of new platforms for democratizing and crowdsourcing the design and evaluation of assessments, along with several methodological strategies for making headway.

In part V, we turn to the most difficult aspect of assessment. Chapter 11 considers issues of fairness, where there is a delicate balance between encouraging and forcing good choices. In fact, before we move forward in our argument, we should clarify what we mean when we use the term *choice*. We take it as foundational that a primary goal of education is to help students develop aspirations and understandings so they can make choices that maximize their chances of succeeding within and beyond school, and we believe, therefore, that choice should be at the heart of assessment. Yet we recognize that not all choices

are in the purview of education. Choice assessments should not be a backdoor way to enforce beliefs that fall outside the domain of publicly sponsored education (such as whether students make the "correct" choice about a political or religious matter). Instead, choice-based assessments should indicate whether students can learn and adapt in productive ways. Our discussion of choice-based assessments thus refers to learning-relevant choices such as how and what to learn, not all choices. Nevertheless, measuring choices—the stuff of agency and freedom—raises difficult questions about the province of education in shaping and assessing children. Choice-based assessments bring issues of fairness into helpful relief. Chapter 12 summarizes our argument.

## References

Anderson, J. R. 2000. *Cognitive Psychology and Its Implications*. 5th ed. New York: Worth.

Anderson, K. T., S. J. Zuiker, G. Taasoobshirazi, and D. T. Hickey. 2011. "Classroom Discourse as a Tool to Enhance Formative Assessment and Practice in Science." *International Journal of Science Education* 29:1721–1744.

Anderson, R. C., and P. D. Pearson. 1984. "A Schema-Theoretic View of Basic Processes in Reading." In *Handbook of Reading Research*, ed. P. D. Pearson, R. Barr, M. L. Kamil, and P. Mosenthal, 255–291. White Plains, NY: Longman.

Appelbaum, E., and R. Batt. 1994. *The New American Workplace*. New York: ILR Press.

Arena, D. A. 2012. "Commercial Video Games as Preparation for Future Learning." PhD diss., Stanford University.

Arena, D., and D. L. Schwartz. 2010. Stats Invaders! Learning about Statistics by Playing a Classic Video Game. In *Proceedings of the Fifth International Conference on Foundations of Digital Games*, ed. I. Horswill and Y. Pisan, 248–249. New York: ACM.

Baker, R.S.J.d., S. Gowda, and A. Corbett. 2011. "Automatically Detecting a Student's Preparation for Future Learning: Help Use Is Key." In Proceedings of the Fourth International Conference on Educational

Data Mining, ed. M. Pechenizkiy, T. Calders, C. Conati, S. Ventura, C. Romero, and J. Stamper, 179–188. Eindhoven, Netherlands.

Bandura, A. 1989. "Human Agency in Social Cognitive Theory." *American Psychologist* 44, no. 9:1175–1184.

Barab, S. A., and J. A. Plucker. 2002. "Smart People or Smart Contexts? Cognition, Ability, and Talent Development in an Age of Situated Approaches to Knowing and Learning." *Educational Psychologist* 37, no. 3:165–182.

Barnett, S. M., and S. J. Ceci. 2002. "When and Where Do We Apply What We Learn? A Taxonomy for Far Transfer." *Psychological Bulletin* 128:612–637.

Barron, F., and D. M. Harrington. 1981. "Creativity, Intelligence, and Personality." *Annual Review of Psychology* 32:439–476.

Bell, P., B. Lewenstein, A. W. Shouse, and M. A. Feder, eds. 2009. *Learning Science in Informal Environments: People, Places, and Pursuits*. Committee on Learning Science in Informal Environments, Board on Science Education, Center for Education, Division of Behavioral and Social Sciences and Education, National Research Council. Washington, DC: National Academies Press.

Benner, C. 2002. *Work in the New Economy: Flexible Labor Markets in Silicon Valley*. Oxford: Blackwell Publishing.

Biederman, I., and M. M. Shiffrar. 1987. "Sexing Day-Old Chicks: A Case Study and Expert Systems Analysis of a Difficult Perceptual Learning Task." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 13:640–645.

Black, P., and D. Williams. 1998. "Inside the Black Box: Raising Standards through Classroom Assessment." *Phi Delta Kappan* 80, no. 2:139–148.

Blackwell, L. S., K. H. Trzesniewski, and C. S. Dweck. 2007. "Implicit Theories of Intelligence Predict Achievement across an Adolescent Tran-

sition: A Longitudinal Study and an Intervention." *Child Development* 78, no. 1:246–263.

Blair, K. 2009. "The Neglected Importance of Feedback Perception in Learning: An Analysis of Children and Adults' Update of Quantitative Feedback in a Mathematics Simulation Environment." PhD diss., Stanford University.

Bransford, J. D., and M. K. Johnson. 1972. "Contextual Prerequisites for Understanding: Some Investigators of Comprehension and Recall." *Journal of Verbal Learning and Verbal Behavior* 11:717–726.

Bransford, J. D., and D. L. Schwartz. 1999. "Rethinking Transfer: A Simple Proposal with Multiple Implications." *Review of Research in Education* 24:61–100.

Broudy, H. S. 1977. "Types of Knowledge and Purposes of Education." In *Schooling and the Acquisition of Knowledge*, ed. R. C. Anderson, R. J. Spiro, and W. E. Montague, 1–17. Hillsdale, NJ: Erlbaum.

Buhrmester, M., T. Kwang, and S. D. Gosling. 2011. "Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data?" *Perspectives on Psychological Science* 6, no. 1:3–5.

Callan, E. 2009. "Why Bring the Kids into This? McLaughlin and Anscombe on Religious Understanding and Upbringing." In *Faith in Education: A Tribute to Terence MacLaughlin*, ed. G. Haydon. London: London Institute of Education Publications.

Chase, C. C. 2011. "Motivating Persistence in the Face of Failure: The Impact of an Ego-Protective Buffer on Learning Choices and Outcomes in a Computer-Based Educational Game." PhD diss., Stanford University.

Chase, C. C., D. B. Chin, M. A. Oppezzo, and D. L. Schwartz. 2009. "Teachable Agents and the Protégé Effect: Increasing the Effort towards Learning." *Journal of Science Education and Technology* 18:334–352.

Chomsky, N. 1966. *Cartesian Linguistics: A Chapter in the History of Rationalist Thought*. New York: Harper & Row.

Collins, A., and R. Pea. 2011. "The Advantages of Alternative Certifications for Students." *Education Week* 31, no. 8:22–23.

Cooper, S., F. Khatib, A. Treuille, J. Barbero, J. Lee, M. Beenen, A. Leaver-Fay, D. Baker, Z. Popović, and Foldit players. 2010. "Predicting Protein Structures with a Multiplayer Online Game." *Nature* 446:756–760.

Costa, P. T., Jr., and R. R. McCrae. 1992. *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) Manual*. Odessa, FL: Psychological Assessment Resources.

Crocker, L., and J. Algina. 1986. *Introduction to Classical and Modern Test Theory*. New York: Harcourt Brace Jovanovich.

Damasio, A. R. 1994. *Descartes Error: Emotion, Reason and the Human Brain*. New York: G. P. Putnam's Sons.

DeYoung, C. G., J. B. Hirsh, M. S. Shane, X. Papademetris, N. Rajevaan, and J. R. Gray. 2010. "Testing Predictions from Personality Neuroscience: Brain Structure and the Big Five." *Psychological Science* 21:820–828.

Dow, S. P., A. Glassco, J. Kass, M. Schwarz, D. L. Schwartz, and S. R. Klemmer. 2010. "Parallel Prototyping Leads to Better Design Results, More Divergent Creations, and Self-Efficacy Gains." *ACM Transactions on Computer-Human Interaction* 17, no. 4:1–24.

Ericsson, K. A., ed. 2009. *Development of Professional Expertise: Toward Measurement of Expert Performance and Design of Optimal Learning Environments*. New York: Cambridge University Press.

Ericsson, K. A., R. Th. Krampe, and C. Tesch-Römer. 1993. "The Role of Deliberate Practice in the Acquisition of Expert Performance." *Psychological Review* 100:363–406.

Falk, J. H., and L. D. Dierking. 2002. *Lessons without Limit: How Free-Choice Learning Is Transforming Education*. Walnut Creek, CA: AltaMira Press.

Feuerstein, R. 1979. *The Dynamic Assessment of Retarded Performers: The Learning Potential Assessment Device, Theory, Instruments, and Techniques*. Baltimore, MD: University Park Press.

Frezzo, D. C., J. T. Behrens, R. J. Mislevy, P. West, and K. E. DiCerbo. 2009. "Psychometric and Evidentiary Approaches to Simulation Assessment in Packet Tracer Software." *ICNS '09: Proceedings of the Fifth International Conference on Networking and Services*, eds. J. L. Mauri, V. Casares, R. Tomas, T. Serra, and O. Dini, 555–560. Washington, DC: IEEE Computer Society.

Gee, J. P. 2003. *What Video Games Have to Teach Us about Learning and Literacy*. New York: Palgrave.

Ghiselli, E. E. 1966. *The Validity of Occupational Aptitude Tests*. New York: Wiley.

Gibson, J. J., and E. J. Gibson. 1955. "Perceptual Learning: Differentiation or Enrichment." *Psychological Review* 62:32–51.

Gough, H. G. 1979. "A Creative Personality Scale for the Adjective Check List." *Journal of Personality and Social Psychology* 37, no. 8:1398–1405.

Greeno, J. G. Forthcoming. "Framing Intellective Identities, Intellective Characters, and Epistemic Values." In *Educating Comprehensively: Varieties of Educational Experiences*, ed. L. Lin, H. Varenne, and E. W. Gordon. Vol. 3. Lewiston, NY: Edwin Mellen Press.

Gresalfi, M. S. 2009. "Taking Up Opportunities to Learn: Constructing Dispositions in Mathematics Classrooms." *Journal of the Learning Sciences* 18:327–369.

Haertel, E. H., and J. L. Herman. 2005. "A Historical Perspective on Validity Arguments for Accountability Testing." In *Uses and Misuses of Data in Accountability Testing. Yearbook of the National Society for the Study of Education*, ed. J. L. Herman and E. H. Haertel, 104:1–34. Boston, MA: Blackwell Publishing.

Harré, R., and L. van Langenhove, eds. 1999. *Positioning Theory: Moral Contexts of Intentional Action*. Oxford: Blackwell.

Hatano, G., and K. Inagaki. 1986. "Two Courses of Expertise." In *Child Development and Education in Japan*, ed. H. Stevenson, H. Azuma, and K. Hakuta, 262–272. New York: Freeman.

Hatano, G., and K. Osawa. 1983. "Digit Memory of Grand Experts in Abacus-Derived Mental Calculation." *Cognition* 15:95–110.

Ito, M., S. Baumer, M. Bittanti, d. boyd, R. Cody, B. Herr-Stephenson, H. A. Horst, P. G. Lange, D. Mahendran, K. Z. Martinez, C. J. Pascoe, D. Perkel, L. Robinson, C. Sims, and L. Tripp. 2009. *Hanging Out, Messing Around, Geeking Out: Kids Living and Learning with New Media*. Cambridge, MA: MIT Press.

Iyengar, S. S., and M. R. Lepper. 1999. "Rethinking the Value of Choice: A Cultural Perspective on Intrinsic Motivation." *Journal of Personality and Social Psychology* 76:349–366.

Jeong, H., A. Gupta, R. Roscoe, J. Wagster, G. Biswas, and D. L. Schwartz. 2008. "Using Hidden Markov Models to Characterize Student Behaviors in Learning-by-Teaching Environments." *Proceedings of the Ninth International Conference on Intelligent Tutoring Systems*, 614–625.

Kaminski, J. A., V. M. Sloutsky, and A. F. Heckler. 2008. "Learning Theory: The Advantage of Abstract Examples in Learning Math." *Science* 320:454–455.

Kendall-Taylor, N., E. Lindland, and A. Mikulak. 2010. *"Faster and Fancier Books": Mapping the Gaps between Expert and Public Understandings of Digital Media and Learning*. Washington, DC: FrameWorks Institute.

Koedinger, K. R., and J. R. Anderson. 1997. "Intelligent Tutoring Goes to School in the Big City." *International Journal of Artificial Intelligence in Education* 8:30–43.

Konold, C. 1989. "Informal Conceptions of Probability." *Cognition and Instruction* 6, no. 1:59–98.

La Paro, K. M., R. C. Pianta, and M. Stuhlman. 2004. "The Classroom Assessment Scoring System: Findings from the Prekindergarten Year." *Elementary School Journal* 104:409–426.

Lederman, L., with D. Teresi. 1993. *The God Particle: If the Universe is the Answer, What Is the Question?* New York: Houghton Mifflin.

Li, C., and G. Biswas. 2002. "A Bayesian Approach for Learning Hidden Markov Models from Data." *Science Progress* 10:201–219.

Martin, L., and D. L. Schwartz. 2009. "Prospective Adaptation in the Use of Representational Tools." *Cognition and Instruction* 27, no. 4:370–400.

Massa, L. J., and R. E. Mayer. 2006. "Testing the ATI hypothesis: Should multimedia instruction accommodate verbalizer-visualizer cognitive style?" *Learning and Individual Differences* 16:321–336.

McDermott, R. 1993. "Acquisition of a Child by a Learning Disability." In *Understanding Practice*, ed. S. Chaiklin and J. Lave, 269–305. London: Cambridge University Press.

Mislevy, R. J., L. S. Steinberg, and R. G. Almond. 2003. "On the Structure of Educational Assessments." *Measurement: Interdisciplinary Research and Perspectives* 1:3–67.

Moss, P. A., D. C. Pullin, J. P. Gee, E. H. Haertel, and L. J. Young, eds. 2008. *Assessment, Equity, and Opportunity to Learn*. Cambridge: Cambridge University Press.

National Research Council. 2007. *Taking Science to School*: *Learning and Teaching Science in Grades K–8*. Washington, DC: The National Academies Press.

National Research Council. 2012. *A Framework for K–12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. Washington, DC: National Academies Press.

Newell, A., and H. A. Simon. 1972. *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall.

Nisbett, R. E., D. Krantz, C. Jepson, and Z. Kunda. 1983. "The Use of Statistical Heuristics in Everyday Inductive Reasoning." *Psychological Review* 90:339–363.

Norris, S. P. 1985. "Synthesis of Research on Critical Thinking." *Educational Leadership* (May):40–45.

Olivier, C., and R. F. Bolwer. 1996. *Learning to Learn*. New York: Fireside.

Ortega y Gasset, J. 1960. What Is Philosophy? Trans. Mildred Adams. new York: W. W. Norton.

Pinker, S. 1994. *The Language Instinct: How the Mind Creates Language*. New York: HarperCollins.

Polyani, M. 1966. *The Tacit Dimension*. Garden City, NY: Doubleday.

Power, D. J. 2002. "What Is the 'True Story' about Data Mining, Beer, and Diapers?" *DSS News* 3, no. 23. http://www.dssresources.com/newsletters/66.php.

Rabiner, L. R. 1989. "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition." *Proceedings of the IEEE* 77:257–286.

Rafferty, A. N. 2007. "Using FACT to Challenge Assumptions: Frequency, Accuracy, Choice, and Timing in Machine Learning." Master's thesis, Stanford University.

Reeves, B., and J. L. Read. 2009. *Total Engagement: Using Games and Virtual Environments to Change the Way People Work and Businesses Compete*. Cambridge, MA: Harvard Business Press.

Resnick, D. P., and L. B. Resnick. 1994. "Performance Assessment and the Multiple Functions of Educational Measurement." In *Implementing Performance Assessment: Promises, Problems, and Challenges*, ed. M. Kane and R. Mitchell, 23–38. Mahwah, NJ: Erlbaum.

Richland, L. E., O. Zur, and K. J. Holyoak. 2007. "Cognitive Supports of Analogies in the Mathematics Classroom." *Science* 316:1128–1129.

Robinson, W. S. 1950. "Ecological Correlations and the Behavior of Individuals." *American Sociological Review* 15, no. 3:351–357.

Rousseau, J. J. (1762) 1947. *The Social Contract, or Principles of Political Right*. Translated by C. Frankel. New York: Hafner.

Schacter, D. L. 1987. "Implicit Memory: History and Current Status." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 13:501–518.

Schwartz, B. 2004. *The Paradox of Choice: Why More Is Less*. New York: HarperCollins.

Schwartz, D. L., and J. B. Black. 1996. "Analog Imagery in Mental Model Reasoning: Depictive Models." *Cognitive Psychology* 30:154–219.

Schwartz, D. L., and J. D. Bransford. 1998. "A Time for Telling." *Cognition and Instruction* 16:475–522.

Schwartz, D. L., J. D. Bransford, and D. Sears. 2005. "Efficiency and Innovation in Transfer." In *Transfer of Learning from a Modern Multidisciplinary Perspective*, ed. J. P. Mestre, 1–52. Greenwich, CT: Information Age Publishing.

Schwartz, D. L., C. C. Chase, D. B. Chin, M. A. Oppezzo, H. Kwong, S. Okita, G. Biswas, R. Roscoe, H. Jeong, and J. Wagster. 2009. "Interactive Metacognition: Monitoring and Regulating a Teachable Agent." In *Handbook of Metacognition in Education*, ed. D. J. Hacker, J. Dunlosky, and A. C. Graesser, 340–358. New York: Routledge.

Schwartz, D. L., C. C. Chase, M. A. Oppezzo, and D. B. Chin. 2011. "Practicing versus Inventing with Contrasting Cases: The Effects of Telling First on Learning and Transfer." *Journal of Educational Psychology* 103, no. 4:759–775.

Schwartz, D. L., and T. Martin. 2004. "Inventing to Prepare for Learning: The Hidden Efficiency of Original Student Production in Statistics Instruction." *Cognition and Instruction* 22:129–184.

Schwartz, D. L., T. Martin, and J. Chang. 2008. "Innovation and Instrumentation: Taking the Turn to Efficiency." In *Handbook of Design Research Methods in Education: Innovations in Science, Technology, Engineering, and Mathematics Learning and Teaching*, ed. A. E. Kelly, R. A. Lesh, and J. Y. Baek, 47–67. New York: Routledge.

Shaffer, D. W. 2006. *How Computer Games Help Children Learn*. New York: Macmillan.

Shavelson, R. J., G. P. Baxter, and J. Pine. 1991. "Performance Assessment in Science." *Applied Measurement in Education* 4:347–362.

Shen, C. 2002. "Revisiting the Relationship between Students' Achievement and Their Self-perceptions: A Cross-national Analysis Based on TIMSS 1999 Data. Assessment in Education: Principles." *Policy and Practice* 9, no. 2:161–184.

Shute, V. J. 2011. "Stealth Assessment in Computer-Based Games to Support Learning." In *Computer Games and Instruction*, ed. S. Tobias and J. D. Fletcher, 503–524. Charlotte, NC: Information Age Publishers.

Skinner, B. F. 1986. "Programmed Instruction Revisited." *Phi Delta Kappan* 68, no. 2:103–110.

Snow, R. 1989. "Aptitude-Treatment Interaction as a Framework for Research on Individual Differences in Learning." In *Learning and Individual Differences*, ed. P. Ackerman, R. J. Sternberg, and R. Glaser, 13–58. New York: W. H. Freeman.

Sternberg, R. J. 2006. "The Nature of Creativity." *Creativity Research Journal* 18, no. 1:87–98.

Stevens, R. H., and V. Thadani. 2007. "Quantifying Student's Scientific Problem Solving Efficiency and Effectiveness." *Technology, Instruction, Cognition, and Learning* 5:325–337.

Stroop, J. R. 1935. "Studies of Interference in Serial Verbal Reactions." *Journal of Experimental Psychology* 18, no. 6:643–662.

Tai, R. H., C. Q. Liu, A. V. Maltese, and X. Fan. 2006. "Planning Early for Careers in Science." *Science* 312:1143–1144.

Tancer, B. 2008. *Click: What Millions of People Are Doing Online and Why It Matters*. New York: Hyperion.

Taylor, J. L., K. M. Smith, A. P. van Stolk, and G. B. Spiegelman. 2010. "Using Invention to Change How Students Tackle Problems." *CBE Life Sciences Education* 8:504–512.

Tobias, S. 2009. "An Eclectic Appraisal of the Success or Failure of Constructivist Instruction." In *Constructivist Instruction: Success or Failure?* ed. S. Tobias and T. Duffy, 335–350. New York: Routledge.

Tobias, S., and T. Duffy, eds. 2009. *Constructivist Instruction: Success or Failure?* New York: Routledge.

Tversky, A., and D. Kahneman. 1974. "Judgment under Uncertainty: Heuristics and Biases." *Science* 185:1124–1131.

Vygotsky, L. S. (1934) 1987. *The Collected Works of L. S. Vygotsky*, ed. R. Rieber and A. Carton. New York: Plenum.

Wainer, H. 2010. "14 Conversations about Three Things." *Journal of Educational and Behavioral Statistics* 35, no. 1:5–25.

Watson, G., and E. Glaser. 2002. *Watson-Glaser Critical Thinking Appraisal—UK Edition: Practice Test*. London: Pearson Assessment.

Williamson, D. M., R. J. Mislevy, and I. I. Bejar. 2006. *Automated Scoring of Complex Tasks in Computer-Based Testing*. Mahwah, NJ: Lawrence Erlbaum.

Wineburg, S. 1991. "On the Reading of Historical Texts: Notes on the Breach between School and Academy." *American Educational Research Journal* 28:495–519.

Wineburg, S. 1998. "Reading Abraham Lincoln: An Expert/Expert Study in the Interpretation of Historical Texts." *Cognitive Science* 22, no. 3:319–346.

Yackel, E., and P. Cobb. 1996. "Sociomathematical Norms, Argumentation, and Autonomy in Mathematics." *Journal for Research in Mathematics Education* 27:458–477.