SPRINGER BRIEFS IN STATISTICS

Robert J. Mislevy · Geneva Haertel
Michelle Riconscente
Daisy Wise Rutstein
Cindy Ziker

Assessing Model-Based Reasoning using Evidence-Centered Design

A Suite of Research-Based Design Patterns

Springer

SpringerBriefs in Statistics

More information about this series at http://www.springer.com/series/8921

Robert J. Mislevy · Geneva Haertel
Michelle Riconscente · Daisy Wise Rutstein
Cindy Ziker

Assessing Model-Based Reasoning using Evidence-Centered Design

A Suite of Research-Based Design Patterns

Springer

Robert J. Mislevy
Educational Testing Service
Princeton, NJ
USA

Daisy Wise Rutstein
SRI International
Menlo Park, CA
USA

Geneva Haertel
SRI International
Menlo Park, CA
USA

Cindy Ziker
SRI International
Menlo Park, CA
USA

Michelle Riconscente
Designs for Learning, Inc.
San Francisco, CA
USA

ISSN 2191-544X ISSN 2191-5458 (electronic)
SpringerBriefs in Statistics
ISBN 978-3-319-52245-6 ISBN 978-3-319-52246-3 (eBook)
DOI 10.1007/978-3-319-52246-3

Library of Congress Control Number: 2017931568

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

To Geneva: A keen researcher, a delightful collaborator, a dear friend
(1947–2016)

Preface

Model-based reasoning is essential to navigating the complexities of the real world. Though the literature on model-based reasoning has been available for many years, the advent of the Next Generation Science Standards (NGSS) has increased interest in how students can effectively develop model-based reasoning abilities. The vision of these standards originated with "The Science Framework for K-12 Education" (NRC Framework, 2012), which advocates for the deep integration of the practices of science with the understanding of core ideas and crosscutting concepts (NRC, 2014). The Framework identifies eight practices of science and engineering that are essential for all students:[1]

1. Asking questions (for science) and defining problems (for engineering);
2. Developing and using models;
3. Planning and carrying out investigations;
4. Analyzing and interpreting data;
5. Using mathematics and computational thinking;
6. Constructing explanations (for science) and designing solutions (for engineering)
7. Engaging in argument from evidence; and
8. Obtaining, evaluating, and communicating information.

Developing assessments that target these practices will require tasks that elicit evidence about how students integrate their knowledge of disciplinary core ideas, apply scientific practices, and build connections across ideas (NRC, 2012). In this manuscript, we focus on model-based reasoning as related to the NGSS practice of developing and using models. The National Science Teachers' Association (NSTA) describes developing and using models as "a practice of both science and engineering … to use and construct models as helpful tools for representing ideas and explanations. These tools include diagrams, drawings, physical replicas,

[1] *Next Generation Science Standards: For States, By States* (2013) Chapter: Appendix F: Science and Engineering Practices in the Next Generation Science Standards

mathematical representations, analogies, and computer simulations" (NSTA, 2016). The practices of defining problems and designing solutions for engineering present new opportunities to engage learners in aspects of model-based reasoning, in both instruction and assessment.

In these sections, we describe an approach to designing assessments of model-based reasoning that draws on recent developments in several areas. The first is research on science learning, and, in particular, learning to reason with and through models. A second is cognitive psychology, which highlights both cognitive and social aspects of how people, as individuals and as communities develop models and use them to solve problems and extend their knowledge. The third area is developments in technology, which enable us to build interactive simulations for students, for both learning and assessment, and to make sense of the rich data that can be captured.

The fourth area, the center of this brief, is advances in assessment theory. The first three developments mentioned above expand our knowledge base, widen our vision of what assessment can be, and give us technologies to create richer and more valid assessments. The challenge is how to effectively realize the potential of these advances in practice.

Educational assessment is itself experiencing a renaissance of sorts. Although technological developments have provided improved task environments and psychometric methods, the key development has been to recognize assessment not as a simple exercise in measurement, but as the construction of an argument: What do the particular, situated, noisy observations in a task tell us about students' understandings? About the knowledge structures and activity patterns they can marshal to address what kinds of situations? How does their performance depend on their previous experiences, and how can we sort out the meanings of complex performances in complex tasks? How can we manage task design when multiple aspects of knowledge and skill are involved, and when they may interact differently with different students? How can we design tasks that integrate the broad array of new insights about learning, about the nature of science, and about the technologies that are available for new forms of assessment?

The framework of this brief is a particular approach to these questions, called evidence-centered assessment design. We describe and illustrate a support tool for task development called an assessment design pattern. The suite of design patterns presented here integrate design issues and design choices for seven aspects of building and working with models in science and engineering: Model Formation, Model Use, Model Elaboration, Model Articulation, Model Evaluation, Model Revision, and Model-Based Inquiry. They can be used as stand-alone patterns, or in combinations for in-depth investigations. The key is to organize them as assessment arguments, to enable clear and coherent reasoning about what students know and can do. Assessment arguments provide much-needed structure for coordinating the moving parts of models, cognition, performances, and technology to render integrated and valid insights into student learning. These are tools to help us do it better.

Our goal for this work is to generate a reference that provides insights on design decisions that must be addressed to develop assessments of model based reasoning. We hope these *de-sign patterns* and the framework they are created in can support the design of model-based reasoning in NGSS-inspired assessments, and assessments to come as long as model-based reasoning remains integral to science.

Princeton, USA	Robert J. Mislevy
Menlo Park, USA	Geneva Haertel
Redwood City, USA	Michelle Riconscente
Menlo Park, USA	Daisy Wise Rutstein
Menlo Park, USA	Cindy Ziker

Acknowledgements

Design patterns were introduced in a series of collaborative projects collectively called Principled Assessment Design for Inquiry, or PADI, with SRI International as prime contractor and Geneva Haertel as Principal Investigator. This volume draws on research supported by (1) the Interagency Educational Research Initiative (IERI) under grants REC-0089122, Principled Assessment Design for Inquiry Planning Grant, and REC-0129331, PADI Implementation Grant; (2) the National Science Foundation under grant DRL-0733172, Application of Evidence-Centered Design to State Large-Scale Science Assessment; and (3) the under Grant No. R324A070035, Principled Science Assessment Design for Students with Disabilities. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation or the U.S. Department of Education.

Material in this publication is based upon work supported by the United States Government under Contract No. W911NF-14-2-0062. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the U.S. Government. The authors would like to thank the following for their contributions to the works described in this brief: Alex Kernbaum, Karen Koenig, Maria-Isabel Carnasciali, Srujal Patel, Ian Chen, Nate Koenig and the Open Source Robotics Foundation.

Contents

About the Authors

Robert J. Mislevy holds the Frederic M. Lord Chair in Measurement and Statistics at Educational Testing Service, and is Professor Emeritus of Measurement, Statistics, and Evaluation at the University of Maryland. He earned his Ph.D. at the University of Chicago in 1981. His research applies developments in statistics, technology, and cognitive science to practical problems in educational assessment. His work includes a multiple-imputation approach for integrating sampling and test-theory models in the National Assessment of Educational Progress (NAEP), an evidence-centered framework for assessment design, and simulation- and game-based assessments, in collaboration with GlassLAB, the Cisco Networking Academy, and ETS colleagues.

Dr. Mislevy has received NCME's Award for Career Contributions, AERA's E.F. Lindquist Award for contributions to educational assessment, the International Language Testing Association's Messick Lecture Award, AERA Division D's Robert L. Linn Distinguished Address Award, and the NCME's Award for Technical Contributions to Educational Measurement three times. He is a member of the National Academy of Education and a past president of the Psychometric Society. His publications include *Bayesian Psychometric Methods* (Chapman & Hall/CRC, 2016) with Roy Levy, *Bayesian Networks in Educational Assessment* (Springer, 2015) with Russell Almond, Linda Steinberg, Duanli Yan, and David Williamson, and the chapter on cognitive psychology in the fourth edition of *Educational Measurement*.

Geneva Haertel, a pioneer in the application of Evidence-Centered Design (ECD) in large-scale assessment and the practical application of using assessments to improve student learning, wrote publications that focus on the development of valid and reliable assessments for use with a wide range of students, including students with special needs. She published over 60 articles and book chapters on assessment design, student learning, and conditions promoting student achievement. Her research explored the inferential validity, reliability and effectiveness of formative assessments embedded within games. From 2003 to 2008, Dr. Haertel was the SRI Principal Investigator of a series of projects funded by the National Science Foundation, the Institute of Education Sciences, and the Gates and

MacArthur Foundations, that resulted in the development of the online assessment design system referred to as PADI (Principled Assessment Designs for Inquiry).

Dr. Haertel earned her Ph.D. in Education/Educational Psychology at Kent State University and a Bachelor's of Science degree in Education from Kent State University. As the Director of Assessment Research and Design at the Center for Technology in Learning, Dr. Haertel was named a Fellow at SRI International in 2015.

Michelle Riconscente is President of Designs for Learning, a consultancy firm at the intersection of learning, assessment, and digital interactive design. An expert in evidence-centered design, she leads and advises innovative projects that run the spectrum from interactive learning design and embedded assessments, to independent research and evaluation. Previously, as Director of Learning and Assessment at GlassLab, she provided strategic leadership that resulted in valid and reliable game-based assessments of complex competencies such as proportional reasoning, problem solving, and argumentation. The author of over 100 publications and presentations, her clients include Michelson Runway, Digital Promise, Harvard University, MIT Education Arcade, Motion Math, UCLA's CRESST, and Teaching Channel. She holds a Ph.D. in Educational Psychology from the University of Maryland, College Park and a bachelor's degree in Mathematics-Physics from Brown University.

Daisy Wise Rutstein is Education Researcher in SRI International's Center for Technology in Learning. Dr. Rutstein's work focuses on the application of Evidence-Centered Design (ECD) to develop assessments. She moves through the different stages of the development process, starting with the initial conception of the assessment and moving through the development of items, the creation of a complete assessment, and the validation of the assessment. Dr. Rutstein's work in this area has included the development of design patterns, scenarios, and items as well as the identification and application of measurement models for these tasks.

Dr. Rutstein is the lead developer of assessments on several projects in computer science, mathematics, and science including the lead developer for Computational Thinking for the ECS curriculum under the PACT grants (NSF awards CNS-1433065 and DRL-1418149). Dr. Rutstein received her Ph.D. in measurement, statistics, and evaluation from the University of Maryland.

Cindy Ziker is Senior Researcher in the Center for Technology in Learning at SRI International. She earned her Ph.D. in Educational Psychology at Arizona State University and a Graduate Certificate in Large-Scale Assessment from the University of Maryland. Her research focuses on the application of evidence-centered design to the development of technology-enhanced performance tasks and the use of simulation as applied to assessment. She leads validity and evaluation studies of formative and summative assessment systems for state departments of education and large urban school districts. From 2003 to 2006, Dr. Ziker served on a Technical Work Group that evaluated the National Assessment of Educational Progress for Congress. Her publications include "An Introduction to the Evaluation of NAEP", with Suzanne Lane, Bruno D. Zumbo, Jamal Abedi, Jeri