Table of Contents

Preface iv

Author Biographies vi

Acknowledgement xii

CHAPTER 1

Introduction to Rasch Measurement in Science Education 1
Xiufeng Liu and William Boone

CHAPTER 2

Constructing a Quality Assessment through Rasch Techniques: The Process of Measurement, Feedback, Reflection and Change 23
Kelly D. Bradley and Shannon O. Sampson

CHAPTER 3

The Geoscience Concept Inventory: Application of Rasch Analysis to Concept Inventory Development in Higher Education 45
Julie C. Libarkin and Steven W. Anderson

CHAPTER 4

Tracking Cognitive Development with the Rasch Model: Empirical Evidence of Growth and Heterogeneity 74
Lorna C. Endler and Trevor G. Bond

CHAPTER 5

Stage-Like Patterns in the Development of Conceptions of Energy 111
Theo L. Dawson-Tunik

CHAPTER 6

Exploring Conceptual Understandings of Diffusion and Osmosis by Senior High School and Undergraduate University Science Students 137
Debra Panizzon and Trevor Bond

CHAPTER 7

Mapping Out Students' Matter Concept Development from Elementary to High School 165
Xiufeng Liu

CHAPTER 8

Using Structured Item Response Theory Models to Analyze Content and Inquiry Reasoning Skills in BioKIDS 188
Han Bao, Amelia Wenk Gotwals, Nancy Butler Songer, and Robert J. Mislevy

Chapter 9

Assessing Students' Level of Knowledge and Analysing the Reasons for Learning Difficulties in Physics by Rasch Analysis 212
Alexander Kauertz and Hans E. Fischer

Chapter 10

Interpreting and Using Multidimensional Performance Data to Improve Learning in Science 247
Cathleen A. Kennedy and Karen Draney

Chapter 11

Is the Paper and Pencil Assessment Mode Appropriate for Assessing the Learning Outcomes of Primary Science? Teachers' Attitudes 285
Iris Lee and Guanzhong Luo

Chapter 12

Increasing Students' Interest: Informal Learning in Authentic Science Labs 315
Katrin Engeln and Jürgen Rost

Author Index ..... ???

Subject Index ..... ???

Preface

Assessment is of paramount importance in any educational system. In recent years, there have been important, highly visible international comparisons of students' progress in science and mathematics at different grade levels through the Trends in Mathematics and Science Studies (TIMSS) and the Programme for International Student Assessment (PISA). These international studies use different kinds of items to compare students in science and mathematics from an increasing number of nations. The results are published not only in scholarly journals and reports but also in the national and international press. As might be expected, educators, educational managers and politicians use these comparisons both within and between nations for different purposes to meet their own agendas. In addition, the findings of these reports have spawned much additional research on measurement of student progress.

At another level of the educational spectrum, in many countries much attention is given to testing to meet proficiency levels or standards or guidelines of attainment at selected years of schooling. This happens to such an extent that there can be financial penalties for those schools not meeting student attainments. So to ensure adequate test scores, teaching for success on the tests becomes a major goal of teaching.

For several years, I have been designing school-based pencil-and-paper tests that can be used by teachers and students to diagnose learning outcomes in selected content domains in physics, biology and chemistry. Similar research has been conducted by other science education colleagues, all with the intended outcomes of producing tests that enable more successful learning and teaching based on feedback from such tests.

The points of departure from these three levels of measurement—international, national, and school—is that educational measurement in science and mathematics education is ubiquitous and the consequences of these measurements are made visible in a wide range of outlets, including local newspapers, such that citizens outside of education are informed about and talk about these results. For researchers working in areas of

measurement, it is important to understand the background behind the measures used in TIMSS and PISA, as well as State tests. For interested non-measurement colleagues reading reports, for example, about TIMSS and PISA, it is usually not evident that the data analyses reported are from Rasch measures. For those who wish to learn more about these kinds of analyses in the context of science education research, the edited text by Xiufeng Liu and William Boone makes a very timely and valuable contribution to the literature.

A key issue raised by Liu and Boone is to bring to our attention what reported research in science education is indeed measuring. Certainly large scale international studies utilise sophisticated sampling and measurement testing techniques but this is not always so in smaller scale studies such as those I have reported about my own research. Many science educators will know the basic underpinnings of Rasch models of measurement. However, I suspect many science educators are like me in that they have not used these measures in their research studies and have limited comprehension of how such measures are developed. This edited volume is most informative because the various authors of the chapters provide a range of approaches to assessment using Rasch measures in science education that can guide and assist others in conducting research using more robust and challenging measurement tools as well as reporting findings in a comprehendible manner. Certainly, this edited book with a wide variety of topics in different content domains has whet my appetite to consider how to apply Rasch models of measurement to my own data collected on diagnostic tests. I commend the book to those interested in learning more about approaches to measurement of outcomes in science education using Rasch models.

David F. Treagust
Curtin University of Technology
Perth Australia

Author Biographies

Steven W. Anderson is a Professor of Geology and Planetary Science at Black Hills State University (SD), and an Associate Research Scientist at the Planetary Science Institute (AZ). His research interests include terrestrial and planetary lava flow surface morphology, the origin of alternative geoscience conceptions, and the development of valid and reliable assessment instruments in the geosciences. In addition to his faculty position at Black Hills State University, he has had visiting faculty appointments at Arizona State University, the University of Arizona, and the University of Pittsburgh. Steven received his B.S. from Cornell College in 1985, M.S. from Arizona State in 1988, and Ph.D. from Arizona State in 1990.

Han Bao is a doctoral candidate in the Department of Measurement, Statistics, and Evaluation at the University of Maryland. Her research focuses on measurement and statistical applications in education.

Trevor G. Bond is Head of the Department of Educational Psychology and Learning Needs at the Hong Kong Institute of Education. He is internationally regarded for his work on the application of Rasch analysis to developmental variables and educational outcomes. His best-selling current book (co-authored with Christine Fox), *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*, reflects his long-standing experience in educational settings as the background for applying fundamental measurement to outcomes in the human sciences. He is Program Chair of the Rasch Measurement Special Interest Group of the American Educational Research Association, Chair of the Assessment and Measurement Special Interest Group of the Australian Association for Research in Education, on the editorial board of the *Journal of Applied Measurement* and the *Australian Educational Researcher* and is a regular presenter on topics of educational measurement in the U. S. and Europe, in particular.

William Boone is the Condit Professor of Science Education at Miami University. He is also currently Co-Director of Ohio's Evaluation and Assessment Center for Mathematics and Science Education. For over a decade he served as a faculty member at Indiana University-Bloomington. In 1991 he received his Ph.D. from the University of Chicago's MESA program. At the University of Chicago Ben Wright was his thesis director. Prior to entering the MESA program Bill was a high school physics and physical science instructor. He also earned a B.S. in Geology from Indiana University Bloomington, and a M.S. in Geophysics from the University of Wisconsin-Madison. His primary research interest concerns how the Rasch model might be optimally used to develop and maintain measurement instruments of use in science and mathematics education efforts. An additional interest concerns improving the participation and performance of under-represented groups in science and mathematics. He has (or currently serves) on a number of editorial boards (*Journal of Research in Science Teaching*, *Science Education*, *Journal of Science Teacher Education*). Although science education is his main area of work, he served as the measurement specialist for the research team evaluating the use of vouchers in Cleveland, Ohio.

Kelly D. Bradley is an Assistant Professor in the Department of Educational Policy Studies and Evaluation at the University of Kentucky (UK) where she teaches research methods and statistics courses. She came to UK directly from doctoral study in quantitative research, evaluation, and measurement at The Ohio State University. Her research is anchored in quantitative research and measurement, with a focus on Survey Research and Rasch analysis. Dr. Bradley holds a M.S. in statistics from the University of South Carolina, a B.S. in mathematics and sociology and a B.A. in mathematics education from Fairmont State College, with certification in grades 5-12.

Karen Draney is a Project Manager for the Berkeley Evaluation and Assessment Research (BEAR) Center in the Graduate School of Education at University of California, Berkeley. Her research focuses on application of the BEAR Assessment System to classroom assessment practice, with an emphasis on teachers' use of assessment data.

Hans E. Fischer is Professor of Physics Education and head of the research group "Teaching and Learning Science" at the University of Duisburg-Essen (Germany). He teaches physics and physics education. His research is anchored in quantitative research and physics education in general with a focus on teaching methods and learning processes. Dr. Fischer holds a first and second teacher exam in mathematics, physics and pedagogy for grades 5-13 of German high school (Gymnasium).

Lorna C. Endler supervises secondary science Credential Candidates in the Teacher Education Program and mentors undergraduates in the Community Mathematics and Science Teaching Initiative at the University of California Santa Barbara. Her Ph.D. and Master's degrees, awarded by James Cook University, Australia, focused on cognitive development and student achievement, specifically of children who experienced the Cognitive Acceleration through Science Education (CASE) program. She taught Biology for 30 years in a variety of schools and colleges in England and Australia before moving to California in 1998.

Katrin Engeln, after graduating in physics, became a mathematics and physics teacher at a senior high school. Since 2001 she has been working at the IPN. Her main research interests are informal learning and the development of students' interest.

Amelia Wenk Gotwals is a doctoral candidate in science education at the University of Michigan. Her research focuses on the development and analysis of assessments for measuring students' scientific content and inquiry reasoning abilities.

Alexander Kauert is holder of a scholarship financed by the German Research Foundation (DFG) and is member of the graduate school, "teaching and learning of science". He finished his teacher exam in 2004 in mathematics and physics at the University of Dortmund (Germany). Currently he is in the second year of his doctoral study at the University of Duisburg-Essen (Germany). His quantitative research deals with content structure, learning processes and tasks in physics education.

Cathleen A. Kennedy is the Director of Technology for the Berkeley Evaluation and Assessment Research (BEAR) Center in the Graduate School of Education at University of California, Berkeley. Her research focuses on the use of technology to advance assessment practice, particularly in the application of multidimensional item response modeling.

Iris Lee is a Ph.D. candidate with the School of Education, Murdoch University. She is currently granted professional development leave from the Ministry of Education (Singapore). Her area of interest lies with Science Education, Assessment and Measurement.

Julie C. Libarkin is an Assistant Professor of Geosciences at Ohio University. Her research encompasses both traditional geology and geoscience education. In particular, Julie is interested in understanding conceptual change in college science classrooms, works on the development of valid and reliable assessment instruments, and is intrigued by the mechanisms responsible for the uplift of high elevation plateaus. Prior to starting her position at Ohio University in 2003, she completed an NSF Postdoctoral Fellowship in Science, Mathematics, Engineering, and Technology Education at the University of Arizona and Harvard-Smithsonian Center for Astrophysics. Julie received a B.S. in Geology and Physics from the College of William and Mary in 1994 and a Ph.D. in Geosciences from the University of Arizona in 1999.

Xiufeng Liu is an Associate Professor of Science Education at the State University of New York, Buffalo. Before taking his current position, he was an assistant and associate professor in science education at St. Francis Xavier University and the University of Prince Edward Island, both in Canada, for 10 years. He received Ph.D. in Science Education in 1993 from the University of British Columbia with dissertation research in Item Response Theory (IRT) models. As a former high school chemistry teacher and now a science educator, he has focused his research on measuring and promoting students' conceptual change in major science concepts from elementary to high school. He has also been doing research related to science curriculum standards and accountability as well as technology-based science teaching and assessment.

Guanzhong Luo is currently heading the research and development division of the Hong Kong Examinations and Assessment Authority while holding two other portfolios concurrently. He remains adjunct professor to Murdoch University and fulfills his professorship at South China Normal University. Psychometrics and assessment reforms are his two major research interests.

Robert J. Mislevy is a Professor of Measurement, Statistics, and Evaluation at the University of Maryland. Previously, Dr. Mislevy was a Distinguished Research Scientist at Educational Testing Service. His research interests apply developments in technology, statistical methods, and cognitive research to practical problems in educational and psychological measurement.

Debra Panizzon is a Senior Lecturer in Science Education in the School of Education at the University of New England (UNE), Australia. Currently, she is Deputy Director for the National Centre of Science, Information and Communication Technology, and Mathematics Education for Rural and Regional Australia (SiMERR National Centre). Her research interests are in the areas of cognition and concept acquisition in secondary and tertiary students. She also works with developmental-based assessment and the ways in which teachers and academics use assessment to improve instruction.

Jürgen Rost is the head of the department of pedagogical and psychological methods at the IPN. His research focuses on the Item Response Theory. He is a member of the German national PISA 2003 and 2006 consortium.

Shannon O. Sampson is a research associate in the Department of Educational Policy Studies and Evaluation at the University of Kentucky, where she completed her Ph.D. in 2005. Her research interests are second language acquisition, assessment of linguistically diverse students, and applications of Rasch measurement. She holds a M.A. in Spanish Literature and Linguistics from Florida State University, and a B.A. in Spanish with a minor in mathematics from Georgetown College (Summa Cum Laude).

Nancy Butler Songer is a Professor of Science Education and Learning Technologies at the University of Michigan. Focusing on students in high-poverty urban settings, Dr. Songer's research characterizes higher-order thinking in science, evaluates simple technologies used as cognitive tools, and develops assessment systems to provide developmental evidence of complex reasoning.

Theo L. Dawson-Tunik received her Ph.D. from the University of California at Berkeley in 1998. She is the President and CEO of the Developmental Testing Service, LLC and a Visiting Assistant Professor at Hampshire College. Dawson-Tunik has designed a novel approach to investigating, promoting, and evaluating conceptual development, called *developmental maieutics*. This methodology is currently being employed to improve instruction and learning in K-12, college, and workplace contexts.

Acknowledgements

This book would not have come into being without the support from Dr. Richard Smith, the JAM Press. Richard's passion for promoting good measurement in education and social sciences and particularly his belief in the value of this book will greatly benefit the science education community.

We are also indebted to chapter contributors of this book who come from around the globe, and share our passion for Rasch measurement and belief in quality measurement in science education. We also acknowledge the contribution of two reviewers, Dr. Linjun Shen of the National Board of Osteopathic Medical Examiners and Dr. Jaekyung Lee of State University of New York at Buffalo, who provided helpful comments and suggestions to select chapters.

Co-Editors,

Xiufeng Liu, Ph.D.
State University of New York at Buffalo

William J. Boone, Ph.D.
Miami University in Ohio