



EDITED BY
CARLA M. EVANS
AND CATHERINE S. TAYLOR

CULTURALLY RESPONSIVE ASSESSMENT IN CLASSROOMS AND LARGE-SCALE CONTEXTS

Theory, Research, and Practice

APPLICATIONS OF
EDUCATIONAL MEASUREMENT
AND ASSESSMENT

NCME

national
council on
measurement
in education

ROUTLEDGE

The Routledge logo, which is a stylized white profile of a person's head facing right, set against a black background.

Culturally Responsive Assessment in Classrooms and Large-Scale Contexts

Culturally Responsive Assessment in Classrooms and Large-Scale Contexts explores how scholars and professionals in educational measurement and assessment can use the unique cultural and social identities of students to shape assessment purpose, design, implementation, use, and validation processes. Despite the sheer diversity of student populations in the United States, the tools used to understand their performance and progress have not substantively changed in decades. Large-scale testing and related policies still privilege “culturally neutral” test content, standardization, and comparability. Classroom assessments often mimic these procedures, even though rich tasks and activities could easily be incorporated into curriculum and instruction to allow students to see themselves, their interests, communities, beliefs, and backgrounds represented and valued. This book collects the theory, research, and best practices that are essential to ensuring cultural responsiveness in classroom assessment and standardized testing procedures, policies, and practices. Chapters by assessment and measurement experts, along with scholars who are experts in sociocultural learning theory, philosophy of education, critical theories, multilingual education, and Indigenous education, summarize findings from existing research and suggest how future research can move classroom and industry practice as well as federal, state, and local policy forward.

Carla M. Evans is Senior Associate at The National Center for the Improvement of Educational Assessment and Adjunct Professor of Education at the University of New Hampshire, USA.

Catherine S. Taylor is Professor Emerita at the University of Washington, USA, and was Senior Vice President for Measurement Services at Measured Progress.

The NCME Applications of Educational Measurement and Assessment Book Series

Editorial Board:

Kadriye Ercikan, Educational Testing Service, Editor
Henry Braun, Boston College
Brian Clauser, National Board of Medical Examiners
Neil J Dorans, Educational Testing Service
Rochelle Michel, Smarter Balanced
Fritz Drasgow, University of Illinois at Urbana-Champaign
Melissa Margolis, National Board of Medical Examiners
Guillermo Solano-Flores, Stanford Graduate School of Education
Roy Levy, Arizona State University

Fairness in Educational Assessment and Measurement

Edited by Neil J. Dorans and Linda L. Cook

Testing in the Professions: Credentialing Policies and Practice

Edited by Susan Davis-Becker and Chad W. Buckendahl

Validation of Score Meaning for the Next Generation of Assessments: The Use of Response Processes

Edited by Kadriye Ercikan and James W. Pellegrino

Preparing Students for College and Careers: Theory, Measurement, and Educational Practice

Edited by Katie Larsen McClarty, Krista D. Mattern, and Matthew N. Gaertner

Score Reporting Research and Applications

Edited by Diego Zapata-Rivera

Classroom Assessment and Educational Measurement

Edited by Susan M. Brookhart and James H. McMillan

Integrating Timing Considerations to Improve Testing Practices

Edited by Melissa J. Margolis and Richard A. Feinberg

Advancing Natural Language Processing in Educational Assessment

Edited by Victoria Yaneva and Matthias von Davier

Culturally Responsive Assessment in Classrooms and Large-Scale Contexts: Theory, Research, and Practice

Edited by Carla M. Evans and Catherine S. Taylor

For more information about this series, please visit: <https://www.routledge.com/NCME-APPLICATIONS-OF-EDUCATIONAL-MEASUREMENT-AND-ASSESSMENT/book-series/NCME>

Culturally Responsive Assessment in Classrooms and Large-Scale Contexts

Theory, Research, and Practice

Edited by Carla M. Evans and Catherine S. Taylor

Designed cover image: © Getty Images

First published 2025

by Routledge

605 Third Avenue, New York, NY 10158

and by Routledge

4 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

Routledge is an imprint of the Taylor & Francis Group, an informa business

© 2025 selection and editorial matter, Carla M. Evans and Catherine S.

Taylor; individual chapters, the contributors

The right of Carla M. Evans and Catherine S. Taylor to be identified as the authors of the editorial material, and of the authors for their individual chapters, has been asserted in accordance with sections 77 and 78 of the Copyright, Designs and Patents Act 1988.

The Open Access version of this book, available at www.taylorfrancis.com, has been made available under a Creative Commons Attribution-Non Commercial-No Derivatives (CC-BY-NC-ND) 4.0 license.

Any third party material in this book is not included in the OA Creative Commons license, unless indicated otherwise in a credit line to the material. Please direct any permissions enquiries to the original rightsholder.

Trademark notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

ISBN: 978-1-032-50490-2 (hbk)

ISBN: 978-1-032-50491-9 (pbk)

ISBN: 978-1-003-39221-7 (ebk)

DOI: 10.4324/9781003392217

Typeset in Minion Pro

by Apex CoVantage, LLC

For his thought leadership and lifetime of service to the field of education, assessment, and human flourishing, we dedicate this book to Dr. Edmund Gordon.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Contents

Editor and Contributor Biographies	x
Acknowledgments	xviii
1 Reimagining the Future of Assessment that Honors Human Diversity	1
CARLA M. EVANS AND CATHERINE S. TAYLOR	
Section 1: Culturally Responsive Assessment: Theoretical Foundations and Concepts	15
2 Building from Sociocultural Learning Theory to Culturally Responsive Assessment	17
ANDREW D. COPPENS AND REBECCA KELLEY	
3 Student Engagement, Motivation, and Assessment: A Critical Triad Requiring Critical Reform	34
MICHAEL NAKKULA AND ERIC TOSHALIS	
4 Voice, Recognition, and Exit: A Rights-Based Response to Cultural Bias in Educational Assessments	52
WINSTON C. THOMPSON AND JOY DANGORA ERICKSON	
5 Using a QuantCrit Lens to Analyze Data from a Culturally Responsive Assessment	70
QUINTIN LOVE	
6 A Framework for Enacting Equity Aims in Assessment Use: A Justice-Oriented Approach	88
SUSAN LYONS, MARIA ELENA OLIVERI, AND MYA POE	
7 Validity Argumentation for Culturally Responsive Assessments	106
SUZANNE LANE AND SCOTT F. MARION	

8	Building the Conceptual Foundations of Culturally Responsive Assessment: An Analysis of a Sample of Conceptual Pieces Across Multiple Assessment Process Components—Commentary on Section 1	124
	GUILLERMO SOLANO-FLORES	
Section 2: Implications of Culturally Responsive Assessment for Large-Scale Assessment Practices		133
9	A Measurement Argument for Culturally Responsive Assessment	135
	CATHERINE S. TAYLOR AND STEVE FERRARA	
10	Reconsidering Theories of Action: Using Cultural and Community Validity to Transform Assessment Development, Implementation, and Use	156
	KERRY ENGLERT AND PŌHAI KŪKEA SHULTZ	
11	Working Toward Culturally Sustaining Workplace Assessment Programs Using an Integrated Design and Appraisal Framework	177
	DAVID SLOMP, MARIA ELENA OLIVERI, AND MYA POE	
12	Antiracist Approaches to Scoring Large-Scale Assessments	202
	LAUREN WHITE, JAYLIN NESBITT, HEATHER ROETERS-SOLANO, SARAH QUESEN, SUE LOTTRIDGE, AND KAREN LOCHBAUM	
13	On Objectivity, Multicultural Education, and Large-Scale Assessments—Commentary on Section 2	222
	WAYNE AU	
14	An Evidentiary-Reasoning Perspective on Culturally Responsive Assessment—Commentary on Section 2	228
	ROBERT J. MISLEVY	
Section 3: Implications of Culturally Responsive Assessment for Classroom Assessment Practices		243
15	Implications of Culturally Responsive Education for Classroom Assessment Practices	245
	KYNDRA V. MIDDLETON AND JANINE A. HAYNES	
16	Manifesting the Mana of Culturally Responsive Assessments	264
	CHELSEA N. K. KEEHNE, IVY K. KELLING, AND KUULEIANUHEA AWO-CHUN	

17	How Culturally Responsive Formative Assessment Practices Can Expand Students' Opportunities to Participate in Science Learning ERIN MARIE FURTAK AND CLARISSA DEVEREL-RICO	284
18	Equitable and Culturally Relevant Grading Practices SUSAN M. BROOKHART AND DWAYNE CHISM	306
19	Teachers' <i>Mis</i> -Assessment of Black Children's Behavior: What's Culture Got to Do With It? MONICA R. BROWN, MONIQUE MATUTE-CHAVARRIA, AND PRICELLA MORRIS	324
20	Reflections on Some Current Approaches to Culturally Responsive Assessment—Commentary on Section 3 ELISE TRUMBULL AND SHARON NELSON-BARBER	337
21	From “But That’s Just Good Teaching” to “An Ethical Imperative”: The Role of Culturally Responsive Practices in Classroom Assessment—Commentary on Section 3 ELENA DIAZ-BILELLO	349
	Conclusion	357
22	Culturally Responsive Assessment: Lessons Learned So Far CATHERINE S. TAYLOR AND CARLA M. EVANS	359
	Index	381

Editor and Contributor Biographies

Editors

Carla M. Evans, Ph.D., is Senior Associate at the National Center for the Improvement of Educational Assessment. She helps state education agencies design and implement assessment and accountability reforms. Her research focuses on the impacts and implementation of assessment and accountability policies on teaching and learning. She is especially interested in policy research related to balanced assessment systems, culturally responsive assessment, performance-based assessments, and assessment literacy. Dr. Evans has recently co-authored a book, *Understanding Instructionally Useful Assessment*, and has published numerous articles in peer-reviewed journals. She also currently serves as an adjunct professor of education at the University of New Hampshire.

Catherine S. Taylor, Ph.D., is Professor Emerita from the University of Washington (UW)-Seattle. While at UW, Dr. Taylor took two leaves to support Washington State's testing program, first as Interim Director of Assessment and later as Director of Assessment Alternatives. After retirement from UW, Dr. Taylor served as Senior Vice President for Measurement Services at Measured Progress. Dr. Taylor conducts research on validity and bias in large-scale testing. She has published in leading journals such as *American Educational Research Journal* and *Applied Measurement in Education*. Dr. Taylor's latest book is *Culturally and Socially Responsible Assessment* published by Teachers College Press.

Chapter Author Biographies

Wayne Au, Ph.D., is currently Dean and Professor in the School of Educational Studies at the University of Washington Bothell. Dr. Au's public and academic scholarship focuses on high-stakes testing, teaching for social justice, critical pedagogy, anti-racist education, and Asian American education, and he is a long-time editor for the social justice teaching magazine *Rethinking Schools*. His recent books include *Unequal by Design: High-Stakes Testing and the Standardization of Inequality* (2nd ed), *Rethinking Multicultural Education* (3rd ed.), and the upcoming *Asian American Racialization and the Politics of U.S. Education*.

Kuuleianuhea Awo-Chun, M.Ed., is from Ko'olaupoko, O'ahu, and currently serves as Assistant School Director of Mālama Honua PCS in Waimānalo. With over 20 years of work in Hawaiian-focused Charter Schools, her experience includes teaching English Language Arts and leading integrated aloha 'āina-based curriculum and assessment with

the wa'a and community partner Kānehūnāmoku. She earned her BA in English, Creative Writing from the University of Puget Sound in Tacoma, WA, and a Masters Degree in teaching from the University of Hawaii at Mānoa's MedT program. She is Native Hawaiian and a mother of three.

Susan M. Brookhart, Ph.D., is Professor Emerita at the Duquesne University School of Education. Dr. Brookhart's interests include the role of both formative and summative classroom assessment in student motivation and achievement, the connection between classroom assessment and large-scale assessment, and grading. Dr. Brookhart was the 2007–2009 editor of *Educational Measurement: Issues and Practice*, a journal of the National Council on Measurement in Education. She received the 2014 Jason Millman Award from the Consortium for Research on Educational Assessment and Teaching Effectiveness (CREATE) and the 2015 Samuel J. Messick Memorial Lecture Award from Educational Testing Service.

Monica R. Brown, Ph.D., is Professor of Special Education in the Department of Early Childhood, Multilingual and Special Education at the University of Nevada, Las Vegas (UNLV). Dr. Brown's research and other scholarship centers on the lived experiences of Black and other racially minoritized children, families, and communities (especially those that include individuals with differing abilities). She utilizes critical frameworks to analyze special education as it intersects with race.

Dwayne Chism, Ed.D., is former Dean of the School of Education at Peru State College. He teaches at the University of Nebraska in Omaha, which honored him with the Distinguished Alumni Award in 2013–2014. Dr. Chism founded Shifting Perspectives, Inc. in 2018 as a company dedicated to growing leaders and helping schools reach equitable outcomes for all learners. He has authored several articles for the Association for Supervision and Curriculum Development (ASCD), including a book in 2022, *Leading Your School Toward Equity*, a finalist for the Excellence in Equity Award presented by the American Consortium for Equity in Education.

Andrew D. Coppens, Ph.D., is Associate Professor of Education in Learning Sciences at the University of New Hampshire. His developmental and cultural psychological work examines informal learning and development among nondominant children, youth, families, and communities, involving rural, middle-class, and Indigenous-heritage communities in the US, Mexico, Ecuador, Germany, India, and Bhutan. Dr. Coppens is a founding PI in the Youth Retention Initiative (YRI), a research collaborative focused on sustainable and strengths-based educational and workforce pathways among rural youth. YRI research has been funded by UNH, the Spencer Foundation, the National Science Foundation, and the New Hampshire Charitable Foundation.

Clarissa Deverel-Rico, Ph.D, is a postdoctoral researcher at BSCS Science Learning, where she is studying the extent to which currently available classroom assessments support the vision of *Framework*-era reforms. She is also developing assessments for broad use in high school biology. Through other studies, she has partnered with teachers and created space to learn directly from students and families how classroom assessment can be spaces that sustain students' interests and identities. Dr. Deverel-Rico is committed to studying innovative approaches for how classroom assessment can support a vision of science education that prioritizes epistemic justice, agency, and student experience.

Elena Diaz-Bilello, Ph.D., is Associate Director of the Center for Assessment, Design, Research, and Evaluation (CADRE) and a Faculty Affiliate of the University of Colorado Boulder School of Education. She collaborates with state education agencies, school districts, and educational organizations to develop practical and sound approaches to address assessment and educational policy challenges. Her work includes conducting mixed-methods research and program evaluations to study the effectiveness of educational reforms and initiatives intended to provide equitable learning and assessment experiences, and to improve teaching and learning. Dr. Diaz-Bilello also serves as an assessment and accountability technical advisory committee member in several states.

Kerry Englert, Ph.D., is President of Seneca Consulting, a small, woman-owned business that provides psychometrics, program evaluation, and data collection and analysis services. She serves as the lead psychometrician for the Hawai'i Department of Education's Native language assessment project (KĀ'EO). In this role, she focuses on the technical aspects of the assessment development process and ensures the integration of the priorities of the Hawaiian language community. Dr. Englert has also partnered with Marzano Research, where she has led projects with state Indian education directors to develop culturally relevant needs assessments that reflect the priorities of Native communities.

Joy Dangora Erickson, Ph.D., is Assistant Professor of education at Endicott College. She received the Student Outstanding Research Award from the Literacy Research Association in 2018 for her work examining young children's reading motivation. In addition to exploring children's motivation to read within intervention programs, Dr. Erickson's scholarship centers on issues of early education for citizenship, including culturally responsive and sustaining pedagogies. Her work has been featured in numerous journals, including *The Reading Teacher*; *Literacy Research: Theory, Method, and Practice*; *Reading & Writing Quarterly: Overcoming Learning Difficulties*; *Journal of Early Childhood Literacy*; *Democracy & Education*; and *Journal of Curriculum Studies*.

Steve Ferrara, Ph.D., had decades of experience conducting research and providing leadership in educational assessment. Early in his career, he was Director of Assessment for the State of Maryland. Later, he served in measurement leadership positions for American Institutes for Research, CTB/McGraw-Hill, Pearson, Cognia, and HumRRO. His research focused on validity issues in large-scale and classroom assessments. He published over 50 chapters and articles. His work has been published in leading journals, including *Journal of Educational Measurement* and *Applied Measurement in Education*. His most recent book is *Assessing Student Learning by Design* with Jay McTighe.

Erin Marie Furtak, Ph.D., is Professor of STEM Education at the University of Colorado at Boulder. A former high school science teacher, Erin transitioned into a career studying how science teachers learn and improve their daily classroom practices through formative assessment. In a series of multiple studies, Dr. Furtak has been partnering with teachers, schools, and districts to learn how teachers can design, enact, and take instructional action on the basis of classroom assessments that they design. She also explores the ways in which the design and enactment of classroom assessments can promote more equitable participation in science learning.

Janine A. Haynes, Ph.D., is a graduate of psychometrics at Morgan State University. With a bachelor's degree in English and a master's degree in Curriculum and Instruction, Dr. Haynes combines her robust academic background with extensive practical experience to advance the

field of educational measurement. Her professional journey includes over 16 years of K–12 classroom experience in Maryland and Washington, D.C., as well as a significant tenure overseas in the United Arab Emirates. Her international experience enriches her research, particularly concerning measurement invariance across large-scale educational assessments and the impact of sociocultural factors on classroom assessment practices.

Chelsea N. K. Keehne, Ed.D., was born and raised in Hilo, Hawai‘i, and obtained her undergraduate degree from the University of Hawai‘i at Hilo and EdD from the University of Hawai‘i at Mānoa. The first decade of Dr. Keehne’s career was as a Hawai‘i DOE teacher in predominately Native Hawaiian communities. Her current role in the Kamehameha Schools Strategy and Experience Division includes supporting Hawaiian-focused Charter Schools to develop culturally relevant assessment systems that provide a counternarrative to the conventional assessment dialogue. She is a Native Hawaiian action researcher who is deeply committed to perpetuating ‘ike kupuna (ancestral knowledge).

Rebecca Kelley is a Ph.D. candidate in Curriculum and Instruction at the University of New Hampshire. Her research examines how teachers design and use classroom assessment, focusing particularly on dynamic and formative assessment in world language classrooms. She is a former middle school French teacher and also serves as the Performance Assessment Coach with the Massachusetts Consortium for Innovative Education Assessment (MCIEA).

Ivy K. Kelling, Ed.D., was born and raised in Makakilo, Honouliuli, ‘Ewa, O‘ahu, and has served in various capacities in Hawaiian Medium Education and the Hawaiian Language revitalization movement over the past 27 years. She currently serves as the Director of Ke Kula ‘o Samuel M Kamakau, Laboratory Public Charter School, where she started as a lower elementary school teacher. Dr. Kelling has earned a BA in Hawaiian language from the University of Hawai‘i at Manoa, an M.Ed. in Early Childhood/Elementary Education from Chaminade University, and an Ed.D. from the University of Southern California.

Suzanne Lane, Ph.D., is Professor Emeritus in the Research Methodology Program at the University of Pittsburgh. Dr. Lane’s research and publications are in educational measurement, with a focus on design, validity, and equity issues. She is a co-author (with Scott Marion) of the Validity chapter in *Educational Measurement* (in press). She was President of NCME (2003–2004), Vice President of Division D of AERA (2000–2002), member of the AERA, APA, and NCME Joint Committee for the Revision of the *Standards for Educational and Psychological Testing* (1993–1999), and appointed to NAGB (sets policy for NAEP (2020–2024)). She has served on numerous technical advisory committees for assessment programs.

Karen Lochbaum, Ph.D., was Vice President of Automated Scoring at Pearson, where she was responsible for the development and delivery of Pearson’s automated writing and math scoring services, including the Intelligent Essay Assessor™ (IEA), Continuous Flow, and the Math Reasoning Engine. She has over 20 years of experience in automated scoring, including design, development, research, and oversight of the entire process. Dr. Lochbaum holds a Ph.D. in computer science from Harvard University and is a co-inventor on several patents related to Latent Semantic Analysis (LSA) and Continuous Flow.

Sue Lottridge, Ph.D., is Chief Scientist of Natural Language Applications at Cambium Assessment, Inc. She has a Ph.D. in Assessment and Measurement from James Madison University and Masters’ degrees in Mathematics and Computer Science from the University of Wisconsin—Madison. In this role, she leads CAI’s machine learning and scoring team on

the research, development, and operation of CAT's automated scoring and feedback software. Dr. Lottridge has worked in automated scoring for 15 years and has contributed to the design, research, and use of multiple automated scoring engines, including equation scoring, essay scoring, short answer scoring, speech scoring, and crisis alert detection.

Quintin Love, Ph.D., is an Independent Consultant. He holds a doctorate in Educational Psychology from the University of Minnesota, both a Master's and Bachelor's degree in Middle Grades Education from Albany State University, and a graduate certificate in Educational Leadership from Lamar University. His professional experiences include employment as a middle school mathematics teacher at Title 1 schools serving predominantly students of color in Georgia and Texas, as a Psychometrician on multiple statewide summative assessment programs, and as Senior Research Associate in Assessment Research and Innovation at WestEd. His research interests include QuantCrit, Culturally Responsive Assessments, and Psychometrics.

Susan Lyons, Ph.D., is Owner and Principal Consultant at Lyons Assessment Consulting. Her firm is a leader in supporting innovation in educational assessment and accountability. Susan is dedicated to transforming traditional systems to serve all students better. Though Dr. Lyons primarily works as a practitioner, she remains active in writing and publishing. Susan frequently releases white papers documenting her applied work, publishes in peer-reviewed journals, and presents at professional conferences. Susan is the co-founder and Executive Director of Women in Measurement, a non-profit organization dedicated to advancing gender and racial equity in the professional field of educational measurement.

Scott F. Marion, Ph.D., is Executive Director of the National Center for the Improvement of Educational Assessment. Marion was recently elected to the National Academy of Education and serves on the National Assessment Governing Board, which sets policy for NAEP (2022–2026). He has served on multiple National Research Council committees and coordinates and/or serves on ten state or district technical advisory committees for assessment and accountability. Dr. Marion is a co-author (with Lane) of the validity chapter in *Educational Measurement* and recently co-authored a book on how to help improve the instructional usefulness assessment.

Monique Matute-Chavarria, Ph.D., is Assistant Professor of Special Education at New Mexico State University (NMSU). Before obtaining her Ph.D. in Special Education, she worked with Nevada Early Intervention Services, working with children from birth to three with disabilities. Dr. Matute-Chavarria identifies as a Black mother, scholar, and Afro-Latina; therefore, her research focuses on Black students and families. Her research centers on the intersections of race, family, and disability. She also researches the use of hip-hop pedagogy practices (i.e., DJing) as a tool for writing with Black students with and without disabilities.

Kyndra V. Middleton, Ph.D., is Professor of Educational Psychology at Howard University, where she teaches statistics and measurement courses. Her research focuses on ensuring validity, equity, and fairness for all students as well as on increasing the number of underrepresented minorities in STEM. Her educational as well as professional training has allowed her to provide service as a psychometric consultant for numerous grants, organizations, universities, and a professional sports league. She received her MA and Ph.D. in Educational Measurement and Statistics from The University of Iowa and her BS in Mathematical Sciences from Florida Agricultural and Mechanical University.

Robert J. Mislevy, Ph.D., is Emeritus Professor of Measurement and Statistics at the University of Maryland and retired Frederic Lord Chair in Measurement and Statistics at ETS. His research applies developments in statistics, technology, and cognitive and social psychology to educational assessment, including a multiple-imputation approach for latent-variable modeling in large-scale assessment, an evidence-centered assessment design framework, and Bayesian networks for game- and simulation-based assessments. His books include *Bayesian Psychometric Modeling* and *Sociocognitive Foundations of Educational Measurement*. He has received NCME's annual award for Contributions to Educational Measurement five times and career contribution awards from AERA, NCME, and the Psychometric Society.

Pricella Morris, Ph.D., is Assistant Professor at the University of Nevada, Las Vegas (UNLV) in the Department of Early Childhood, Multilingual, and Special Education. Dr. Morris identifies as a Black mother scholar who focuses on the impacts of race and ability on young Black children and their families. In 2022, she was awarded the J. David Sexton Doctoral Award from the Council for Exceptional Children Division of Early Childhood for her commitment to service to young children with disabilities and their families. Her research focuses on racial and dis/ability identity, dismantling antiblackness in education and teacher preparation, disproportionate discipline, and culturally responsive pedagogy.

Michael Nakkula, Ph.D., is Chair of the Human Development and Quantitative Methods Division at the University of Pennsylvania Graduate School of Education. Over the past 30 years, he has helped develop and direct counseling and human development programs at Harvard and Penn. A licensed counseling psychologist, he studies the intersection of mental health and optimal development among children, youth, and young adults in school and community settings. He is the lead editor of the three-volume series *Adolescent Psychology in Today's World* (Praeger Press, 2018) and co-author with Eric Toshalis of *Understanding Youth: Adolescent Development for Educators* (Harvard Education Press, 2006).

Sharon Nelson-Barber, Ph.D., a sociolinguist, directs WestEd's Culture and Language in STEM Education. Her research centers on understanding how the sociocultural contexts in which students live influence ways in which they make sense of STEM schooling. Working side by side with other Indigenous researchers whose communities span the lower 48 US states, Alaska, Hawai'i, Micronesia, and areas of Polynesia, she explores how aspects of cultural knowledge can become visible in assessment and evaluation, making schooling more equitable for all students. She earned a doctorate in human development from Harvard University and completed postdoctoral work at Stanford University as a Spencer Fellow.

Jaylin Nesbitt, MA, is Research Associate on the Assessment Research & Innovation team at WestEd. Ms. Nesbitt conducts quantitative and qualitative analyses and assists in designing educational research studies centered around culturally and linguistically responsive assessments and early learning assessments. She holds a BS in psychology from Spelman College and an MA in quantitative psychology from James Madison University. Her research experience and interests are centered around quantitative data analyses and the research and development of culturally and linguistically responsive assessments to better represent Black students who speak African American Vernacular English (AAVE) in the US K-12 education system.

Maria Elena Oliveri, Ph.D., is Research Associate Professor and Director of Psychometric Consulting at the Buros Center for Testing, University of Nebraska-Lincoln. Her expertise is

on assessing workplace-relevant competencies, educational measurement, fairness, validity, and culturally responsive assessment. She has over 100 publications and presentations on these topics, with a focus on the design, development, and use of international assessments. She has served in multiple leadership roles, including the Chair for International Test Commission (ITC) Guidelines for the Fair and Valid Assessment of Linguistically Diverse Populations, ITC Technology-Based Assessment Guidelines (steering committee members); ITC Chair of Publications and President: Learning Assistance Teachers Association.

Mya Poe, Ph.D., is Professor of English at Northeastern University. For more than 20 years, she has studied college-level writing assessment practices and advocated for eliminating assessment practices that result in unnecessary structural barriers for students and perpetuate deficit narratives of historically marginalized students. In addition to teaching undergraduate courses on writing research methods and scientific writing, she also teaches graduate courses on writing assessment and the teaching of writing. She has directed three writing programs and worked extensively with faculty across the US to improve the teaching of writing. She is co-editor of the international writing research journal *Written Communication*.

Sarah Quesen, Ph.D., serves as Director of Assessment Research and Innovation, where she directs research studies and supports evidence-based technical assistance to states and districts. She is committed to identifying ways to enhance diversity, equity, inclusion, and accessibility considerations in assessment. Dr. Quesen has over 25 years of experience in education and continues to teach at the University of Pittsburgh in the Department of Statistics. Dr. Quesen has a Ph.D. in Research Methodology from the University of Pittsburgh, an MPH from West Virginia University School of Medicine, and a BS in Culture and Communication from New York University.

Heather Roeters-Solano, M.Ed., is Principal Assessment Specialist at Pearson, where she works on the development of assessments, accommodations, and assessment policies for special populations. Her areas of interest include language acquisition and language proficiency assessment. Dr. Roeters-Solano earned a BS in Applied Learning and Development and an M.Ed. in Curriculum and Instruction from The University of Texas at Austin.

Pōhai Kūkea Shultz, Ph.D., is a Native Hawaiian daughter, wife, and mother and currently resides with her family in the beautiful valley of Waimalu on the island of O‘ahu. As a former teacher, Dr. Kūkea Shultz’s experiences include working in teacher education for Hawaiian communities, supporting initiatives in student services with a focus on the recruitment, retention, and graduation of Hawaiian students, and currently serving as the principal investigator of the Kaiapuni Assessment of Educational Outcomes (KĀ‘EO). Dr. Kūkea Shultz’s research interests focus on the potential and promise of educational self-determination in assessments for native students.

David Slomp, Ph.D., is Professor of Literacy and Assessment and Associate Dean of Graduate Studies and Research in the Faculty of Education at the University of Lethbridge, where he has also served as the Board of Governors’ Teaching Chair. He is Editor-in-Chief of *Assessing Writing*. He served as a department editor for the *Journal of Adolescent and Adult Literacy*. He has guest-edited special issues for the *Journal of Writing Assessment*, *Assessing Writing*, and *Writing Analytics*. His work research focuses on the social consequences of assessment design and use.

Guillermo Solano-Flores, Ph.D., is Professor of Education at Stanford University. His research focuses on assessment, cultural diversity, and fairness. Main areas of interest include

translation in the testing of multiple linguistic groups, the use of generalizability theory in the testing of English learners, the use of Boolean logic in the analysis of classroom interactions in linguistically diverse classrooms, and the design of platforms for online and iPad-based testing. He is a member of the National Assessment Governing Board and has served on multiple technical advisory boards, including the Smarter Balanced Assessment Consortium and the Department of Education of Hawaii.

Winston C. Thompson, Ph.D., is Associate Professor of Educational Studies and Associate Professor of Philosophy (by courtesy) at The Ohio State University. He is the incoming Director of OSU's Center for Ethics and Human Values. Thompson's scholarship explores ethical/political dimensions of educational policy and practice. His work on justice and the role of education in pluralistic, democratic societies has appeared in leading journals and volumes collecting scholarship in the philosophy of education, moral education, and civic education. He is the editor of *Philosophical Foundations of Education* and co-editor (with John Tillson) of *Pedagogies of Punishment: The Ethics of Discipline in Education*.

Eric Toshalis, Ph.D., authored the award-winning book *Make Me!* (2015), and is co-author of *Understanding Youth* (2006), both published by Harvard Education Press. Toshalis also co-authored the 13th edition of *Comprehensive Classroom Management* (Pearson, 2024). Over the last three+ decades, Dr. Toshalis has served in a variety of roles in and around public secondary education, including middle and high school teacher, mentor, union president, researcher, professor, and community activist. He received his Bachelor's, teaching credential, and M.Ed. from the University of California, Santa Barbara, and his second Master's and Doctorate from Harvard. He provides independent scholarship and education consulting at EngagingResistance.com.

Elise Trumbull, Ph.D., is an applied psycholinguist whose research has taken her into the realms of sociolinguistics, cross-cultural cognitive psychology, and anthropology in service to understanding culturally responsive education and culturally valid assessment. Collaboration with colleagues from American Indian and Alaska Native communities and specialists in bilingual education and assessment have been essential to her own learning. Dr. Trumbull has studied ten languages and participated in research across the United States and in US entities in Micronesia. She is a co-author of numerous books and journal articles on the role of language in assessment, cultural and linguistic issues in the education of immigrant students, and culturally responsive evaluation and assessment in Indigenous communities.

Lauren White, Ph.D., is Senior Research Scientist at Pearson, where she serves as the lead psychometrician on a number of large-scale school assessment contracts. Dr. White has over a decade of experience in educational measurement, specifically large-scale licensure and credentialing assessments, school assessments, and assessment policy. Her research interests include culturally responsive assessment practices, policy-informed assessments, and the influence of large-scale accountability systems on educator quality. Dr. White holds a Ph.D. in Measurement and Statistics, an MS and EdS in Mental Health Counseling, and a BS in both Psychology and Sports Management from Florida State University.

Acknowledgments

This book is the culmination of several years of dreaming, planning, researching, and writing. We are deeply indebted to the 43 authors who contributed to this volume! Their dedication, perseverance, and willingness to submit multiple drafts and revisions are what makes this effort so special. We appreciate our authors' willingness to develop and share their work, subject it to our review, respond to questions, and turn around edits/revisions on an ambitious production schedule. Obviously, this volume would not have taken shape without their thoughtful and timely contributions.

We also want to thank Kadriye Ercikan for her support as the NCME Book Series Director and the other members of the NCME Editorial Board—Guillermo Solano-Flores and Rochelle Michel—for their feedback.

We would also like to thank The National Center for the Improvement of Educational Assessment (Center for Assessment) for supporting Carla's time spent on this work. Finally, we would like to acknowledge our spouses and families who shared their time with us so we could focus on completing this book the way we had envisioned it.

Reimagining the Future of Assessment that Honors Human Diversity

Carla M. Evans and Catherine S. Taylor

Why This Volume?

Students in the United States's educational system are quite diverse. According to the Digest of Education Statistics (National Center for Education Statistics, 2024), there were almost 50 million Pre-K to grade 12 public students in the 2021–2022 school year, and less than half of those students were White. We also know from the results of large-scale educational assessments, such as the National Assessment of Educational Progress and state-specific achievement tests, that there are performance gaps between White and non-White student groups that have persisted over time. If we believe that all students can learn and achieve high academic standards, then the systematic disparities in K-12 student performance are rooted in differential access to resources and opportunities to learn at a minimum. The question measurement scholars are becoming more and more aware of is how score differences may also be attributable to assessment development and implementation practices. Even before the murders of George Floyd, Breeana Taylor, and Rayshard Brooks in 2020, public consciousness of racial injustice has been growing. There has also been a natural questioning and reawakening of our collective awareness of the inequities in outcomes, access, and opportunities in our educational systems, including how educational measurement and assessment systems contribute to or reify existing inequalities.

Educational measurement has a troubled past, one that shows a lack of respect for the diversity of human characteristics. We see this most clearly in the history of intelligence testing and early measurement scholars' beliefs in Eugenics (Russell, 2023). Early measurement specialists (Galton, Terman, Spearman) designed and analyzed intelligence tests based on these beliefs, which were carried over to achievement testing. Recently, influential authors in the measurement field have raised a clarion call for interrogating our current measurement and assessment practices based on questions about biased test design, review processes, scoring, and post hoc analytical policies, practices, and processes (Dixon-Román, 2020; Randall, 2021; Russell, 2023). As with any other social policy and practice, educational measurement can continue to be used to reify existing unjust and racist systems and structures, or it can be used to actively disrupt such systems and structures (Kendi, 2019).

The purpose of this edited volume is to help educational measurement and assessment specialists, researchers, practitioners, policymakers, and users explore a reimagined future where

2 • Culturally Responsive Assessment in Classrooms and Large-Scale Contexts

the unique cultural and social identities of students shape assessment purpose, design, implementation, use, and validation processes. The phrase “cultural and social identities” is intentionally broad and inclusive. Just as the word “culture” itself is multifaceted, dynamic, fluid, and changing (Organization for Economic Cooperation and Development [OECD], 2016), so too are people’s identities. In other words, this book is about more than just racial or ethnic diversity and how that interacts with assessment. This book is about human diversity in all of its forms and how diversity interacts with assessment.

The chapter authors use theory, research, and practice to explore key issues and possible solutions within classroom assessment and large-scale standardized testing contexts. There are chapters by assessment and measurement experts, as well as scholars and practitioners with expertise in sociocultural learning theory, philosophy of education, critical theories, multilingual education, and Indigenous education. The chapters differ somewhat in terms of length and the depth with which topics are discussed. We see this as an expected outcome of bringing together a wide variety of authors from different disciplinary perspectives. The authors summarize what is known from previous research, what is currently being explored, and what future research is needed to move measurement and assessment policy and practice forward in the shorter and longer term with the aim of ensuring cultural responsiveness in assessment.

One feature of this volume is the inclusion of commentators for each section. Commentators were solicited due to their recognized expertise and experience in the field of culturally responsive assessment. Commentators provide a critical lens on the chapters included in this volume and the larger field of culturally responsive assessment. They do an excellent job of highlighting themes and tensions across the chapters in their section and filling in any gaps that may not have been addressed by the individual chapters. For this reason, we strongly encourage every reader to read not only the individual chapters in each section but also the commentator chapter(s) as readers will benefit from doing so with a more complete understanding than reading any chapter as a stand-alone requiem on the topic.

Positionality Statement

We, the editors of this volume, are White women who have held different positions of power and privilege in our professional careers. We are aware of our positionality as White scholars and practitioners in this space and, as such, are not seeking to question the rich tradition of scholarship related to culture, language, human diversity, pluralism, and testing from those who have come before us. Rather, our goal is to highlight, summarize, and extend this scholarship to the field of educational measurement and assessment to support changes so that all students’ ways of knowing and being are reflected in large-scale and classroom assessment policies, practices, and processes.

Background on Culturally Responsive Assessment

The first use of the term “culturally responsive assessment” that we could locate in the scholarly literature in the field of education was in the 1998 special issue of *The Journal of Negro Education*. In the introduction and overview of that special issue, Stafford Hood (1998b) explained that the idea for the special issue emerged from a symposium at the 1997 annual meeting of the American Educational Research Association. The symposium focused on

preliminary discussion on building bridges between culturally responsive pedagogy and performance-based assessments. The panelists felt that it was necessary for us to “think out

loud” and engage in professional discourse about whether the notion of culturally responsive pedagogy could be manifested in such testing.

(Hood, 1998a, p. 184)

It was, after all, only a couple of years after Gloria Ladson-Billings published two trailblazing articles about culturally relevant pedagogy (Ladson-Billings, 1995b, 1995a), which encompasses three components: academic achievement, cultural competence, and sociopolitical or critical consciousness. Gloria Ladson-Billings was one of the panelists in the 1997 symposium along with Stafford Hood and other authors whose work was then featured in the special issue (Boodoo, 1998; Hood, 1998a; Lee, 1998; Qualls, 1998; Smith-Maddox, 1998).

Yet, questions and concerns about the implications of human diversity and culture on the design, implementation, and validation of educational assessments have been raised in the scholarly literature since the early 1940s, although it wasn’t called “culturally responsive assessment.” For example, just before World War II, psychologists debated intelligence tests and challenged prevalent hereditary ideas, exploring different factors that account for differences in intelligence—including cultural bias in test items (Eels et al., 1951). Eells presented his view that “variations in opportunity for familiarity with specific cultural words, objects or processes required for answering the test items seem to the writer . . . the most adequate general explanation for most of the findings” (cited in Sargent, 1952, p. 209–210). Edmund Gordon (1977) recounts the history of this line of research and explains why “culture-free” IQ testing was ultimately deemed impossible because of normative testing approaches that worked at cross purposes with accurate inferences about minority student group performance.

During the Civil Rights Movement of the 1950s and 1960s, attention was once again turned to the “the inappropriateness of standardized tests for the assessment of minority group members” (Gordon, 1977, p. 3). Psychometricians during this time would routinely “add five to 15 points to the scores of minority subjects to compensate for the assumed artificial depression in test scores resulting from the inappropriateness of the test” (Gordon, 1977, p. 3). However, arbitrarily inflating test scores did not address the underlying issue or solve the cultural divide between the test content, design, and development and the lived experiences of minority student groups. Therefore, the practice was ultimately abandoned. By the early 1970s, some were calling for a moratorium on the use of standardized tests for minority student groups (Gordon, 1977). Yet, standardized testing practices have continued unchanged, even though questions about the impact of culture on testing continued (Miller-Jones, 1989).

There is also a long line of research on issues related to multicultural assessment (Stobart, 2005) and cross-cultural assessment for Indigenous students around the world (Castagno & Brayboy, 2008; Klenowski, 2009; Kūkea Shultz & Englert, 2023; McCarty & Lee, 2014; Nelson-Barber & Trumbull, 2007; Preston & Claypool, 2021). In the United States, concerns about the validity, reliability, and fairness of standardized test results for Indigenous students led to the provision in the 1988 Indian Education Act of “a program of research and development to provide accurate and culturally specific assessment instruments to measure student performance in cooperation with Tribes and Alaska Native entities” (Chavers & Locke, 1989, p. 18–19). Elise Trumbull (formerly Elise Estrin) and Sharon Nelson-Barber have raised important questions and discussed key issues in cross-cultural assessment pertaining to American Indian and Alaska Native students since the mid-1990s (Estrin & Nelson-Barber, 1995; Nelson-Barber & Trumbull, 2007; Trumbull et al., 2015; Trumbull & Nelson-Barber, 2019). For example, Trumbull and Nelson-Barber argue, along with others, that it is impossible to make accurate inferences about what students know and can do without considering students’ language and culture (Mislevy, 2018; Nortvedt et al., 2020; Preston & Claypool, 2021; Kūkea Shultz & Englert, 2021; Solano-Flores, 2011; Solano-Flores & Nelson-Barber, 2001).

Behind the scenes in these discussions about assessment in the context of culture are historical developments in educational measurement, learning sciences, and pedagogy. For example, concerns and questions about the accuracy and appropriateness of norm-referenced testing—where student achievement is compared to other students—began in the 1960s and 1970s, which led to a movement towards criterion-referenced testing since that time—where student achievement is compared against performance criteria (Glaser, 1963; Hambleton et al., 1978; Popham & Husek, 1969). Additionally, advances in our understanding of how people learn highlight the disconnect between test design and sociocultural and socio-cognitive theories of learning and cognition (National Academies of Sciences, Engineering, and Medicine, 2018; National Research Council, 2000, 2001). Our better understanding of how people learn and understand, as well as the increased complexity of learning goals or academic standards, challenges the appropriateness and adequacy of “one-for-all” standardized testing for high-stakes accountability purposes.

Related to theories of learning and cognition, there is an important history and context behind the development and theory of culturally responsive education and other associated concepts (e.g., culturally relevant education, culturally sustaining education), which many of the chapter authors highlight or discuss. We refer interested readers to the following starter readings: Aronson & Laughter, 2016; Gay, 2000; Ladson-Billings, 2014; Paris, 2012; and Paris & Alim, 2014.

State of the Field: The Need for This Edited Volume

The current conversation around culturally responsive assessment builds from these historical foundations in various ways. Scholars and researchers often question the language, content, format, administration, scoring, and/or use of large-scale and classroom assessments based on the diverse, pluralistic, and multicultural society in which we live. As others have said previously, there are difficult technical, developmental, conceptual, and other challenges to both understand and operationalize culture in the context of assessment (Bennett, 2023; Gordon, 1992; Poe et al., 2023; Sireci, 2020; Taylor, 2022). There are also racist and oppressive systems and policies that infiltrate every part of society, including educational measurement and assessment (Dixon-Román, 2020; Randall, 2021; Russell, 2023).

As signaled in the preceding section, there are a plethora of terms and meanings appearing in the scholarly literature related to the influence of culture, lived experiences, and language on student assessment. The plethora of terms and meanings reflects the complexity of the problem and the necessary contextualized nature of assessment that reflects culture. A sample of terms used in the measurement and assessment literature within the last few years include culturally responsive assessment (Evans, 2023; Nortvedt et al., 2020; Walker et al., 2023), culturally relevant assessment (O’Dwyer et al., 2023), culturally sustaining assessment (Kūkea Shultz & Englert, 2023), culturally and socially responsible assessment (Taylor, 2022), anti-racist assessment (Randall, 2021; Randall & Poe, 2021), socioculturally responsive assessment (Bennett, 2023), and so on. There is not one set definition of any of these terms from a teaching and learning perspective (Aronson & Laughter, 2016; Franco et al., 2023), let alone as applied to student assessment. Additionally, there is little to no agreement on principles or paradigms that can be used to ensure the vision and goals of any selected term are operationalized in large-scale and classroom assessment design, implementation, or validation in consistent and similar ways. For example, in a recent systematic review of the research on the quantitative measurement of culturally responsive practice, the researchers noted that a “challenge in the field continues to be that the terminology related to culturally responsive practice is so expansive and often changes, which can prevent researchers and practitioners from identifying the instruments that would be most useful or theoretically

aligned” (Franco et al., 2023). The lack of clear and consistent terminology leads to the jingle-jangle fallacy—an erroneous belief that two different things are the same because they share the same name (jingle fallacy) or two identical or almost identical things are different because they are labeled differently (jangle fallacy) (Kelley, 1927).

If the larger educational research field has not yet settled on common definitions or theoretical frameworks for the plethora of terms related to culturally responsive education, we should not expect the educational measurement and assessment field to be at a place at this moment in time where there is consensus or shared approaches to defining what constitutes culturally responsive assessment and how it should be applied in practice. With others, we acknowledge that the lack of shared understandings and clear definitions is a challenge and stumbling block to moving policy and practice forward. If researchers and scholars are not using the same terms in the same way, then it is difficult to create future lines of research to examine what is effective, for whom, under what conditions, and why. Additionally, because assessment should be coherent with curriculum and instruction, we recognize the larger need for shared definitions across the educational research and measurement fields, which is no easy feat. We believe that the field needs a broad overview of how these terms are being defined and operationalized at this moment in time—a state of the field—to help support continued conversations that lead to shared understandings and drive distinct, even if somewhat related, lines of research.

Our Approach to Definitional Issues

Based in the previous discussion, it should come as no surprise that our approach as editors has not been to define culturally responsive assessment or other related terms for our authors. We thought such a top-down approach was inappropriate given our positionality and did not serve to describe the current state of the field. Rather, we view this book as an opportunity to explore how a sample of researchers and scholars who are actively doing work related to culturally responsive assessment are defining and operationalizing the term or other terms they may employ (e.g., anti-racist assessment, culturally relevant assessment, culturally sustaining assessment). Our approach was to ask each of the chapter authors to select terms and define their meanings carefully.

Continuum of Culturally Responsive Assessment Meanings

We noticed early on in our review of the chapters in this volume that the jingle-jangle fallacy (discussed in the previous section) was present across the way authors used terms in this volume. To help our readers make sense of what they are reading, we created a figure that illustrates the continuum of meanings typically found behind the various terms used in this edited volume. Figure 1.1 shows how authors’ work tends to fall along a continuum from representing culturally responsive assessment as meaning inclusivity and engagement at the minimum to taking informed action for social justice at the maximum. The terms in the figure are briefly defined below.

Inclusivity and Engagement

Assessments that engage students through multiple means of engagement, representation, expression and/or action (e.g., applying principles of Universal Design for Learning to assessment). Assessments that use contexts or scenarios that students may find familiar, interesting, or engaging (e.g., music, art, sports).



Figure 1.1 Continuum of Terms Related to Culturally Responsive Assessment Reflected in this Volume: From Inclusivity to Action

Representation and Exposure

Assessments where students either (a) see themselves, their communities, and cultures represented in some form or fashion or (b) are exposed to other cultures’ lived experiences, conceptions, and belief systems in the items/tasks, stimulus materials, and/or images. Representation and exposure sometimes focus on surface-level features of culture (e.g., food, fashion, festivities).

Perspective Taking

Assessments that require students to explain how an issue, problem, or scenario may be perceived or viewed from different cultural perspectives. Perspective-taking requires students to possess a level of cultural competence wherein students understand, research, or surmise about different cultural beliefs and how they differ from their own. Here, cultural competence is defined as knowing one’s own culture and at least one other. Cultural competence requires students to understand how culture affects expected norms, ways of being and knowing in the world, etc.

Taking Informed Action

Assessments that include contexts or scenarios that identify social justice or community issues. Students must apply their disciplinary content knowledge and skills to address those issues by planning for or taking informed action (e.g., addressing food insecurity, scarcity, or food deserts; creating solutions to local homelessness; rising costs of medical care, prescription medications, or affordable housing). Chapter authors who implicitly or explicitly operationalize culturally responsive assessment as “Taking Informed Action” tend to focus on the critical consciousness aspects of the larger educational movement. Critical consciousness is defined here as the extent to which the assessment accounts for the systemic oppression of marginalized people in our country and takes action against those oppressive elements (Freire, 2000).

Chapter Authors' Use of Culturally Responsive Assessment Meanings

Some chapter authors define and operationalize culturally responsive assessment using two or more of these meanings. This approach represents a tacit understanding that the continuum is more of a step ladder where an assessment that requires students to take informed action also requires students to be able to see an issue from multiple perspectives and is designed so that multiple cultures are represented and inclusivity and engagement are maximized. Yet, on the other hand, an assessment intended to maximize inclusivity and engagement using principles of Universal Design for Learning does not necessarily also include perspective-taking or require students to take informed action. This understanding is roughly congruent with broader understandings about the relationship between culturally sustaining pedagogy and Universal Design for Learning (Alim et al., 2017; Waitoller & Thorius, 2016), as well as culturally sustaining assessment and its relationship with culturally responsive, relevant, and sensitive assessment (Evans, 2021).

Additionally, some chapter authors discuss culturally responsive assessment from a systems perspective. In particular, the two chapters that focus on Native Hawaiian assessment systems (Englert and Kūkea Shultz—large scale assessment Section 2; Keehne, Kelling, and Awo-Chun—classroom-level assessment Section 3) illustrate how, at minimum, every assessment within a culturally responsive assessment system should be designed so that all students can show what they know and can do because the assessment is designed with inclusivity and engagement in mind. However, not every assessment within an assessment system—from informal probes of student thinking to more formal standardized assessments—will necessarily require students to take informed action or explain an issue from multiple perspectives. As these two sets of chapter authors show, an assessment system perspective may be a useful framework for thinking about the implementation of culturally responsive assessment in the future.

Overview of Contributions from the Chapters in This Volume

Section 1: Culturally Responsive Assessment: Theoretical Foundations and Concepts

The chapters in Section 1 of this volume discuss the theoretical and conceptual foundations of culturally responsive assessment. This section of the book is wide in scope and diverse in perspectives. **Coppens and Kelley's** chapter opens this section by framing culturally responsive assessment as an integration of sociocultural approaches to learning and assessment. This exposition is important because it challenges what might be tacit assumptions about the relationship between learning theory and assessment design, implementation, and use processes. The next chapter by **Nakkula and Toshalis** explores one key concept underpinning culturally responsive assessment, which is that the ways we assess student performance impact student engagement in school and overall student achievement motivation. The authors emphasize the importance of student choice, voice, affirmation, and belonging as critical to understanding the roots of student engagement and motivation. Recommendations from this chapter locate culturally responsive assessment as the key to improving student engagement, motivation, and academic achievement. In the next chapter, **Thompson and Erickson** zoom out and provide a philosophical perspective on educational assessment generally: what it is and ought to be to ensure equity and fairness across diverse cultural contexts. They shed light on the normative judgments embedded in assessment tools and processes and assert that assessment processes operate within a non-ideal social context where there are intricate webs of historical injustices, systemic biases, and cultural nuances that shape educational outcomes. This chapter helps us further understand and explore how assessments are not as neutral as they might seem and carry many implicit assumptions about individuals' capabilities and potentials.

The next three chapters provide a critical lens on current test data analysis, validation, and use practices and how these practices can be reshaped so they support the ultimate goals of culturally responsive assessments—equity and social justice. **Love’s** chapter presents Quantitative Critical Race Theory (QuantCrit) as a theoretical framework that can be used to guide quantitative and psychometric analysis intended to contribute to efforts of equity and social justice using data from culturally responsive assessments. This chapter provides important insights into the types of validity evidence and argumentation that should be conducted to support, rather than work at cross purposes with, a justice-oriented use of culturally responsive assessments. **Lyons, Oliveri, and Poe’s** chapter continues this call for an approach to educational measurement that explicitly prioritizes the aims of advancing social justice. They introduce a framework for justice-oriented assessment use to evaluate the appropriateness of proposed uses of educational assessments within culturally and linguistically diverse societies. **Lane and Marion** then present their perspective on how an argument-based approach to validation research provides a mechanism for attending to the validity of score interpretations and uses for students with diverse cultural and linguistic backgrounds. They propose an interpretation and use argument and validity argument for culturally responsive assessments, drawing on insights from validity theorists, situated learning theories, and principled assessment design practices. This chapter challenges prevailing notions about validity and argues that interpretation and use arguments and validity arguments for assessments must be shaped to account for the cultural diversity in the student population. Together, these chapters contribute to a growing discourse on the role educational measurement can play in both perpetuating and disrupting systems of oppression in pluralistic and multicultural societies. Section 1 ends with a commentary by **Solano-Flores** on the chapters therein, in which he discusses the theoretical underpinnings of culturally responsive assessment—particularly the social nature of cognition and learning and how such notions intersect with fairness and social justice considerations.

Section 2: Implications of Culturally Responsive Assessment for Large-Scale Assessment Practices

The chapters in Section 2 of the volume constitute a set of examples about how the different authors have thought about and operationalized the implications of culturally responsive assessment for large-scale assessment practices. There are more practical resources in Section 3 in the context of classroom assessment, which may reflect where we are as a field in terms of the challenges and limitations of our understanding of culturally responsive assessment within the context of large-scale assessment.

Taylor and Ferrara’s chapter opens this section by presenting an examination of measurement practices in large-scale testing. They identify ways in which test development and psychometric processes can result in assessments that are not relevant to the lived experiences of many students. To do so, they examine each aspect of the assessment development process, including construct definition, specifications, item writing, test construction, item analysis, and scale development to reveal practices that introduce unintended bias into these assessments. Taylor and Ferrara’s chapter foregrounds the need for additional development and investigation into new psychometric models that are designed with the intent of supporting culturally responsive approaches to large-scale assessment rather than retrofitting traditional psychometric models. Current psychometric models focus on the construction of numeric scales that can be used to measure and compare. While these models served the purposes of Eugenics quite well, rank ordering people by race and “proving” hereditarian beliefs about intelligence, they were not built for an education system where all students are expected to achieve high standards. New models would pave the way for assessment focused on what successful performance looks like and for making fair, valid, and comparable claims for all students and student groups.

The next two chapters focus on specific testing contexts and provide useful examples of how the authors' definitions of culturally responsive assessment are currently being applied in similar and different ways to large-scale assessment practices in these settings. **Englert and Kūkea Shultz's** chapter highlights a development process that is grounded in the culture and worldview of students using a Native language assessment as an example. They contend that for an assessment to reflect cultural relevancy, there needs to be a critical interrogation of its foundations, theory of action, and purpose. They argue this cannot be accomplished except through partnerships with communities. In the next chapter, **Slomp, Oliveri, and Poe** demonstrate the application of a framework they created and have refined over time to support the design of culturally sustaining assessment programs. They use a case study detailing the development of a scenario-based writing assessment program and what they learned through multiple iterations over time.

Next, **White, Nesbitt, Roeters-Solano, Quesen, Lottridge, and Lochbaum** provide a critical examination of the assumptions underlying the scoring of spoken and written tasks in large-scale assessments. The authors focus on the experiences, values, and identities of historically underrepresented students. They emphasize that the fairness and validity of constructed response scores depend on decisions made at each step in the process such as item development, rubric development, the selection of anchor and training papers, and rater selection. Additionally, they stress the significance of considering the underlying context and cultural assumptions that inform both human and automated scoring processes. Section 2 concludes with two commentaries—one by **Au** and the other by **Mislevy**. **Au** compares the culturally responsive assessment approaches discussed in the chapters with multicultural education. He also raises important questions about how culturally responsive assessment could reify stereotypes and provides a critical perspective on the large-scale testing enterprise and positivist objectivity in relation to the goals of culturally responsive assessment. **Mislevy** organizes his comments from the perspective of educational assessment as an evidentiary argument and then discusses cross-cutting themes and insights from across the chapters.

Section 3: Implications of Culturally Responsive Assessment for Classroom Assessment Practices

The chapters in Section 3 of the volume focus on the implications of culturally responsive education for classroom assessment practices. **Middleton and Haynes's** chapter opens this section of the book by discussing the similarities and differences between culturally relevant and culturally responsive pedagogies. They then present a variant of ecological systems theory to provide a lens through which to view the development of culturally responsive practices in educational settings. The chapter provides suggestions and recommendations for educators on how to integrate culturally responsive practices into their educational frameworks, thereby promoting equity and excellence in education for all students. The next chapter from **Keehne, Kelling, and Awo-Chun** provides a complementary lens and shows how an ecological and systems approach to educational change was designed and implemented in a network of culturally grounded charter schools in Hawai'i. They describe the creation of culturally responsive assessment tools that manifest the power of the individual school missions and the collective advocacy of the schools. The chapter concludes with implications for this work on policies and assessment systems in a wider context, which can inform others who are attempting to enact change at the systems level.

The next two chapters address the implications of culturally responsive assessment for specific assessment processes that occur within classrooms—namely, formative assessment and grading practices. **Furtak and Deverel-Rico** explain how culturally responsive approaches to science teaching encourage us to consider how classroom assessment—both in its design and

enactment—can create space for learners to make connections with their prior experiences as they make sense of phenomena that are interesting to them and relevant to their lives. They examine the enactments of two phenomenon-based high school biology formative assessment tasks to understand the ways in which variations in teachers' responsiveness to student ideas through their classroom practices can enable students' learning. This research highlights the interconnectedness between formative assessment processes and instructional strategies in the classroom. **Brookhart and Chism** combine insights from two areas of research—research about the relationship between culture and pedagogy and research about grading practices—to provide and support recommendations about equitable and culturally relevant grading practices. They provide practical suggestions to teachers about how to apply their insights and create conditions where students receive grades that reflect their academic achievement. The next chapter addresses culturally responsive assessment from a completely different perspective because it focuses on the assessment of students' behavior in classrooms rather than academic assessments.

All preceding chapters in this volume have focused on academic assessments. However, this volume includes a chapter on behavioral assessment and how that interacts with culturally responsive academic assessment. **Brown, Matute-Chavarría, and Morris** explain the disproportionate identification of Black students for disciplinary referrals, suspensions, and special education because of the *mis*-assessment of the behaviors of these students. They frame the issue using research and then delve into recommendations for supporting the culturally responsive assessment of Black students' behavior. This is an important topic for two primary reasons. First, assessments of students' behaviors are routinely incorporated into teachers' thinking as they facilitate learning, provide students with opportunities to learn, and assign grades. Second, this chapter expands the boundaries of what educational measurement and assessment experts and others should consider when interrogating how students' unique cultural and social identities shape all assessment processes, including behavior assessments. Section 3 concludes with two commentaries—one by **Nelson-Barber and Trumbull** and the other by **Diaz-Bilello**. **Nelson-Barber and Trumbull** identify themes across the chapters related to culturally responsive assessment and then suggest an additional topic or theme that they believe is critical but not addressed in the chapters—the role of language in the assessment process. **Diaz-Bilello** contextualizes her remarks in the context of how culturally responsive assessment can strengthen classroom activity systems, the need for practical implications of the ideas given the district locus of control over the instructional work in schools, and areas for future development needed.

Conclusion

We invite you to lean into the complexity of these issues and the confusion around terms with us. We don't anticipate you will agree with everything that is said in this edited volume or that all of your questions will be answered about culturally responsive assessment from a theoretical/conceptual standpoint or how the theory/concept is applied in large-scale or classroom assessment contexts. However, we urged the chapter authors to substantiate their arguments and claims with reasons and evidence. We invite you into this conversation and ask for you to do the same. Although the state of the field is currently dispersed in terms of how terms are being used and operationalized in practice, this does not eliminate the educational measurement and assessment community's responsibility to respond to and wrestle with the complex realities of students' contextual, cultural, and linguistic pluralism. For decades, scholars have questioned the validity of classroom and large-scale assessments for students from diverse cultural and social backgrounds. This book represents the best thinking of scholars and researchers at this time. We hope it will encourage educational measurement and assessment specialists

to work toward common definitions (even if they represent a range of ideas as in Figure 1.1) and to continue and expand research on effective, fair, and valid culturally responsive assessment tools and practices.

References

- Aronson, B., & Laughter, J. (2016). The theory and practice of culturally relevant education: A synthesis of research across content areas. *Review of Educational Research*, 86(1), 163–206.
- Bennett, R. E. (2023). Toward a theory of socioculturally responsive assessment. *Educational Assessment*, 28(2), 83–104. <https://doi.org/10.1080/10627197.2023.2202312>
- Boodoo, G. M. (1998). Addressing cultural context in the development of performance-based assessments and computer-adaptive testing: Preliminary validity considerations. *Journal of Negro Education*, 67(3), 211–219.
- Castagno, A. E., & Brayboy, B. (2008). Culturally responsive schooling for Indigenous youth: A review of the literature. *Review of Educational Research*, 78(4), 941–993. <https://doi.org/10.3102/0034654308323036>
- Chavers, D., & Locke, P. (1989). The effects of testing on Native Americans. Retrieved from <https://eric.ed.gov/?id=ED336445>
- Dixon-Román, E. (2020). A haunting logic of psychometrics: Toward the speculative and indeterminacy of blackness in measurement. *Educational Measurement: Issues and Practice*, 39(3), 94–96. <https://doi.org/10.1111/emip.12375>
- Eels, K., Davis, A., Havighurst, R. J., Herrick, V. E., & Tyler, R. W. (1951). *Intelligence and cultural differences—a Study of cultural learning and problem-solving*. The University of Chicago Press.
- Estrin, E. T., & Nelson-Barber, S. (1995). Issues in cross-cultural assessment: American Indian and Alaska Native students. In *Knowledge brief* (vol. 12). Far West Laboratory.
- Evans, C. M. (2021). *Culturally sensitive, relevant, responsive, and sustaining assessment: Are there limits to making large-scale standardized testing culturally responsive?* (Blog). The National Center for the Improvement of Educational Assessment. <https://www.nciea.org/blog/culturally-responsive/culturally-sensitive-relevant-responsive-and-sustaining-assessment>
- Evans, C. M. (2023). Applying a culturally responsive pedagogical framework to design and evaluate classroom performance-based assessments in Hawaii. *Applied Measurement in Education*, 36(3), 269–285. <https://doi.org/10.1080/08957347.2023.2214655>
- Franco, M. P., Bottiani, J. H., & Bradshaw, C. P. (2023). Assessing teachers' culturally responsive classroom practice in PK–12 schools: A systematic review of teacher-, student-, and observer-report measures. *Review of Educational Research*. <https://doi.org/10.3102/00346543231208720>
- Freire, P. (2000). *Pedagogy of the oppressed, 30th anniversary edition*. Bloomsbury Academic. <http://www.amazon.com/Pedagogy-Oppressed-30th-Anniversary-Edition/dp/0826412769>
- Gay, G. (2000). *Culturally responsive teaching: Theory, research, and practice*. Teachers College Press.
- Glaser, R. (1963). Instructional technology and the measurement of learning outcomes: Some questions. *American Psychologist*, 18, 519–521.
- Gordon, E. W. (1977). Human diversity, program evaluation and Pupil assessment. *IRCD Bulletin*, 12(1), 2–8.
- Gordon, E. W. (1992). *Implications of diversity in human characteristics for authentic assessment (CSE Technical Report 341)*. National Center for Research on Evaluation, Standards, and Student Testing (CREST), University of California Graduate School of Education.
- Hambleton, R. K., Swaminathan, H., Algina, J., & Coulson, D. B. (1978). Criterion-referenced testing and measurement: A review of technical issues and developments. *Review of Educational Research*, 48(1), 1–47.
- Hood, S. (1998a). Culturally responsive performance-based assessment: Conceptual and psychometric considerations. *Journal of Negro Education*, 67(3), 187–196.
- Hood, S. (1998b). Introduction and overview: Assessment in the context of culture and pedagogy: A collaborative effort, a meaningful goal. *The Journal of Negro Education*, 67(3), 184–186. <https://doi.org/10.2307/2668187>
- Kelley, E. L. (1927). *Interpretation of educational measurements*. World.
- Kendi, I. (2019). *How to be an antiracist*. One World.
- Klenowski, V. (2009). Australian Indigenous students: Addressing equity issues in assessment. *Teaching Education*, 20(1), 77–93.
- Kūkea Shultz, P., & Englert, K. (2021). Cultural validity as foundational to assessment development: An Indigenous example. *Frontiers in Education*, 6. <https://doi.org/10.3389/educ.2021.701973>
- Kūkea Shultz, P., & Englert, K. (2023). The promise of assessments that advance social justice: An Indigenous example. *Applied Measurement in Education*, 36(3), 255–268. <https://doi.org/10.1080/08957347.2023.2222031>
- Ladson-Billings, G. (1995a). But that's just good teaching! The case for culturally relevant pedagogy. *Theory Into Practice*, 34(3), 159–165.
- Ladson-Billings, G. (1995b). Toward a theory of culturally relevant pedagogy. *American Educational Research Journal*, 32(3), 465–491. <https://doi.org/10.3102/00028312032003465>

12 • Culturally Responsive Assessment in Classrooms and Large-Scale Contexts

- Ladson-Billings, G. (2014). Culturally relevant pedagogy 2.0: A.K.A. the remix. *Harvard Educational Review*, 84(1), 74–85.
- Lee, C. D. (1998). Culturally responsive pedagogy and performance-based assessment. *Journal of Negro Education*, 67(3), 268–279.
- McCarty, T. L., & Lee, T. S. (2014). Critical culturally sustaining/revitalizing pedagogy and Indigenous education sovereignty. *Harvard Educational Review*, 84(1), 101–136.
- Miller-Jones, D. (1989). Culture and testing. *American Psychologist*, 44(2), 360–366.
- Mislevy, R. J. (2018). *Sociocognitive foundations of educational measurement*. Routledge.
- National Academies of Sciences, Engineering, and Medicine. (2018). *How people learn II: Learners, contexts, and cultures*. <https://doi.org/10.17226/24783>
- National Center for Education Statistics. (2024). *Digest of education statistics*. https://nces.ed.gov/programs/digest/d23/tables/dt23_203.65.asp
- National Research Council. (2000). *How people learn: Brain, mind, experience, and school: Expanded edition*. National Academies Press.
- National Research Council. (2001). *Knowing what students know: The science and design of educational assessment* (J. Pellegrino, N. Chudowsky & R. Glaser, Eds.). National Academy Press.
- Nelson-Barber, S., & Trumbull, E. (2007). Making assessment practices valid for Indigenous American students. *Journal of American Indian Education*, 46(3), 132–147.
- Nortvedt, G. A., Wiese, E., Brown, M., Burns, D., Mcnamara, G., Hara, J. O., Altrichter, H., Fellner, M., Herzog-punzenberger, B., & Nayir, F. (2020). Aiding culturally responsive assessment in schools in a globalising world. *Educational Assessment, Evaluation and Accountability*, 32, 5–27.
- O'Dwyer, E. P., Sparks, J. R., & Nabors Oláh, L. (2023). Enacting a process for developing culturally relevant classroom assessments. *Applied Measurement in Education*, 36(3), 286–303. <https://doi.org/10.1080/08957347.2023.2214652>
- OECD. (2016). *Global competency for an inclusive world*. Author.
- Paris, D. (2012). Culturally sustaining pedagogy: A needed change in stance, terminology, and practice. *Educational Researcher*, 41(3).
- Paris, D., & Alim, S. H. (2014). What are we seeking to sustain through culturally sustaining pedagogy? A loving critique forward. *Harvard Educational Review*, 84(1).
- Poe, M., Oliveri, M. E., & Elliot, N. (2023). The standards will never be enough: A racial justice extension. *Applied Measurement in Education*, 36(3), 193–215. <https://doi.org/10.1080/08957347.2023.2214656>
- Popham, W. J., & Husek, T. (1969). Implications of criterion-referenced measurement. *Journal of Educational Measurement*, 6(1), 1–9.
- Preston, J. P., & Claypool, T. R. (2021). Analyzing assessment practices for Indigenous students. *Frontiers in Education*, 6, 1–11. <https://doi.org/10.3389/educ.2021.679972>
- Qualls, A. L. (1998). Culturally responsive assessment: Development strategies and validity issues. *Journal of Negro Education*, 67(3), 296–301.
- Randall, J. (2021). “Color-neutral” is not a thing: Redefining construct definition and representation through a justice-oriented critical antiracist lens. *Educational Measurement: Issues and Practice*, 40(4), 82–90. <https://doi.org/10.1111/emip.12429>
- Randall, J., & Poe, M. (2021). Ain't oughta be in the dictionary: Getting to justice by dismantling anti-black literacy assessment practices. *Journal of Adolescent and Adult Literacy*, 64(5), 594–599. <https://doi.org/10.1002/jaal.1142>
- Russell, M. (2023). Shifting educational measurement from an agent of systemic racism to an anti-racist endeavor. *Applied Measurement in Education*, 36(3), 216–241. <https://doi.org/10.1080/08957347.2023.2217555>
- Sargent, S. S. (1952). Review reviewed work(s): Intelligence and cultural differences—a study of cultural learning and problem-solving by K. Eells, A. Davis, R. J. Havighurst, V. E. Herrick, & R. W. Tyler. *American Journal of Sociology*, 58(2), 209–210. <https://www.jstor.org/stable/2772197>
- Sireci, S. G. (2020). Standardization as UNDERSTANDARDIZATION in educational assessment. *Educational Measurement: Issues and Practice*, 39(3), 100–105.
- Smith-Maddox, R. (1998). Defining culture as a dimension of academic achievement: Implications for culturally responsive curriculum, instruction, and assessment. *Journal of Negro Education*, 67(3), 302–317. <https://doi.org/10.2307/2668198>
- Solano-Flores, G. (2011). Cultural validity in assessment—basic concepts. In M. R. Basterra, E. Trumbull & G. Solano-Flores (Eds.), *Cultural validity in assessment: A guide for educators* (pp. 3–21). Routledge.
- Solano-Flores, G., & Nelson-Barber, S. (2001). On the cultural validity of science assessments. *Journal of Research in Science Teaching*, 38(5), 553–573. <https://doi.org/10.1002/tea.1018>
- Stobart, G. (2005). Fairness in multicultural assessment systems. *Assessment in Education: Principles, Policy and Practice*, 12(3), 275–287. <https://doi.org/10.1080/09695940500337249>
- Taylor, C. S. (2022). *Culturally and socially responsible assessment: Theory, research, and practice*. Teachers College Press.

- Trumbull, E., & Nelson-Barber, S. (2019). The ongoing quest for culturally-responsive assessment for Indigenous students in the U.S. *Frontiers in Education*, 4, 1–11. <https://doi.org/10.3389/feduc.2019.00040>
- Trumbull, E., Sexton, U., Nelson-Barber, S., & Johnson, Z. (2015). Assessment practices in schools serving American Indian and Alaska Native students. *Journal of American Indian Education*, 54(3), 5–30.
- Waitoller, F. R., & Thorius, K. A. K. (2016). Cross-pollinating culturally sustaining pedagogy and universal design for learning: Toward an inclusive pedagogy that accounts for dis/ability. *Harvard Educational Review*, 86(3), 366–389.
- Walker, M. E., Olivera-Aguilar, M., Lehman, B., Laitusis, C., Guzman-Orth, D., & Gholson, M. (2023). Culturally responsive assessment: Provisional principles. *ETS Research Report Series*, 2023(1), 1–24. <https://doi.org/10.1002/ets2.12374>



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Section 1

Culturally Responsive Assessment

Theoretical Foundations and Concepts



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Building from Sociocultural Learning Theory to Culturally Responsive Assessment

Andrew D. Coppens and Rebecca Kelley

Building from Sociocultural Learning Theory to Culturally Responsive Assessment

The main purpose of this chapter is to articulate key features of sociocultural theories of learning and to position them in relation to the conceptualization, design, and use of classroom assessment practices. Although our discussion of sociocultural learning theory and its implications for culturally responsive assessment focuses mostly on classroom assessment practices, many of the implications of sociocultural learning theory also extend to large-scale educational assessment.

It will be helpful to state our view on relations between learning and assessment early and clearly, which is firmly rooted in sociocultural assumptions: learning and assessment are distinguishable but also inseparable and co-dependent processes in *any* educational activity (including large-scale assessment involving direct student participation). By co-dependent, we mean that assessment cannot take place without students *learning to* take part and *learning as* they take part in assessments. Similarly, learning cannot take place without the contribution of some form of assessment in supporting knowledge acquisition, motivation, and other related processes.

Conventionally, “learning” and “assessment” are terms signaling different activities in educational spaces. Learning, for example, is the ostensible purpose of teachers’ and students’ work together, and there are many ways to organize curricular activities to promote and support student learning (e.g., discussions, content presentations, collaborative group work). Assessment, again conventionally, refers to activities and processes designed to determine whether and to what extent learning has occurred (e.g., tests, quizzes, test-like questioning).¹ By contrast, a baseline contribution of sociocultural learning theory is that both within activities designed for learning and within activities designed for assessment, learning and assessment are present, active, and interwoven processes. Both are essential to the functioning of each activity (see Rogoff & Mejía-Arauz, [2022] for two contrasting cultural models of learning integrating assessment as a core process).

Due to institutional conventions regarding what formally “counts” as assessment activity in schools, this deep relation and interpenetration between learning and assessment may be

more obvious in informal ways of organizing learning (e.g., in socially collaborative work endeavors or play), where “assessment occurs in support of contributing to the activity, not for external purposes” (Rogoff et al., 2016, p. 360; see also Callanan et al., 2011). However, the lack of recognition of the pervasive role of assessment processes in all ways of organizing learning creates problems. As has been argued and documented at length, the hidden-in-plain-sight presence of learning-in-assessment and assessment-in-learning can make tacit social norms and trajectories for growth unnecessarily opaque, especially for students from marginalized or underrepresented cultural backgrounds (e.g., Delpit, 1995; Gutiérrez et al., 1995; Philips, 1983).

For example, Mehan (1979) and Cazden (1988) used sociocultural sensibility, as well as ethnographic methods, to “see” assessment in the form of IRE discourse deeply embedded in talk occurring during classroom activities and interactions designed for learning. They described a pervasive form of assessment embedded in everyday classroom talk between teachers and students: The I-R-E interaction sequence in which a teacher *initiates* interaction with students with a test-like known-answer question, seeks a *reply* that demonstrates students’ knowledge, and then *evaluates* the reply, often with contingent reinforcements such as “good” (positive) or “does anyone else know the answer?” (negative). In classrooms where this practice is common, students’ competent and willing participation in this discourse pattern is crucial to the validity of teachers’ understanding of students’ knowledge as well as to their appraisals of student behavior (e.g., cooperative, defiant, etc.).

The IRE classroom discourse pattern—at once a learning and an assessment practice—is well known and continues to be pervasive (Mehan & Cazden, 2015), as do its behaviorist motivational assumptions more generally. Although the IRE sequence is rarely taught to students explicitly, learning it and being willing to take part in it is a prerequisite to being seen and positioned as knowledgeable in such interactions. Relatedly, teachers deploy IRE and similar forms of talk as a means of organizing classroom interactions that are ostensibly about learning but that also reinforce control and asymmetrical power relations between teachers and students (see Paradise et al., 2014). Deployment of assessment practices is a significant political act in and beyond classrooms that has consequences for student equity. For this reason, the connection of assessment practices to a theory that can account for variation in cultural processes for learning across social groups is crucial. The sophistication of how sociocultural learning theory accounts for “context” can permit just this kind of connection.

Metaphorically, we think of the relation between learning and assessment as sourdough bread, which consists of mere flour, water, yeast, and salt. Outside of bread-making, these four ingredients are obviously distinct and separable. Yet, nutritionally speaking, their separation is a problem: Although each ingredient is essential to making bread, none can sustain human life on their own. The process of bread-making reorganizes their relations into a supraordinate unit of analysis (Cole, 1995) in which each must be understood in terms of other ingredients and in terms of a larger whole. Salt, for example, improves flavor but is also entailed in the proper regulation of fermentation. Similarly, yeast-driven fermentation “unlocks” the nutritional potential of flour. We offer (and have found in our research and teaching) that understanding sociocultural perspectives on learning involves a paradigm shift that often strays off the mark on this point: understanding learning in terms of activity-level relations among different components—in short, seeing learning as bread (see Rogoff, 2003 for a detailed discussion).

Underdevelopment in the understanding of relations between learning and assessment characterizes significant portions of educational theory and research. For example, the relatively recent National Academies consensus report *How People Learn II* (HPL II) (National Academies of Sciences, Engineering, and Medicine, 2018) adopted an explicit sociocultural perspective on learning. Yet just nine pages of HPL II were written with a focus on classroom assessment, and assessment was not a part of the forward-looking research agenda advanced

in the report. A *separate* National Research Council report, *Knowing What Students Know* (Pellegrino et al., 2001), focuses on the assessment of student learning. To be sure, we applaud the contributions of the HPL II report, which, for example, includes a conclusion consistent with the aims of this chapter: “When grounded in well-defined models of learning, assessment information can be used to identify and subsequently narrow the gap between current and desired levels of students’ learning and performance” (National Academies of Sciences, Engineering, and Medicine, 2018, p. 7). Nonetheless, such an imbalance may inadvertently advance a common view that assessment is an important but distinct process from learning (i.e., an assessment *of* learning view) rather than a perspective more closely aligned with sociocultural theory that assessment is a deep and essential aspect of the architecture of learning itself (i.e., an assessment *in* learning view).

The rest of this chapter is organized as follows. We first outline an overarching view of sociocultural theories of learning, a necessarily brief discussion. We offer justifications for taking up a sociocultural perspective on learning, highlighting its advantages for situating learning as more than an individual process of growth and knowledge accumulation and thereby permitting the inclusion of relational, institutional, historical, and political processes. We then outline an important sub-feature of sociocultural learning theory—artifact mediation—and discuss its implications for classroom assessment. Last, we discuss the degree to which formative assessment can be viewed as a socioculturally informed approach, as is commonly assumed.

What Is a Thoroughly Sociocultural View of Learning?

L. S. Vygotsky was a Soviet psychological scientist who is likely most associated with sociocultural perspectives on learning. Vygotsky’s work was a radically creative alchemy of Darwin and Marx, aiming to theorize the origins of human psychological functioning in the material realities and histories of community social life (see Daniels et al. [2007] and Newman & Holzman, [1993] for excellent discussions). His contributions continue to inspire contemporary social science theory in and beyond psychology (e.g., Cole, 1996; Nasir et al., 2020). Nonetheless, although sociocultural theories of learning are rooted in the contributions of Vygotsky and his students, many aspects of them have been further developed, reworked, and even refuted in the 90 years since Vygotsky’s death in 1934. Vygotsky died prematurely of tuberculosis at the age of 37, so even his own writing contained underdeveloped theoretical features. Moreover, Vygotsky, like any of us, wrote with the aim of contributing to the issues of his own time and place, which, at least in terms of his primary texts, are the proper context for interpreting their meaning.

A good example is the notion of learning as “internalization.” Vygotsky’s core insight was to locate the *origin* of knowledge in the social and cultural world; however, internalization as a *mechanism* of learning and cognitive development is problematic in its reification of dualities such as internal/external, mind/body, and individual/social (Packer, 1993). Vygotsky’s early notion of learning as internalization contradicts more recent commitments to theorizing the mind as *mutually constituted* by both individual and societal/cultural processes (Rogoff, 2003). Contemporary sociocultural learning theory conceptualizes learning as a fundamentally social and shared endeavor patterned via individuals’ engagement in cultural and institutional practices. We later discuss the learning-as-social-practice perspective.

This path of evolution in thinking regarding mechanisms of learning in sociocultural theory can be recapitulated by someone new to its concepts and assumptions, a helpful guide as several of its features (such as *mutual constituting relations* between individual and cultural processes) and histories (such as its roots in Soviet psychology) can seem more complex than, for example, cognitivist alternatives. Educators can also be guided in understanding sociocultural learning theory by the dilemmas they may encounter in practice despite much educational

discourse being rooted in pervasive cognitivist metaphors. For example, an internalization view of learning assumes that known information or skills are located “internal” to individuals, reducing the task of assessment to elicitation and performance measurement. Yet, many classroom teachers can readily identify students for whom this elicitation and measurement process is not straightforward—students who consistently demonstrate more extensive skills and knowledge in collaborative, practical activities than in individualized testing. The intractable assessment dilemma that the example of this student creates—“Is the problem with the test or with the student’s understanding?”—is rooted in an assumption of learning as internalized knowledge and personal know-how. Sociocultural learning theory provides an alternative path.

Sociocultural as a “Family” of Varying Commitments

At present, the term sociocultural does not refer to a single or unified theory of learning but rather to a “family” (Cole, 2005) of varying commitments to understanding both what learning is and what kinds of processes or factors are necessary to consider when researching or designing for learning. Although this plurality has advantages, we believe the term sociocultural has been overapplied, often used in reference to any consideration of contextual processes thought to “influence,” “impact,” or have an “effect on” learning processes internal to and primary in individuals. As an example of the varied use of the term, in 2020, the Society for Research in Child Development (SRCD) announced a *Sociocultural Policy* requiring authors to disclose:

(1) the dates of data collection (if applicable); (2) the theoretically relevant characteristics of the particular sample studied, for example, but not limited to: race/ethnicity, socioeconomic status, language, sexual orientation, gender identity (inclusive of non-binary options), religion, generation, family characteristics; and (3) the place(s) from which that sample was drawn, including country, region, city, neighborhood, school, etc. and all other context variables that are relevant to the focus of the publication. (<https://www.srkd.org/news/new-sociocultural-policy-enacted-across-all-srkd-journals>)

We strongly support the spirit of this policy, like many SRCD members. However, it is difficult to reconcile what “sociocultural” means in the context of this policy statement from the leading international professional society on childhood learning and development and the following statement from a leader in the development of sociocultural theory:

key theoretical insights that inform sociocultural work [include] its recognition of the centrality of culture in mediating psychological experience, its emphasis on the need for genetic and historical analyses of cognitive development, and its attention to cognition as it is embedded in culturally organized everyday activities.

(Cole, 2005, p. 1)

With us, Cole (2005) noted risks in the generic application of the term sociocultural related to the second two facets mentioned in the second quote: the loss of sensitivity to developmental (i.e., genetic) and historical aspects of learning and the danger of developing insights on learning that are unmoored from the everyday lives of people. Our view is that generic use of the term sociocultural also risks undermining Cole’s first facet of sociocultural work: culture as a *mediating* process in learning, which is distinct from an “influence view” (Rogoff, 2003) where flour, yeast, water, and salt “influence” each other but retain their status as individual ingredients. It is instructive to note that one could comprehensively conform a manuscript to SRCD’s *Sociocultural Policy* requirements without commitment to any of Cole’s (2005) three facets of sociocultural work.

Definition of a Sociocultural View of Learning

In the landmark 2003 text, *The Cultural Nature of Human Development*, Rogoff defines development as humans' "changing participation in the sociocultural activities of their communities, which also change" (p. 11). Although this is a definition of development,² it is consistent with canonical definitions of sociocultural perspectives on learning. For example, Lave and Wenger's (1991) treatise *Situated Learning* defines learning as "increasing participation in communities of practice" and an "an evolving, continuously renewed set of relations" (p. 49).

Contrast with a "Cognition Plus" View of Learning

Elsewhere, in the same year that *Situated Learning* was published, Lave (1991) contrasts sociocultural views of learning with two theoretical orientations to learning (see also McDermott, 1993). In the first contrast, Lave (1991) summarizes the "cognition plus" position as follows:

Researchers have for years analyzed the individual, internal business of cognitive processing, representations, memory, and problem solving, and . . . should now attend to other factors as well. People process, represent, and remember in relation to each other and while located in a social world. Therefore, researchers should extend the scope of their intraindividual theory to include everyday activity and social interaction. For proponents of [the cognition plus] view, social factors become conditions whose effects on individual cognition are then explored . . . A proponent of this position is likely to argue that a person thinking alone in a forest is not engaged in social cognition.

(p. 66)

Cole (1996) notes that the "cognition plus" view contextualizes or situates learning superficially, retaining the notion that individuals and the social world are fundamentally distinct (like is suggested by the concept of "internalization" as learning) even though they may "act on" each other. In other words, the view that sociocultural learning theory rejects is that context "surrounds" but is not integrated with individual learning or cognitive functioning or vice versa.

Approaches to educational assessment calibrated to a cognition plus view would be designed to elicit the knowledge, skills, and cognitive processes that students "have" or have developed. Arguing from a sociocultural perspective, Packer (2018) critiques the "cognition plus" view as naïve in its understanding of language and cognition as a largely individual process rather than fundamentally relational, cultural, historical, and political (all of which both transcend and include individual processes).

Contrast with an "Interpretative" View of Learning

In the second contrast, Lave (1991) summarizes an "interpretative view" of learning as arguing that learning takes place:

In a pluralistic world composed of individuals who have perspectively unique experience . . . meaning is negotiated, the use of language is a social activity rather than a matter of individual transmission of information . . . In this position there is no world independent of agents' construction of it.

(p. 67)

In this interpretative view, knowledge and the learning processes that support its growth are construed as fundamentally about individual and internal subjective meaning. Social and cultural aspects of learning, then, are understood as a matter of negotiating and renegotiating

personal understanding in the context of a separate-but-together jigsaw puzzle of interactive learning. Moreover, this view of learning requires culture to be conceptualized as a kind of individual cognitive process, a structuralist notion of culture that made important contributions to social anthropology but that may also be outdated (Lévi-Strauss, 1978).

Approaches to educational assessment grounded in this interpretative view allow for multiple ways of knowing and forms of knowledge, a plurality that would seem like a firm footing for developing assessment approaches responsive to a variety of cultural backgrounds. However, this interpretative view offers little to no imperative for conceptualizing learning as a process *made possible* by structures and processes “outside” of individuals’ efforts to understand and know, providing minimal implications or suggestions for curricular design.

Sociocultural as a “Situated Social Practice” View of Learning

Lave (1991) offers a third view—the “situated social practice” or “situated learning” view—which shares some features of the interpretative view but also makes important departures.

[The *situated social practice* view] emphasizes the relational interdependency of agent and world, activity, meaning, cognition, learning, and knowing. It emphasizes the inherently socially negotiated quality of meaning and the interested, concerned character of the thought and action of persons engaged in activity. But, unlike the first two approaches, this view also claims that learning, thinking, and knowing are relations among people engaged in activity *in, with, and arising from the socially and culturally structured world*. This world is itself socially constituted. Thus, from this point of view, “nature” is as much socially generated as afternoon tea. And its generation, according to this perspective, takes place in dialectical relations between the social world and persons engaged in activity; together these produce and re-produce both world and persons in activity.

(p. 67)

The paradigm shift of this “situated social practice” view of learning is that learning is not “influenced” or “shaped” by features of learning or assessment activities that are external to individuals’ processes of “learning, thinking, and knowing.” Instead, these processes are inherently social and relational, both *constituted by* learning and assessment activities and centrally *involved in constituting* those same learning and assessment activities. In this sociocultural view, Rogoff’s (2003) position that learning is the transformation of participation in cultural activities comes into clearer view. Rather than learning being understood as “internalizing” knowledge or skill from outside (e.g., from teachers or expert others) or internally making sense of an outside world, sociocultural perspectives understand learning as changes in forms of social and cultural participation.

Approaches to educational assessment firmly grounded in this situated social practice view—a perspective we argue is most faithfully “sociocultural”—would note problems with the individualistic and socially isolated arrangement of many educational tests. If learning is fundamentally a relational process, and if knowledge is also never merely a property of individual understanding (see Wells, 2008), why are most assessments of learning at any scale explicitly organized to impoverish social interaction?

Addressing “Culture” in Sociocultural Views of Learning

So far, our definition of sociocultural learning theory has focused on the social aspects of the term; attention also needs to be focused on the cultural aspect. Because cultural values are also inseparable from learning and knowing (see Sternberg, 2004, 2015; Sternberg & Preiss, 2022),