What Your Colleagues Are Saying . . .

Surfacing Brilliance Through Meaningful Science Assessment is the assessment guide we've been waiting for. Aneesha Badrinarayan offers a clear, compassionate, and visionary roadmap for designing assessments that affirm student brilliance, honor cultural assets, and spark real sensemaking. Grounded in both research and practice, this book translates equity into action—making it an essential companion for educators, leaders, and developers who believe science learning should be rigorous, relevant, and relentlessly student-centered.

—Linda Darling-Hammond
Charles E. Ducommun Professor of Education Emeritus,
Stanford Graduate School of Education
Stanford, CA

This book is grounded in "the fierce belief that all students are capable of brilliance." Badrinarayan's passion, expertise, and humility shine through her writing. She expertly balances providing practical strategies for assessment design and carefully crafted opportunities for deep personal reflection. Her thoughtful approach to assessment inspires educators to see and nurture the potential in every learner.

—Tricia Shelton
Chief Learning Officer, NSTA
Clarksville, TN

This book is an essential resource for anyone involved in designing science assessments. It presents a clear, accessible approach that makes complex concepts easy to understand and apply. Unlike other resources, it bridges the gap between theory and practice, offering valuable insights and practical strategies for realizing a better future for students through improved assessment practices. A must-read for science educators and designers.

—Daniel Alcazar-Roman
Associate Director, University of California Berkeley
Belmont, CA

Rooted in how we learn, Badrinarayan masterfully weaves together research, practical tips, classroom scenarios, design examples and non-examples in a book that is sure to become a foundational text for assessment developers, teachers, and anyone interested in creating meaningful learning experiences. Anchored in science, Badrinarayan's advice will engage, push, and excite anyone working to understand and build upon students' brilliance.

—Sasha Klyachkina
Skyline Assessment Manager, Chicago Public Schools
Chicago, IL

After years of wrestling with science assessments, this book is the one-stop shop we've all been waiting for. Whether you're designing tests, analyzing data, or looking to make assessments more equitable and engaging, this book covers it all. Packed with actionable insights, real-world examples, and practical tools, it's the ultimate guide to transforming every aspect of classroom science assessments. If you're serious about improving science education, start here!

—TJ Heck
Director, Learning Solutions Content, Science
Cognia, INC
Durand, MI

Badrinarayan is somehow both visionary and practical in ways that will challenge the reader and help them to advance their practice forward. This book will serve as an anchor point as we claw our way forward toward better teaching and learning experiences for all students.

—Matt Krehbiel
Senior Director, Development & Innovation, OpenSciEd
Derwood, MD

Surfacing Brilliance Through Meaningful Science Assessment is Aneesha Badrinarayan's "love letter to science assessment," and it is a refreshing and insightful read. Based on her two decades of expertise, she proves science assessment can be meaningful for students and approachable for educators. Her book will help science educators better understand assessment systems for the Next Generation Science Standards, including innovative uses of genAI, and put this knowledge into practice.

—Okhee Lee
Professor, New York University
New York, NY

Given contemporary expectations for students' science achievement, no one has thought more deeply about science assessment design and use than Badrinarayan. She elevates the discussion by emphasizing the role of assessment in surfacing student brilliance. She details what that implies and the many challenges that need to be overcome to attain that vision, while simultaneously providing examples of tasks and design principles to advance the effort.

—James W. Pellegrino
Emeritus Professor of Psychology and Learning Sciences,
University of Illinois Chicago
Chicago, IL

This book is designed to impact how we surface brilliance in our students, and I think at the same time it helps surface brilliance in its readers, making it both immediately useful and something educators will come back to again and again to spur important conversations about what science assessment can and should be now and into the future.

—Katie Van Horne
Principal Researcher & Founder, Concolor Research
Lynnwood, WA

This book shows what high-quality three-dimensional science assessment can and should be. In doing so, it sets the standard for all future development efforts. It summarizes, synthesizes, and builds on the varied efforts in the field, including the authors' own pioneering work. This book is a must read.

—Nathan Dadey
The National Center for the Improvement of Educational Assessment, Inc
Dover, NH

Badrinarayan is fierce in her belief in all students. This is a must-read for anyone interested in unlocking the transformative power of assessment. With her engaging tone, timely ideas, and relevant examples, Badrinarayan's contributions transcend science. This is a much-needed guidebook for addressing the enduring challenge of re-imagining assessment to better serve learners. The future is bright!

—Susan Lyons
Principal Consultant, Lyons Assessment Consulting
Wayland, MA

Surfacing Brilliance Through Meaningful Science Assessment

Dedication

For **Rishi** and **Aliya**. You are my daily reminders that the wonder that fuels learning is a gift, and being in a position to cultivate it—as your mom and your aunt, respectively, and as an architect of some small part of how you experience school—is an enormous responsibility. I hope that we succeed in building an education system—yes, even its assessments—that empower you to recognize and grow your unique brilliance. May school always be a place that ignites your love for learning, may assessments always make you feel capable and motivated, and may you always know that the true measure of your success is the one you set for yourselves (but also, Mama has some thoughts).

For **Nimai**. You were my first window into the brilliance of children. I am so proud of you.

For **GQ**. I'll never stop being grateful that you found me on the very day I entered the education world—it is the closest thing to proof of fate that I have ever experienced. I love you.

For passionate teachers, everywhere. You have the hardest, most important job. The world may not say it enough (or ever), but thank you for shouldering the awesome task of shaping the future. Where would we be without you?

Surfacing Brilliance Through **Meaningful Science Assessment**

Shifting Practice
to Reflect
Today's Students
and
Today's Standards

Aneesha Badrinarayan

CORWIN

CORWIN

FOR INFORMATION:

Corwin
A SAGE Company
2455 Teller Road
Thousand Oaks, California 91320
(800) 233-9936
www.corwin.com

SAGE Publications Ltd.
1 Oliver's Yard
55 City Road
London EC1Y 1SP
United Kingdom

SAGE Publications India Pvt. Ltd.
Unit No 323-333, Third Floor, F-Block
International Trade Tower Nehru Place
New Delhi 110 019
India

SAGE Publications Asia-Pacific Pte. Ltd.
18 Cross Street #10-10/11/12
China Square Central
Singapore 048423

Vice President and
Editorial Director: Monica Eckman
Senior Acquisitions Editor,
STEM: Debbie Hardin
Senior Editorial Assistant: Nyle De Leon
Production Editor: Tori Mirsadjadi
Copy Editor: Sheree Van Vreede
Typesetter: C&M Digitals (P) Ltd.
Cover Designer: Janet Kiesel
Marketing Manager: Margaret O'Connor

Printed in the United States of America

Paperback ISBN 978-1-0719-6808-6

This book is printed on acid-free paper.

25 26 27 28 29 10 9 8 7 6 5 4 3 2 1

CONTENTS

Acknowledgments xiii
About the Author xix
Author's Note xxi

PART 1: UNDERSTANDING SCIENCE STANDARDS IN FUTURE-FOCUSED WAYS 1

INTRODUCTION: Setting the Stage 2
- A Quick Primer on the Next Generation Science Standards 4
- Grounding Principles and Needed Shifts in Science Assessment 6
- This Is Not Your Typical Science Assessment Book 12
- Overview of This Book 14
- Language Use in This Book 15

CHAPTER 1: Future-Focused Science Assessments: Interpreting the NGSS and Similar Standards for Meaningful Assessments in a Rapidly Evolving World 17
- Understanding Our Foundation: Current Features of High-Quality Science Assessments 18
  - Landscape Shift 1: Local and Cultural Relevance and Personalization 19
  - Landscape Shift 2: Competencies and Durable Skills 24
  - Landscape Shift 3: Technology and Its Impact on Teaching, Learning, and Assessment in Science 28
- Future-Focused Science Assessment: Evolution 30
- Chapter Summary 31
- Reflection Questions 32

PART 2: FOUR PRIORITIES FOR SURFACING SCIENCE THINKING IN TRUSTWORTHY AND USEFUL WAYS 33

CHAPTER 2: Priority 1: Sensemaking 34
- Defining Sensemaking in Assessment 35
- Why Is Sensemaking So Important? 37

What Leads to Sensemaking in Assessment Tasks 40
Pathways to Sensemaking: Phenomena, Problematizing, and Prompts 44
Choosing a Phenomenon to Drive Sensemaking 44
Problematizing a Phenomenon to Surface Sensemaking 49
Crafting Prompts to Elicit Sensemaking 50
Scaffolding Sensemaking in Assessment Tasks 55
Deep Dive Into Choosing Meaningful Phenomena: Right-Sizing Phenomena and Problems to a Given Assessment Task 59
Using Phenomena to Understand Transfer and Generalizability of Science Learning 64
Caution: Ensuring Transfer Doesn't Become a Pass for Discrimination and Inequity 70
Chapter Summary 70
Reflection Questions 71

CHAPTER 3: Priority 2: Relevance 72
Making Assessment Tasks *Relevant*—Not Just Interesting—to Learners and Their Communities 73
Using Culture as a Concrete Path to Relevance in Science Assessments 79
Understanding Our Students: An Example 83
Attending to Culture and Relevance at the Scale of Many Students 88
Windows and Mirrors: A Strategy for Approaching Relevance at Scale 88
Features of Culturally Conscious Assessments 92
Motivation and Affect: Reimagining One of Assessment's Cardinal Rules 95
Chapter Summary 99
Reflection Questions 99

CHAPTER 4: Priority 3: Assets. 100
Looking for Assets in Student Sensemaking 102
Key Features of Asset-Based Assessment Practices in Science 107
Moving Beyond Misconceptions: Reframing Our Orientation to Wrong and Incomplete Answers 117
Implications of Asset-Based Approaches for Assessment Tasks and Data Dashboards 123

Habits and Beliefs of Asset-Based Assessors 125
Celebrating Assets of Students Who Think Differently Than We Do 128
Chapter Summary 139
Reflection Questions 140

CHAPTER 5: Priority 4: Opportunity and Experience 141
The Relationship Between Opportunity, Experience, and Performance 149
Making the Right Information Available 151
Impact on Students 154
A Window Into Practices and Durable Skills 159
What Experiences Should We Measure? 160
Turning Experience Information Into Insights About Opportunity to Learn That Influence Our Interpretations of Student Performance and Next Steps in Instruction 162
Using Opportunity and Experience Data to Foster New Conversations With Parents 168
Chapter Summary 171
Reflection Questions 172

PART 3: PUTTING THE PRIORITIES TO WORK 173

CHAPTER 6: Choosing and Designing Assessments That Center the Four Priorities 174
Reviewing Tasks: Criteria 175
Reviewing Tasks: Process 177
Some Tips to Help Quick Task Evaluation 179
Developing Assessments That Are Designed for the Science Standards in Ways That Center the Four Priorities 180
Step 1: Understand and Define Your Assessment Targets Within a Future-Focused Framework 181
Step 2: Unpack Standards and Other Learning Goals With Learners' Culture, Interests, and Identities in Mind 183
Step 3: Write Claims for Your Assessment That Synthesize Your Focal DCIs, SEPs, CCCs, and Durable Skills Into a Statement About Student Learning 191
Step 4: Choose a Phenomenon or Problem That Centers Specific Communities, Interests, and Reasons for Engaging 193

Step 5: Create a Problematized Scenario Based on the Phenomenon 202
Step 6: Develop Complete Student Explanations or Models That Specifically Account for How Different Perspectives and Ideas Can Be Valued in Complete Responses 203
Step 7: Develop Prompts That Elicit Sensemaking With Multiple Dimensions and Lived Experiences 204
Step 8: Develop Asset-Oriented Interpretation Guidance for Student Responses 205
Step 9: Invite Feedback and Review From Those Closest to Students You Are Least Connected To 206
Step 10: Pilot and Revise, Focusing on Learners at the Margins 207
Chapter Summary 207
Reflection Questions 208

CHAPTER 7: Using Generative AI in Responsible Ways to Support Science Assessment Design and Use 209
Defining Terms 209
Understanding the Current State of Generative AI in Science Assessment 212
Ways Generative AI Might Support Science Assessment 216
Approaching AI as a "Collaborator" in Science Assessment Activities 218
A Use-Case: AI in Science Assessment Review 220
Guiding Principles for Using Generative AI in Science Assessment Processes (AKA Aneesha's Laws of Assessment Machines) 224
Supporting Social Sensemaking 233
Some Final Caveats 237
Chapter Summary 239
Reflection Questions 240

CHAPTER 8: Epilogue: An Invitation to Imagine 241

References **245**
Index **251**

ACKNOWLEDGMENTS

I am so lucky to have "grown up" in education and assessment, and science assessment specifically, with so many mentors, teachers, and co-conspirators. Our community has been a living, breathing example of how much better our service to students can be when done in collaboration and with shared learning. I am a huge believer in social learning and sensemaking, and the ideas reflected in this book would not exist without the opportunity to think, figure out, and both fail and succeed forward with so many people—both those named here and the many, many others I have learned from and collaborated with over the last 20 years. I hope every one of you reads this book and sees it as the ultimate homage and flattery to *your* brilliance. I am so grateful to each of you. I am smarter and more compassionate about all of this work because of you.

Thank you to **Katie Van Horne, Jill Wertheim, T. J. Heck**, and **Sara Cooper** for being my partners in crime in all things science, assessment, parenting, and life. I don't think a single idea I've ever had about science assessment hasn't been made better by you. You are living, breathing examples of what it looks like to believe in the brilliance of every learner, and of measuring what truly matters.

So much of my thinking has been shaped in collaboration with state and district leaders. Thank you to **Daniel Alcazar-Roman, Jeff Grieg, Rachael Manzer, Dawn Cope, Ellen Ebert, Maya Garcia, Angela Landrum, Paul Dumas, Tyler Belanga, John Spiegel, Breigh Rhodes, Corrine Steever, Sonia Neuberger, Katie Bowler, Lizette Navarrete-Burks, Jill Grace, Maria Simani**, **Sasha Klyachkina, Peter Leonard,** and all of the SCASS, SPA-LC, and other state and local leaders over the last decade who have conspired with me to figure what on earth new science assessments should look like: For trusting me, for teaching me, for telling me when an idea was insane, for pushing me to center *your* students—what an honor it is to think and serve with you.

Thank you to **Nathan Dadey** and **Susan Lyons** for creating a space for me in the measurement world. You have been so generous with your time, your expertise, and most importantly, your friendship. I'm still not totally sure why you decided I'm worth the effort, but I am grateful each day that you're not tired of me yet.

Over the last decade in particular, I have been lucky enough to be part of a community of bold thinkers and humble doers who have enriched my ideas about expertise; have shown me what it takes to build agency and collective ownership in complex teaching, learning, and assessment systems; and have demonstrated what it looks like to support others in building their own capacity with true grace. **Dawn Novak, Kelley Aitken, Charlie Thompson, Paul Leather, Rita Harvey, Gretchen Morgan, Doannie Tran, Ellen-Hume Howard, Molly Minnick, Taunya Nesin, Charlie Brown, Chris Lazzaro, Ted Willard, Trish Shelton**, and so many other SPA-LC collaborators: Thank you for always pushing us all to think bigger and be bolder. I am so lucky you've let me be part of your circles and learn from you.

I have had so many mentors take me under their wing, offer advice, and open doors I never imagined walking through.

Linda Darling-Hammond, thank you for trusting me with your assessment vision and for showing me that it is actually possible to be a brilliant scholar, a caring leader, a savvy policymaker, and a visionary voice all at the same time—for being a concrete example of what it can look like to do, and be, it all. I am under no delusions that I will make even half the mark you have made on the education field, but I will be forever grateful and honored that you've taught and believed in me as we move the work forward together.

Okhee Lee, thank you for being my forever champion and for giving me firsthand experience in the power of high-expectations. **Jim Short**, thank you for being willing to really engage with a total newbie once upon a time, and for creating opportunity after opportunity for me thereafter. **Joe Krajcik**, you were one of the very first experts to engage with me seriously—thank you for being willing to engage with me then, and for every one of our spirited debates since. You have made me so much smarter. **Bill Penuel**, meeting you was a career-core memory—you have been the kindest guide, fiercest teacher, and most steadfast collaborator in this work. **Jim Pellegrino**, thank you for being so generous with your time and brilliance on nearly every science assessment project I've ever led—your insights have been invaluable to me, and I cherish every opportunity I have to learn from you. **Scott Marion**, I don't think I've ever been as nervous to meet an expert as I was when I reached out to you—and I am so glad that I got over my nerves. Thank you for supporting and believing in me, and for pushing every assessment idea I've ever had. **Erin Furtak,** your scholarship and grace have taught me how to lead with brilliance and care. **Elena Diaz-Bilello**, your thinking continues to expand mine—you help me reimagine what balance and

justice can truly mean in assessment. I am smarter at what I do every time I talk to each of you.

Not a single word in this book would have been possible without my O.G. Achieve community—truly the network that never stops giving. Thank you to **Teresa Eliopoulos** and **Molly Ewing** for helping me navigate a completely new-to-me world when I first started. Thank you to **Izzy Huston and Vanessa Wolbrink** for being my perfect complements in those early days. Thank you to **Matt Krehbiel** for being the other half of our science scheming and dreaming for years, and for being an example for what it looked like to lead forward. Thank you to **Miah Daughtery** and **Ted Coe** for always bringing me along as the science counterpart to your math and English language arts brilliance (and for animal hour, the best hour of the day). Thank you to **Jacob Mishook** for guiding me along the assessment and accountability path when all I had was a love of science and a zest for learning more. Thank you to **Jenn Sattem** and **Mike Cohen** for your unwavering support of me as I sought to move into new spaces.

Thank you to **Jennifer Self** and **Stephen Pruitt** for taking a leap of faith by hiring me, for your steadfast support as I figured out what it meant to make the bench science-to-science education policy leap, and for believing I could do things that, honestly, you (nor I!) had any reason to believe I could do. Jennifer, your trust in my ability to learn and make a meaningful contribution to a field I was so new to set my career on a completely different trajectory than any I had imagined. Stephen, you once told me you had a plan for me to move into assessment, before I had any idea of what that would entail—how did you know? Thank you for your mentorship at and since Achieve. I am so lucky you are in my corner.

Thank you to **Jamie Ricci, Ramya Kasturi, Caitlin Orsini, Ann Hernandez, and Rachel Gershman** for being anchor examples of what brilliance looks like, and the different ways it can manifest. I love you all.

Thank you to my parents **Renu** and **Badri**, my sister **Reema,** and my grandmother (everyone's **Paati**) for being my earliest teachers. Thank you to my amazing extended family for always believing I can do anything. My aunts, uncles, cousins, and in-laws: I have been so lucky to have you as cheerleaders and support throughout the winding path that's gotten to me to today.

And perhaps most importantly, thank you to **Dan Gershman** for being the best partner in all things. I'm sorry there are no sexy vampires in this book, but hopefully you think it's worth it all anyway.

PUBLISHER'S ACKNOWLEDGMENTS

Corwin gratefully acknowledges the contributions of the following reviewers:

Kristina Fulton
Curriculum & Instruction Specialist, Lancaster-Lebanon Intermediate Unit 13
Lancaster, PA

Casey McCormick
Grades 5–8 Advanced Math Teacher
Citrus Heights, CA

Alicia Burdess
Numeracy Lead Teacher and Instructional Coach, Grande Prairie and District Catholic Schools
Alberta, Canada

Enya Granados
High School Life Science Teacher, Clarke County School District
Athens, GA

John F. Mahoney
Member, National Teachers Hall of Fame
Past member, College Board's AP Statistics Committee and AP Calculus Committee
Bethesda, MD

Dawn Novak
MSED and TLEP Adjunct Faculty, Northwestern University
Grayslake, IL

Kelley Aitken
K–12 Science Supervisor, Winchester Public Schools
Martinsburg, WV

TJ Heck
Director, Learning Solutions Content, Science
Cognia, Inc.
Durand, MI

Daniel Alcazar-Roman
Associate Director, University of California Berkeley
Belmont, CA

Katie Van Horne
Principal Researcher & Founder, Concolor Research
Lynnwood, WA

Toni B. Ramey
Teacher, Mobile County Public School System
Mobile, AL

ABOUT THE AUTHOR

Aneesha Badrinarayan is a principal consultant at Education First, where she supports states and districts in developing meaningful teaching, learning, and assessment systems across content areas and strategies. Prior to joining Education First, Aneesha led assessment initiatives and strategy at the Learning Policy Institute and Achieve. She is a national expert on instructionally relevant assessment systems, high-quality assessments, and science teaching, learning, and assessment. Over the last 20 years, Aneesha has led work related to high-quality instructional materials, assessment system development from classroom assessments to state assessments to NAEP, and professional learning. Trained as a behavioral neuroscientist, Aneesha comes to education policy via informal science education and bench science. She received her B.A. in biology from Cornell University and her M.S. in neuroscience from the University of Michigan.

Aneesha lives in Maryland with her husband Dan and their fabulous toddler Rishi. When she isn't playing the role of science assessment superhero, you can find her alter ego with her nose in a romance novel, her hands whisking something in the kitchen, and the rest of her body otherwise occupied with slowly transforming her house into a toddler play place (who needs a dining room, really?).

AUTHOR'S NOTE

I'm finishing edits on this book in spring 2025, in the middle of what can only be described as a tumultuous time for both science and education in the United States. Amid growing conversations about eliminating initiatives designed to lift up communities that have been systematically denied opportunity and access again and again, denying science ideas for which there is abundant evidence, and questioning both the worth and work of science in our local and global communities, I can see those in my circles being made smaller by the realities of our world. My friends, I cannot think of a time when it has been more important to figure out ways to value our humanity and sophisticated science understanding together—to model how perspective-taking and caring discourse makes us smarter, to seek evidence of brilliance in the diversity of our lived experiences, and to arm our youth with fully earned confidence in their ability to make meaning—using robust science ideas and practices!—where it seems like there is only chaos. Teachers, leaders, researchers, developers: The work you do to see the potential in every learner and help them meet it is so important. We need a generation of mentors shaping a generation of thinkers and doers—those who will see opportunities for economic mobility as part of the same system as reducing impacts on climate; who will understand how to look for cause versus correlation when presented with patterns in linking environmental conditions with disease rates; and who will understand that rash and inhumane changes made now will have consequences at tremendous scales, over time. It's just good science.

PART 1

UNDERSTANDING SCIENCE STANDARDS IN FUTURE-FOCUSED WAYS

Introduction: Setting the Stage

Chapter 1: Future-Focused Science Assessments: Interpreting the NGSS and Similar Standards for Meaningful Assessments in a Rapidly Evolving World

INTRODUCTION

Setting the Stage

My toddler's school recently sent out a note asking for parents to come in and talk about their careers. When talking with the other parents, we joked about who had the most 3-year-old appropriate jobs: A doctor was a no-brainer, and given our kids' shared obsession with space, my NASA scientist husband was volunteered by the group almost immediately. When it was my turn to offer up my chosen career—to work on meaningful assessment systems—you might have thought I said I spend my days stealing puppies or creating a particularly vicious poison. Each person had an immediate reaction to the word "assessment," and they painted a pretty bleak picture.

> "Ugh tests—I bet you don't even want to tell other parents you work on that!"
>
> "Good grief, I'm getting anxious flashbacks just thinking about tests!"
>
> "That would be really depressing to have to go tell kids your job is to figure out what they're bad at!"
>
> "Ha, talk about wasting time on things that literally never mattered again—just ask me how many times I've needed to be able to recite the periodic table in my job selling insurance!"

My friends' experiences with assessment were unsurprising. Most people in the United States have overwhelmingly experienced assessment as demoralizing, dehumanizing, and disconnected from anything that really mattered to them—including, and maybe most disturbingly, learning.

Many scholars and assessment experts lament that the shift toward assessments that are increasingly decoupled from meaningful learning means that assessments can't help shape learning. I fear that what we are seeing is worse than "no impact": It's *negative* impact. Assessments are being used to encourage "drill and kill" memorization over deep inquiry and practicing the kinds of limited performances students see on tests over rich and extended engagement with content situated in real-world contexts. Assessments are also limiting which students see themselves as knowers, doers, and learners—just consider how many people self-report that they are "not a science person" or "bad at math" based on grades and test scores.

These kinds of assessment experiences took root particularly easily in content-rich disciplines like science and social studies, where it is easy to imagine getting useful information about students' learning by asking about discrete facts, like the phases of mitosis or steps of the Kreb's cycle. As these kinds of assessments have persisted, so too have the associated teaching and learning practices and student experiences: The easiest and most time-efficient way to recite the phases of mitosis is to memorize them, not to deeply understand how cell division and genetic information transfer leads to tissue growth.

The release of A *Framework for K-12 Science Education* (National Research Council [NRC], 2012) launched a movement in science education, including the development of completely new state standards that described a new vision for what we hope students will know and be able to do in science: one that centers on figuring out meaningful, real-world phenomena and addressing problems that matter to students and the communities they are a part of. For this vision to become lived experience for every student, we need to reimagine what our assessments look like, how they are experienced, and what we do with them. This is about more than needing new assessments to match new standards—we need new assessments to galvanize a *movement* in science education, one that reclaims love of the natural world and impossibly brave imaginations of what the future can look like from the claws of bolded words in a textbook.

This book is my love letter to science assessment—what it can be, and what it needs to be to support our youth in embracing science as a part of their daily lives and the solutions of a better future. I focus on the major lessons learned in science assessment since the release of the *Framework* and standards based on it, like the Next Generation Science Standards (NGSS). Importantly, I've tried to focus on the highest leverage features for assessing student learning in science, strategies that teachers can implement right now in their classrooms to better signal and support ambitious goals for science learning.

This book is my love letter to science assessment—what it can be, and what it needs to be to support our youth in embracing science as a part of their daily lives and the solutions of a better future.

A QUICK PRIMER ON THE NEXT GENERATION SCIENCE STANDARDS

The NGSS were a collaborative effort across 26 lead states to develop science standards that truly reflected what students need to know and be able to do in science (NGSS Lead States, 2013). The NGSS were themselves based on *A Framework for K-12 Science Education,* a 2012 NRCl consensus report that detailed the research base for how students learn science, what science teaching and learning should look like, and what content and practice should be prioritized. Nearly every state in the United States has adopted standards based on the *Framework*, either NGSS verbatim or something that looks quite similar.

A hallmark of both the *Framework* and standards based on its recommendations—and something central to new science assessments—is the expectation that *students will develop and use the three dimensions of science education together to make sense of phenomena and address problems.* Figuring out how to operationalize three dimensions (Table I-1) and their use together has been the primary work of curriculum developers, assessment developers, professional learning providers, educators, and leaders since the release of the NGSS in 2013.

Table I-1 • *Operationalizing Three Dimensions in Science Education*

Dimension	What it is	What this looks like in the NGSS
Science and engineering practices (SEPs)	The skills, behaviors, and knowledge scientists use as they investigate and build models and theories of the world.	The NGSS expects students to regularly engage in eight practices with in increasingly sophisticated ways over the course of grades K–12: **1.** Asking Questions and Defining Problems **2.** Developing and Using Models **3.** Planning and Carrying Out Investigations **4.** Analyzing and Interpreting Data

Dimension	What it is	What this looks like in the NGSS
		5. Using Mathematics and Computational Thinking **6.** Constructing Explanations and Designing Solutions **7.** Engaging in Argument From Evidence **8.** Obtaining, Evaluating, and Communicating Information
Disciplinary core ideas (DCIs)	A limited set of science conceptual ideas that meet at least two of the following criteria: **1.** Have broad importance across multiple sciences or engineering disciplines, or be a key organizing principle of a single discipline. **2.** Provide a key tool for understanding or investigating more complex ideas and solving problems. **3.** Relate to the interests and life experiences of students or be connected to societal or personal concerns that require scientific or technological knowledge. **4.** Be teachable and learnable over multiple grades at increasing levels of depth and sophistication. That is, the idea can be made accessible to younger students but is broad enough to sustain continued investigation over years.	In the NGSS, the core ideas are represented across four domains: ▸ Life Sciences ▸ Physical Science ▸ Earth and Space Sciences ▸ Engineering Each domain is divided into two to four major areas, with specific conceptual expectations defined for each.

(Continued)

(Continued)

Dimension	What it is	What this looks like in the NGSS
Cross-cutting concepts (CCCs)	Concepts that have application across all science domains.	The NGSS expects students to develop and demonstrate seven cross-cutting concepts with increasing levels of sophistication across grades K–12: **1.** Patterns **2.** Cause and Effect **3.** Scale, Proportion, and Quantity **4.** Structure and Function **5.** Energy and Matter **6.** Systems and System Models **7.** Stability and Change

The NGSS specifies a set of expectations for each dimension at every grade band (NGSS, 2013, Appendices E, F, and G), as well as provides examples of what students should know and be able to do in the form of the NGSS performance expectations. We'll dive more into how the structure of the NGSS relates to assessment later in this book, but a major headline is that NGSS assessments must surface what students know and can do with elements of all three dimensions in service of making sense of phenomena and addressing problems.

GROUNDING PRINCIPLES AND NEEDED SHIFTS IN SCIENCE ASSESSMENT

When the NGSS were first released, the field was inundated with talk about "shifts": What were the major conceptual shifts of the standards? What were the needed instructional shifts in the classroom? What were the biggest shifts we wanted to see in instructional materials? In professional learning?

While huge collaborative efforts were underway to change the way teachers and leaders viewed the teaching of science, there was much less clarity about what this meant for assessment. Stephen Pruitt, one of the authors of the

Framework and a major architect of the NGSS, famously said that a goal of the NGSS was to "break assessment"—but for many years, the field of science assessment sought to do the exact opposite, and continued to force science assessments to play by the same rules assessment had been playing by for the last 30 years.

Unsurprisingly, that didn't work. Time and time again, educators and leaders would see "new" science assessments and say, "That's not it." They told us that the assessments were disconnected from what teachers were trying in the classroom—at best, this made the assessments somewhat useless, but at worst, the assessments were being used to limit or even reject better teaching and learning practices that were having real impact on students.

By working directly with teachers and leaders, researchers and curriculum and assessment developers, it became clear that a few foundational principles needed to ground our work in science assessment. These principles have emerged from work in science, but they hold true for assessment more broadly as well:

- **Principle 1. All students are capable of brilliance.** So many traditional assessment instruments in the United States have been designed around the central idea that only some students can succeed. We see this in our obsession with labeling students as "basic" or "not proficient," in red-yellow-green reporting dashboards that focus on which standards students are failing at, and in tests designed to be "gotcha" measures that seek to identify failure points rather than learning edges. We need to stop designing and using science assessments that operate from a perspective of failure, and instead, we need to begin with an assumption that every student is brilliant and can find success. We see this in assessments that:
    - give students an opportunity to show what they *do* know and *can* do,
    - focus on providing information about how to support learners in taking their next steps in learning,
    - recognize diverse ways of knowing and thinking as assets in science learning, and
    - provide ways for students to demonstrate progress toward shared learning goals (i.e., standards) while highlighting unique gifts they have to offer.

We need to stop designing and using science assessments that operate from a perspective of failure, and instead, we need to begin with an assumption that every student is brilliant and can find success.

- **Principle 2. Assessments communicate power and value.** Assessments communicate more than information about what students know and can do—they communicate what a system values. We hear this every time a teacher feels compelled to change their instruction because of what is on "the test." By putting something on an assessment, we are saying, "This matters—and this matters more than something that is not being prioritized on this assessment." Because assessments ultimately convey values, they are tools of power—who gets to decide the content of an assessment is a sign of who has power in systems.

  We can use this knowledge of how assessments communicate who is in power and what they value to our advantage, by being intentional about how assessments are developed, who gets a say in what they look like and how they are used, and what the actual content and design of the assessment looks like.

- **Principle 3. Assessments must be coherent with learning**. Assessments are only useful and trustworthy if they mirror and support how students learn. If students are learning in deeply collaborative ways that privilege meaning-making, knowledge-in-use, and contextualized knowledge and practice development (as we hope science students are), the most productive assessments will surface how their thinking is progressing in similar ways. This means assessments that specifically focus on (a) what and how students have had opportunity to learn, and (b) surfacing students' learning edges so that instruction can meet them where they are at and support their next step in learning.

  One question I often hear is what to do when instruction has not yet shifted to the kind of ambitious teaching and learning the NGSS demands—if students are still experiencing "sit and get" instruction, should we use new science assessments or the old ones? This question deserves careful attention, and we'll discuss this issue more throughout the chapters in this book focused on NGSS assessment tasks and practices, but for now I want to say this: If your assessment truly embodies how we want science teaching, learning, and performance to look and feel, the answer to incoherence with instruction is not to change the test—it's to use student performance on the assessment to reflect on and change *instruction* rather than to judge students. Better yet, incorporating meaningful science assessment practices may itself be a path to improving science instruction!

- **Principle 4. Students lived experiences and cultural assets are a necessary ingredient in effective assessments.** This is really more like Principle 3.5. We know a great deal from brain and cognitive science, as well as from practical research, about how students learn. We learn *everything* by activating and building on what we already know. This is rooted in our lived experiences—including both our experiences from schooling as well as the vast amount of knowledge and practice we have built up and cultivated in our social and cultural interactions. For assessments to truly surface what students know and can do, and provide insight into how we support learning moving forward, they must attend to and account for the cultural and linguistic assets students bring to the table.

For assessments to truly surface what students know and can do, and provide insight into how we support learning moving forward, they must attend to and account for the cultural and linguistic assets students bring to the table.

- **Principle 5. Assessments must have a "net positive" impact on teaching and learning.** A lot of assessment decisions get justified as "necessary evil." I honestly find this notion astounding—what could possibly make an assessment practice or test we would label "evil" necessary? I'm not suggesting that every assessment has to be something every student is excited to participate in—as lovely as that would be, sometimes we do have to do things we might prefer not to. But assessments that demean, demoralize, dehumanize children? That deprofessionalize teachers? That deny students opportunity for meaningful and rich learning because something more narrow and superficial "is what we need to do to score well on the test"? There is nothing "necessary" about those kinds of assessments. We can have valid and reliable measures that help us support every learner without doing harm.

Given this foundation, we can now begin to dive specifically into the features of science assessments, as well as the major shifts and non-negotiables that distinguish high-quality NGSS assessments from other assessments you might find.

	Shift *away* from science assessments that look like . . .	**Shift *toward* science assessments that look like . . .**
Phenomena and problems	▸ Questions that focus on decontextualized science facts ▸ "story problems" that include a context as an interesting but unnecessary "hook"	▸ Tasks centered around a specific event or occurrence ▸ Tasks in which the scenario is necessary for completing the task
Multiple, integrated dimensions	▸ Questions about science ideas OR science practices ▸ Focus on isolated factoids or specific lab skills	▸ Asking students to demonstrate conceptual understanding through engagement with a science practice
Sensemaking	▸ Questions that focus on discrete science topics ▸ Tests and tasks that assume only advanced students can apply learning ▸ View scaffolding as only appropriate in instruction, not assessment ▸ Strict boundaries between assessments "for" learning and assessments "of" learning	▸ Tasks that ask all students to demonstrate what they know and can do by figuring out something about a phenomenon or problem ▸ View appropriate scaffolds as an opportunity to guide and reveal scientific sensemaking ▸ View all assessments as opportunities to learn and make thinking visible
Multimodal	▸ Primarily comprising multiple choice and written questions	▸ Recognize and value many different modalities for both the task and student responses, including selecting, writing, speaking, drawing, photographs, models, simulations, etc.

	Shift *away* from science assessments that look like . . .	Shift *toward* science assessments that look like . . .
Windows and mirrors	▸ Assume all students interact with every assessment in the same way ▸ Seek neutrality—neutral contexts, no emotional engagement with the tasks	▸ Recognize that every student *will* interact with a given task in different ways, and design assessments such that they act as windows for some students and mirrors for others ▸ Seek to engage *productive* affect—tasks seek to engage all students without traumatizing anyone
Progressions	▸ Assessments reveal whether students have achieved grade-level standards ▸ Focus on end-of-instruction target alone	▸ Assessments reveal how students are thinking along within- and across-year learning progressions
Disciplinary literacy	▸ Assessments seek to limit reading, writing, speaking, and listening as “not part of science”	▸ Assessments recognize that literacy elements are expected parts of science performance, and seek to support students in engaging with texts, discourse, and writing in science

Taken together, these shifts have some important implications for what we should expect of new science assessments.

1. **The assessments are going to look different.** Sometimes, they will be so indistinguishable from good instruction it might be confusing to even figure out where the "assessment" is, and what to do about it. End-of-unit assessments might look like one coherent task or project that samples across many standards, rather than like a test that "covers" every fact from the unit. Some assessments might take days for students to complete, and others might involve iteration and revision throughout learning. Almost all assessments will actually be learning experiences in their own right.

2. **More emphasis on student work analysis and feedback than on evaluation and grading.** NGSS assessments take learning, growth, and the brilliance of every learner seriously—and nothing quite stifles learning as much as grades and scores. This doesn't mean that we shouldn't grade or score assessments, or that those can't be important and communicative—but most NGSS assessment resources focus on interpreting student thinking and planning forward from where students are.
3. **Understanding good science assessment practice should also mean understanding better teaching and learning practices.** Nearly all the shifts described above—emphasis on sensemaking, the centrality of phenomena and problems, attention to students' cultural and linguistic assets—are the same as what we want to see in instructional shifts. This means that learning how to assess the NGSS doesn't need to be separate from, or compete with, learning how to teach the NGSS. These efforts should be mutually reinforcing and coherent.

THIS IS NOT YOUR TYPICAL SCIENCE ASSESSMENT BOOK

Leading the collaborative development of most of the foundational criteria and review processes used to determine the quality of new science assessments means I have probably looked at thousands of NGSS assessment tasks over the last decade, developed by researchers and assessment vendors, NGSS writers, and newly minted teachers. We have come a long way since the tasks that popped up in 2014—the field has learned so much about how to assess science learning. There are guides for how exactly to write NGSS assessments (e.g., Harris et al., 2024), about how to use formative processes in science classrooms to assess students in real time, and about how to interpret student thinking relative to performance expectations (Furtak, 2023; Wertheim, 2024; Wertheim et al., 2023). And people are putting these ideas into practice—my colleagues in the field have led countless assessment development efforts over the years, seeking to put their hard-learned lessons about how to write 3D science tasks into practice for educators, assessment vendors, and curriculum developers. Yet, every day I look at assessment tasks that emerge from these processes and think, "Well, this isn't quite what we want to see." Nearly every expert review panel I have led has seen far more assessment tasks that miss the mark than ones that approach what our shared vision for science learning looks like. I hear this echoed every time I work with educators and leaders too—they look at technically well-developed assessment tasks or items and say, "This isn't what we want science performance to look like." Then, they follow the careful development processes we as a field have laid out and come up with tasks that are extremely similar to the ones they rejected as not good enough.

Why does this keep happening?

When writing this book, I did what I often have teachers, leaders, and developers do in my professional learning workshops: I took a wide sample of tasks—along with notes about their development processes and implementation efforts, conversations with students and teachers about their reactions, and examples of student work and educator next steps—and laid them out in ranked order from "least like what we want to see in science assessment" to "most like what we want to see." Then I set about the process of making sense of the patterns I was seeing—what was it about the tasks themselves, their development, and their implementation that led to tasks that were closer to one end of the spectrum versus the other? Why was it possible to follow really high-quality development processes—the same ones I have used countless times—and sometimes get excellent results and sometimes get terrible ones? Was it really true that there was just some kind of magic some assessment writers have and others don't, as I heard so often from my colleagues when trying to make sense of this together?

As I went through this process, I realized that in an effort to make extremely practical the processes and lessons learned through a decade of assessing the NGSS, we have missed some of the most fundamental aspects of assessment development: our orientations, assumptions, and trade-offs. These were the things, time and time again, that distinguished meaningful tasks from ones that checked all the boxes and still fell short. The bright side is that if we center these priorities and values, it is much easier to achieve greatness in science assessment design and use—this is a much more manageable path for every teacher, leader, and developer to pursue than needing to invest 40+ hours in every assessment development effort, to get a somewhat mediocre task out of it.

That's why this book focuses on a slightly different frame of science assessment.

Of course, there are chapters on the specific shifts for assessment (Chapter 1), how to choose and design meaningful multidimensional tasks (Chapter 7), and how to make sense of student learning (Chapter 5). But this book is built on the assumption that if we want assessments to help us win the long game, we need to take seriously an orientation to student learning and performance that privileges motivation, engagement, critical thinking, and meaning-making. I am very sure that not being able to recall the diagram of mitosis or knowing exactly how to calibrate a scale is never going to be something we regret outside of the walls of a biology (or maybe chemistry) classroom—and, when needed, I'm also very sure that any student can learn those things, if we do our jobs right. But if they never find themselves in the situations where they need that knowledge at all because they have systematically avoided science throughout their K–12 and postsecondary careers—because their assessments, grades, and scores perpetually told them "this is not for you"—I think our hearts will be heavy with that burden for the rest of our days.

If we want assessments to help us win the long game, we need to take seriously an orientation to student learning and performance that privileges motivation, engagement, critical thinking, and meaning-making.

Instead, let's use sophisticated, concept- and practice-rich assessments that push students to excel and grow. Let's use assessments to develop and demonstrate that our learners are the most scientifically savvy generation to emerge from K–12 schools yet. But let's do assessment in ways that empower learners, light that inner fire to learn more and do more, and help them imagine a future where they are an essential part of the tapestry that is science.

OVERVIEW OF THIS BOOK

This book sets out a roadmap for this kind of orientation to science assessment. First, let's understand current science standards, and how they relate to the kinds of durable skills and competencies future-focused educational efforts are working on. Then, let's dive into four priorities for approaching science assessment in ways that honor learning for all and build on each other: sensemaking, relevance, assets, and opportunity. Finally, let's consider how to do this in practice: how we select, adapt, and design assessments that do these things, and how we might make use of technologies like artificial intelligence to bring these core values to life more readily in our assessment processes. Because I'm focusing on those most fundamental features of assessment design, it is my sincere hope that this book is useful to everyone working to serve K–12 students, from kindergarten teachers to those teaching Advanced Placement (AP) and International Baccalaureate (IB) classes to high school seniors, and from teachers and coaches to assessment developers.

LANGUAGE USE IN THIS BOOK

Words matter. To the best of my ability, I've tried to use language in this book that is clear, direct, asset oriented, and tends to be preferred by members of the communities I am talking about at any given moment. For example, some things you will see in this book:

- I use "disability" instead of more indirect terms to describe students with physical, mental, cognitive, or developmental conditions that interfere with one or more major life activities (Americans With Disabilities Act, 1990).
- In some instances, I use language that is specific to how U.S. education law categorizes and supports learners because that categorization is relevant to both case-making for support and the direct supports available. In other instances, I use language that is consistent with scholarship in a given area or with a particular community—in these instances, I have tried to privilege the language used within inclusive research–practice partnerships where research has been conducted with, not to, teachers, learners, and the communities they are part of.
- I have capitalized terms that indicate racial and ethnic identities (e.g., Black, LatinX).
- I use "Indigenous" most often in the book because much of the work and scholarship I draw on emerges from many different communities from within and outside the continental United States, each of whom have unique identities, communities, and language heritages. Where an example draws from a more specific geographical and Nation/Tribal identity, I try to be more specific in my language use as well.

At the same time, language about identity is deeply personal *and* influenced by context. As a result, it is entirely possible that some readers or people will disagree with the language choices I've made here. If you disagree with my language choices here, then I am wrong in how I am describing your experiences. I welcome the opportunity to learn forward and be better about using words that invite you into the conversation.

I also want to note that I use "NGSS" here largely as a shorthand for multidimensional standards based on *A Framework for K–12 Science Education*. If you have standards that include DCIs, SEPs, and CCCs, welcome, this book is for you—and honestly, even if you don't, I think these ideas are still important, no matter what your standards framework looks like.

Assessment is rapidly evolving—it's one of the reasons I have loved working in this field and have been so honored and humbled by the minds I get to work with. At the same time, no matter how much technologies and ideologies about what and how we assess change, there are some things I am very sure are constant. If students hate science, they will not learn it. If students think they are innately bad at science, they will not learn it. If students don't see the value and magic in science, they will not learn it. And if students don't learn how to think like scientists—in the broadest, most inclusive definition of the term—they will not be able to meet their potential. They simply will not be the best versions of themselves as the future's scientists but also as the future policymakers, advocates, community organizers, inventors of the yet-to-be-imaged, and champions of responsible, balanced, creative solutions. And we need them.

CHAPTER 1

FUTURE-FOCUSED SCIENCE ASSESSMENTS

Interpreting the NGSS and Similar Standards for Meaningful Assessments in a Rapidly Evolving World

When the Next Generation Science Standards (NGSS; NGSS Lead States, 2013) were first released in 2013, they were a revelation. They turned science education into a future-focused pursuit, setting forth expectations for higher order thinking in science, and a roadmap to getting there via carefully constructed expectations for the integrated use of science and engineering practices (SEPs), disciplinary core ideas (DCIs), and cross-cutting concepts (CCCs). The standards, and the research base (National Research Council [NRC], 2012) it grew from, acknowledged that the *goal* was a critically thinking, problem-solving, scientifically literate generation—and that the way to get there was not to ignore content or skills as superfluous to that goal but to pursue them in intentional ways that build deep, conceptual understanding along with the behaviors and habits that distinguish scientists.

Now, we're living in the future that the NGSS imagined over a decade ago. While there is still a long way to go to achieve the vision for all learners, and there have been many bumps along the road (particularly in assessment—remember when we tried to assess one dimension at a time using multiple choice questions?), it's time to take stock of how the field has shifted, and what that means for science assessment. Assessments, along with instructional materials and professional learning, are how standards come to life for students, teachers, leaders, families, and communities. It stands to reason, then, that features of high-quality assessments should account for the context within which the NGSS and related standards are operating.

While there is still a long way to go to achieve the vision for all learners, and there have been many bumps along the road . . . it's time to take stock of how the field has shifted, and what that means for science assessment.

From my vantage point working across levels in the system, from national and state policy conversations to discussions about classroom practice, we seem to be in the middle of major shifts that are relevant to how we think about science assessments.

1. A call for locally, culturally relevant curriculum and assessments that are meaningful to the students being taught.
2. An acknowledgment of competencies and durable skills that are going to be increasingly essential for every learner in current and future economies.
3. A vibrant interest in technologies and how they could and should be leveraged in classrooms, and what their emergence means for academic goals.

When I think of future-focused science assessments, I think about how assessments surface the brilliance of young people in science in ways that attend to each of these major emerging priorities.

UNDERSTANDING OUR FOUNDATION: CURRENT FEATURES OF HIGH-QUALITY SCIENCE ASSESSMENTS

Before we think about how future-focused science assessments might require an evolution from current practice, let's ground ourselves in what those current practices are. Here are some basic tenets and non-negotiable features of current NGSS assessments (Achieve, 2018, 2019b, 2019d).

1. Assessments ask students to demonstrate their understanding and ability to use the SEPs, CCCs, and DCIs by applying them to making sense of phenomena or solving problems.
2. Like instruction, high-quality phenomena and problems must drive assessment tasks—from the perspective of students, the purpose of their activities during an assessment should be to figure out the phenomenon or address the problem identified or posed.
3. Assessments must be multidimensional—that is, they must assess multiple dimensions *together.* This is true for every part and type of assessment, from individual prompts to complex tasks. These dimensions are most often the DCI and SEP, with DCI/CCC and CCC/SEP combinations often rounding tasks.
4. Assessments must be accessible to the full range of learners intended to engage with the assessment.

5. Assessments must support their intended uses. This includes providing appropriate information at the right time and grainsize, as well as being educative to teachers, building their science teaching and assessing practices.

In a nutshell, assessments should surface student understanding of standards in ways that are similar to how we teach science: phenomenon and problem driven, focused on integrating the three dimensions in service of sensemaking, supportive of a range of learners, and useful for its intended purpose. Now let's think about how the changing conversations in teaching, learning, and assessment might shape how we think about these goals moving forward.

> **Assessments should surface student understanding of standards in ways that are similar to how we teach science: phenomenon and problem driven, focused on integrating the three dimensions in service of sensemaking, supportive of a range of learners, and useful for its intended purpose.**

Landscape Shift 1: Local and Cultural Relevance and Personalization

Equity has always been at the center of the NGSS and similar standards, and all curriculum, professional learning, and assessment efforts surrounding them (NGSS Lead States, 2013, Appendix D). The NGSS development process included an equity team that was embedded in every step of development, and the standards documents imbue equity implicitly and explicitly throughout. The very design of the standards was intended to disrupt inequities in science teaching and learning. Some examples include:

- Setting *standards for all students.* This disrupted the too-common practice of keeping some science disciplines (e.g., physics) and some experiences (e.g., authentic engagement with science practices and sensemaking) reserved for certain students, while barring access for others.
- Elevating *science and engineering practices* as the mechanism by which science understanding was made visible. This effectively removed language and communication differences among learners as a barrier to science learning and performance.

- Making *cross-cutting concepts* an explicit and expected part of both teaching and learning, as well as of assessment. This made these incredibly powerful thinking tools that support not only sensemaking but also broader competencies like transfer, flexible reasoning, and agency something, that was an intentional learning goal for all students, not something that some students happened to develop over time (Badrinarayan & Cooper, 2023*a*).
- Building robust standards that spiral such that *ideas and practices are revisited in more and more depth from kindergarten through high school.* This meant that (*a*) all young children are exposed to and developing meaningful content knowledge that will serve them throughout their lives (e.g., in reading comprehension along with science); and that (*a*) students were given the structures to learn and be successful by making sure they always have the necessary prior knowledge to build future, deeper learning (more on this in the next chapter!).

In the early days of standards implementation, equity was at the heart of every conversation. In those days, the focus was on *all students*—this idea that our most important charge was to make sure everyone had opportunity to meet these common, aspirational standards. Questions about equity often revolved around how to make sure everyone got access to the full breadth of standards, with the assumption that this wouldn't sacrifice depth because the standards were *designed* for depth by integrating sophisticated practices and thinking lenses with conceptual understanding and an expectation for sensemaking in each standard. "All standards, all students" was the shared mantra of researchers, state leaders, classroom teachers, and those of us responsible for tending to and stewarding the standards and their implementation. When the NGSS were first released and being adopted by states, equity conversations focused on questions like:

- How do we ensure all students have opportunities to experience rigorous, challenging, and meaningful science learning experiences?
- How do we make sure that multilingual students and students with disabilities—student groups who had notoriously often been left out of science education all together—had regular and rich access to science classes?
- How do we advocate for, secure, and protect resources (time, money, and human capital) for science in a world that can feel dominated by math and English language arts (ELA)?

That conversation is shifting. Amid an ever-changing landscape—and one that has brought the importance of science, if not science education, to the forefront—parents, teachers, leaders, and students are asking more pointed questions about how we spend our time and energy in education. As notions of what it means to equitably support student learning in science are evolving, people are asking different questions, like:

- Is it actually important that students demonstrate mastery of all the DCIs? Important for what? And to whom?
- Wouldn't it be better if high school students could choose the area of science that they are passionate about, rather than having to pursue all three science disciplines at a level of sophistication that seems unneeded?
- If students find science irrelevant to their own lives within their own communities, does it matter if they "learned" it all?

When the NGSS were first released, the major equity transformation was from "no science" or "memorized facts and cookbook labs" to "robust access to disciplinary concepts, practices, and sensemaking for all." We're in a different world now. Part of this comes out of lessons learned from initial efforts to implement the NGSS, as well as from lessons learned from developing curriculum and assessment, and doing the work of shifting instruction. In implementing, we have seen the need for relevance to students and communities become increasingly apparent.

But I think there's more to it than that. Since the release of the NGSS, our teachers, families, and students have survived a global pandemic, an alarming slew of natural hazards and climate-related devastations, a sickening reckoning with the ever-presence of racism and injustices within our public institutions, explicit confrontation of the violence women and nonbinary people face, and tumultuous political and economic times.

Since the release of the NGSS, our teachers, families, and students have survived a global pandemic, an alarming slew of natural hazards and climate-related devastations, a sickening reckoning with the ever-presence of racism and injustices within our public institutions, explicit confrontation of the violence women and nonbinary people face, and tumultuous political and economic times.

I was talking with a teacher at a workshop, and she described the change she has seen in students:

> They are the [coronavirus-19] COVID and social media generation. They know more about what's going on in countries across the world than I knew about the next county over when I was growing up. Their sense of community is so much broader than we are used to—their communities aren't just the places they live, they're people they relate to through shared interests and passions and political views and priorities, that they talk to more than the kid sitting next to them. And they lived through a pandemic, are craving relationships, and don't know how to form them. Mitosis is the farthest thing from their minds, and there's really nothing I can say in good faith to convince them to spend effort learning something that honestly, they probably won't use, when they are trying to rebuild human connections.

We are exhausted. And, we are, all of us, questioning: What is the best way to use our time, our gifts, our relationships with young people to improve their lives and our world?

The answer has been, increasingly, to rebalance efforts in favor of local contexts and needs, as well as relevance and personalization to learners and their communities. We've seen this shift across research, practice, and policy discussions. An increasing number of scholars have called for reimagining science education in service of culturally sustaining and justice-oriented efforts, for centering youth voice and community perspectives (Bang, 2016; Lee & Grapin, 2024; Morales-Doyle, 2019, 2024). At the same time, practitioners and curriculum developers have been focusing efforts on locally relevant implementation, ranging from adapting high-quality instructional materials (HQIMs) to leverage local community assets to a reinvigoration of local curriculum and assessment efforts (National Academies of Science, Engineering, and Medicine, 2022). Even at the state and national levels, policymakers are prioritizing supporting local relevance and needs. For example, some states are adopting new science instructional material frameworks that center relevance to students and the ability to customize or adapt materials for local implementation (Maryland State Department of Education, 2024), while other states are exploring how to incorporate local contexts and priorities into science standards adoption and revision processes. (Some examples include the New Hampshire Performance Learning and Assessment Consortium for Educators, the Massachusetts Department of Elementary and Secondary Education's science performance tasks, and the science work conducted as part of Hawai'i's Performance Assessment Development Initiative.)

What does this mean for science assessments moving forward (Table 1.1)? Assessment design has always been a game of trade-offs—so, as we look to assessments that can serve students now and into the future, we make some different trade-offs:

- Relevance and engagement shift from "nice to have" to must-have features of science teaching, learning, *and* assessment. We take seriously the time trade-off that is inherent here. Since the standards were released, there has been a concern that there is "too much" in the NGSS to cover within standard school years, given the amount of time needed for students to learn in ways that are likely to lead to sustained understanding. This is not going to get better as we try to co-design more with students, and create space for student-driven projects and inquiries. There will be a tension, but one we must figure out how to overcome.
- We make sampling decisions—that is, what we emphasize and de-emphasize in assessment design—that lean into and honor the aspects of science standards that serve students and communities best, rather than trying to develop assessments that are "neutral" or agnostic to community interests.
- We prioritize assessments that measure what matters to students and their communities, in ways that are humanizing and asset oriented and make available information to those closest to student learning, rather than giving all of the power to assessments that are largely used by external actors.

Table 1.1 • *What Changes for Future-Focused Science Assessments?*

Some things stay the same . . .	**Some things change . . .**
▸ Assessments ask students to demonstrate and integrate DCIs, SEPs, and CCCs to make sense of phenomena and problems. ▸ Phenomena and problems drive assessment activities. ▸ Students are expected to develop the kinds of sophisticated understanding and practices described by the elements of the standards.	▸ Relevance moves beyond familiarity; assessments ask students to make sense of truly meaningful and impactful phenomena and problems. ▸ Students demonstrate their understanding of science through their reasoning and sensemaking, rather than through whether they can produce the "right answer." ▸ Assessments provide more opportunities for students to bring themselves, and their unique perspectives, cultural and linguistic assets, and ideas to the table as a meaningful part of sensemaking.

Landscape Shift 2: Competencies and Durable Skills

Alongside calls for teaching, learning, and assessment that is more relevant and meaningful to students and the communities they are part of, there has been a rising call to reconsider what knowledge, practices, and abilities students need to be developing in school in general, and what that should mean for teaching and learning in each of the disciplines—science included—specifically. Employers, community leaders, and scholars suggest that skills like communication, problem-solving, adaptability, and creativity—skills we might consider "durable"—are going to be essential for all students to develop and cultivate to be successful in the coming decades (Liu et al., 2023). Schools, districts, and states are taking this to heart. For example, many schools and districts have begun adopting and implementing portraits of a graduate, a framework of competencies that, holistically, schools should be helping students achieve. At the state level, a growing number of states are pursuing similar frameworks, along with exploring competency- and skills-based assessment models to make different kinds of data, about different kinds of outcomes, available to support teaching and learning.

Employers, community leaders, and scholars suggest that skills like communication, problem-solving, adaptability, and creativity—skills we might consider "durable"—are going to be essential for all students to develop and cultivate to be successful in the coming decades.

Although we often talk about skills like communication or critical thinking as content-agnostic, transferrable skills, they don't actually develop that way. Put plainly, we don't learn to communicate as a generalized skill—like all learning, we learn in context, and over many opportunities to develop and practice these skills in many different contexts, we begin to be able to generalize them to more and more contexts that we haven't been in before. For example, consider communication. Most people find that they are effective communicators to specific audiences, in specific contexts, and they become so by (intentionally or not) working with those audiences repeatedly over time. While they may be amazing communicators in those contexts, they can often be somewhat ineffective in less familiar or very different contexts.

As the call for durable skills grows stronger, it is important that we think about what this means for science teaching, learning, and assessment. On the one hand, it seems foolish and shortsighted to ignore the emphasis on durable skills, or to hope students will develop them without specific attention to them in instructional design and assessment. This is not in students' best interests. First, I am very sure that everyone calling out the importance of durable skills is right—there has been a steady growth in how employers rate the importance of these kinds of skills for an increasing proportion of job functions, and this is unlikely to change. Second, amid calls to reduce the number of standards students are expected to develop, and a renewed interests and investment in math and ELA following pandemic-related decline in performance, ignoring the relationship between durable skills and science will likely only result in students getting access to and learning *less science* in the long run.

The good news is that durable skills don't have to be "another thing" that we now have to fit into science instruction and assessment. The NGSS and *Framework for K-12 Science Education* (NRC, 2012) intentionally center these kinds of skills within the SEPs, CCCs, nature of science, and connections to technology and society (see examples in Table 1.2).

Table 1.2 • *Analyzing Relationships Between Durable Skills and the NGSS*

Durable skill	NGSS/*Framework* connection
Communication: Use of context-relevant strategies, domain-specific codes and tools when interacting with others, including active listening, asking questions, synthesizing messages, storytelling, and public speaking	The NGSS and *Framework* prioritize communication both of science and engineering ideas as well as in service of sensemaking throughout K–12. This is particularly prevalent in the following practice expectations: ▸ Developing and Using Models ▸ Engaging in Argument From Evidence ▸ Obtaining, Evaluating, and Communicating Information
Reasoning: Logic-based thinking processes of an inductive or deductive nature that are used to draw evidence-based conclusions from data, facts, or premises	Reasoning and scientific sensemaking are at the heart of the entire NGSS and *Framework*. Given expectations for students to make their thinking visible and for all dimensions to be used in service of sensemaking, detailed information about students' abilities to leverage logic and disciplinary knowledge together, to draw conclusions from specific and varied forms of evidence, to update their thinking based on new information, and other dimensions of reasoning will routinely be surfaced.

(Continued)

(Continued)

Durable skill	NGSS/*Framework* connection
Systems Thinking: Mental analyses of any system to understand system elements, the interconnections among the elements that drive the system to work as a whole, and how its constituent elements function both individually and in relation to each other	At every grade band, systems thinking is emphasized through specific DCIs, as well as through specific SEPs and CCCs. For example, the following SEPs and CCCs are central to the framework and provide deep insight into students' ability to identify components of systems, their interactions, and implications for function: **1.** Developing and Using Models **2.** Mathematics and Computational Thinking **3.** Systems and System Models **4.** Relationships Between Structure and Function These SEPs and CCCs can work across domains to provide insight into students' ability to engage in systems thinking. In addition, some specific DCIs also surface systems understanding and can be used, both directly and in transfer contexts, to provide information about this durable skill. These include DCIs related to topics such as: ▸ Ecosystem Functioning ▸ Earth Systems Interactions ▸ Human Impacts on Earth Systems

Therefore, the bones of integrating durable skills and NGSS in science assessments are there—we just have to fill it out by considering how we interpret the NGSS expectations, and design assessments, in ways that provide information about, and support for, cultivating these skills. Consider the following example, which seeks to consider how the NGSS relates to a learner outcome described by a portrait of a graduate (Table 1.3).

Table 1.3 • *Seeing the NGSS in Portraits of a Graduate*

<table>
<tr><td colspan="2">Graduate profile goal: Original thinkers for an uncertain world. Students are sense-makers, problem-solvers, generative thinkers. They respond to changing circumstances with new solutions and look for options even when no clear path exists. Students seek, make sense of, and build on global learnings to meet and balance the needs of many different kinds of people, industries, and perspectives.</td></tr>
<tr><th>Example of an NGSS element that could be used to surface this skill</th><th>Considerations for assessment design</th></tr>
<tr><td rowspan="2">Asking questions and defining problems: Define a design problem that involves the development of a process or system with interacting components and criteria and constraints that may include social, technical, and/or environmental considerations</td><td><u>Ways this could support the competency</u>
▸ The context for the design problem is a complex system with changing or dynamic factors and environments.
▸ Tasks require students to define a design problem that intentionally focuses on an area that people have different perspectives, interpretations of evidence, and values related to. Therefore, defining the problem specifically requires contending with these features.
▸ The situation is authentically uncertain, and likely an example of an authentic, intersectional problem, for which science is part of, not the full, solution.
▸ Problem definition has collaborative components (likely needed to really seriously contend with complex problems).</td></tr>
<tr><td><u>Ways this would not support the competency</u>
▸ Straightforward context without real complexity to navigate.
▸ Emphasis in task on isolated skills, like defining criteria and constraints, without attending to the broader context and orientation.
▸ Efforts to limit the complexity to only focus on a certain kind of data or perspective.</td></tr>
</table>

Like contending with local relevance, the field-wide shift toward competencies and durable skills is not in competition with meaningful NGSS assessments, but it does change the nature of how we interpret the standards for assessment, and what trade-offs are reasonable given this emphasis (Table 1.4).

Table 1.4 • *What Changes for Future-Focused Science Assessments Are More Locally Relevant?*

Some things stay the same . . .	Some things change . . .
▸ Assessments ask students to demonstrate and integrate DCIs, SEPs, and CCCs to make sense of phenomena and problems. ▸ Phenomena and problems drive assessment activities. ▸ Students are expected to develop the kinds of sophisticated understanding and practices described by the elements of the standards.	▸ The nature of the phenomena, problems, and tasks change to emphasize features of the targeted competency. ▸ We may think differently about how we define success and outcomes, and what information is important to pay attention to. ▸ Assessments pay more attention to application in real-world contexts.

Landscape Shift 3: Technology and Its Impact on Teaching, Learning, and Assessment in Science

Science education and assessment have always had to contend with evolutions in technology. Indeed, one major argument for shifting toward prioritizing conceptual understanding and sensemaking has been the substantial growth in access to facts about the world, and how integrated science and technology is in our daily lives. Given how much information and misinformation is available to people, and how many daily and existential decisions about health, privacy, productivity, and entertainment we make that are rooted in science-based technologies, it has been imperative that we reimagine what success in science looks like from "how many facts about space can you tell us?" to "how do we better understand the trade-offs between using land for increased agricultural production vs. the impact on climate and environmental health?" Science teaching, learning, and assessment has been in dialogue with computer science, tech-focused career and technical education, and STEM career pathways for decades—what is shifting is not the fact that there is a relationship between emerging technologies and the science needed to understand, engage, and competitively shape them, but the *nature* and *relative urgency* of figuring out how we do so.

It feels impossible to say anything about the future and not talk about tech generally, and artificial intelligence (AI) specifically. I'll be honest and say I'm fairly neutral on the topic of AI—I don't think AI is either going to doom science education nor do I think it's a solution to all of our challenges. What I do think is that it will have real consequences because we are giving the tech that power by investing so much time, resources, hope, and suspicion in it. I say that with no judgment—I really think what's so interesting about generative AI (genAI) lies in how accessible it is.

I don't think AI is either going to doom science education nor do I think it's a solution to all of our challenges.

When it comes to technology in science assessment, there has been a long-standing interest in leveraging everything from game-based assessments to immersive virtual and augmented reality spaces for more meaningfully surfacing student understanding (Brown, 2020; Li et al., 2023). Right now, two major conversations are influencing how we think about science assessments:

1. Whether the widespread availability of technology like ChatGPT changes the nature of what students need to know and be able to do. For example, if genAI-powered tools can develop research plans and proposals, analyze data, and present data in many different ways, is it still important for students to learn those skills and practices?
2. To what extent will the availability of increasingly effective tech change what our assessments themselves can look like and accomplish. For example, the Programme for International Student Assessment (PISA) is introducing an AI chat bot to support student performance within future administrations of the assessment.

Both conversations have the potential to tremendously impact how we approach assessing student learning in science. At the time I'm writing this, I don't think we know yet what exactly those trade-offs or changes will look like—there are a lot of ideas out there, and I discuss some of this specifically later in the book. For now, the conversations around technology dovetail with both of the prior trends to suggest that while the core tenets of science assessment likely won't immediately change, we may indeed see a call for emphasizing more tech-relevant interpretations of SEPs, CCCs, DCIs, and sensemaking (Table 1.5).

Table 1.5 • *What Changes for Future Focused Science Assessments Attend to Technology?*

Some things stay the same . . .	Some things change . . .
▸ Assessments ask students to demonstrate and integrate DCIs, SEPs, and CCCs to make sense of phenomena and problems. ▸ Phenomena and problems drive assessment activities. ▸ Students are expected to develop at least some of the sophisticated understanding and practices described by the elements of the standards.	▸ Students may begin to consider AI as a tool used in science disciplines as part of assessments. ▸ More science assessments may provide students with opportunities to engage directly with technologies, including with AI and other branches of computer science. ▸ High school pathways may influence the nature of the phenomena, problems, and what aspects of science are foregrounded. ▸ Students may have access to AI-based accommodations, support tools, and data processing tools. ▸ Students may be able to synthesize information across many different performances for scores/outcomes measures.

FUTURE-FOCUSED SCIENCE ASSESSMENT: EVOLUTION

So, what does this all mean for science assessment moving forward?

A great deal of what we have learned about assessing students in science is of course still relevant and forms the foundation of this next phase in our evolution. We still want to understand how well students can use their DCIs, SEPs, and CCCs to figure out phenomena and problems. We want to do this in ways that are equitable and support learning moving forward.

What is clearly changing is what we prioritize and value in science assessments, and how we change our orientations and approaches to match those. Future-focused science assessments empower students and teachers with knowledge and practice but also with confidence and an understanding of themselves as a critical element of the science community. Doing so in ways that reflect the shifting priorities of the field and coherently support evolving conversations about learning goals requires that we reimagine what we hold up as non-negotiable in science assessment, and what might be something that we foreground in different ways throughout students' K–12 experiences. The four priorities described in the upcoming chapters describe how we might accomplish this.

CHAPTER SUMMARY

- Standards based on *A Framework for K-12 Science Education* (NRC, 2012)—like the Next Generation Science Standards—posed a bold and ambitious vision for science education. Assessments that are designed to not only measure learning aligned with that vision but also signal and incentivize teaching and learning that prepares students for the future must look different and be responsive to major landscape shifts.
- Our foundation for better science assessments is grounded in assessments that reflect best practices for teaching and measuring multidimensional standards: Assessments should focus on surfacing students' ability to make sense of phenomena and problems using their DCIs, SEPs, and CCCs through instruments designed for equity and purpose.
- At the same time, future-focused assessments have to contend with three major landscape shifts: a shift toward local and cultural relevance, cultivating durable skills and competencies through science (and measuring these competencies in science assessments), and a fresh take on the interaction between evolving technologies and science education.

(Continued)

(Continued)

REFLECTION QUESTIONS

1. What are the major changes you have made to your science assessments since beginning to teach new science standards?
2. How have the three major landscape shifts shown up in your teaching, learning, and assessment practice? What trade-offs are you currently making in response to them?
3. How do you think about using science, and science assessments, as a way to empower learners to have agency in their futures?