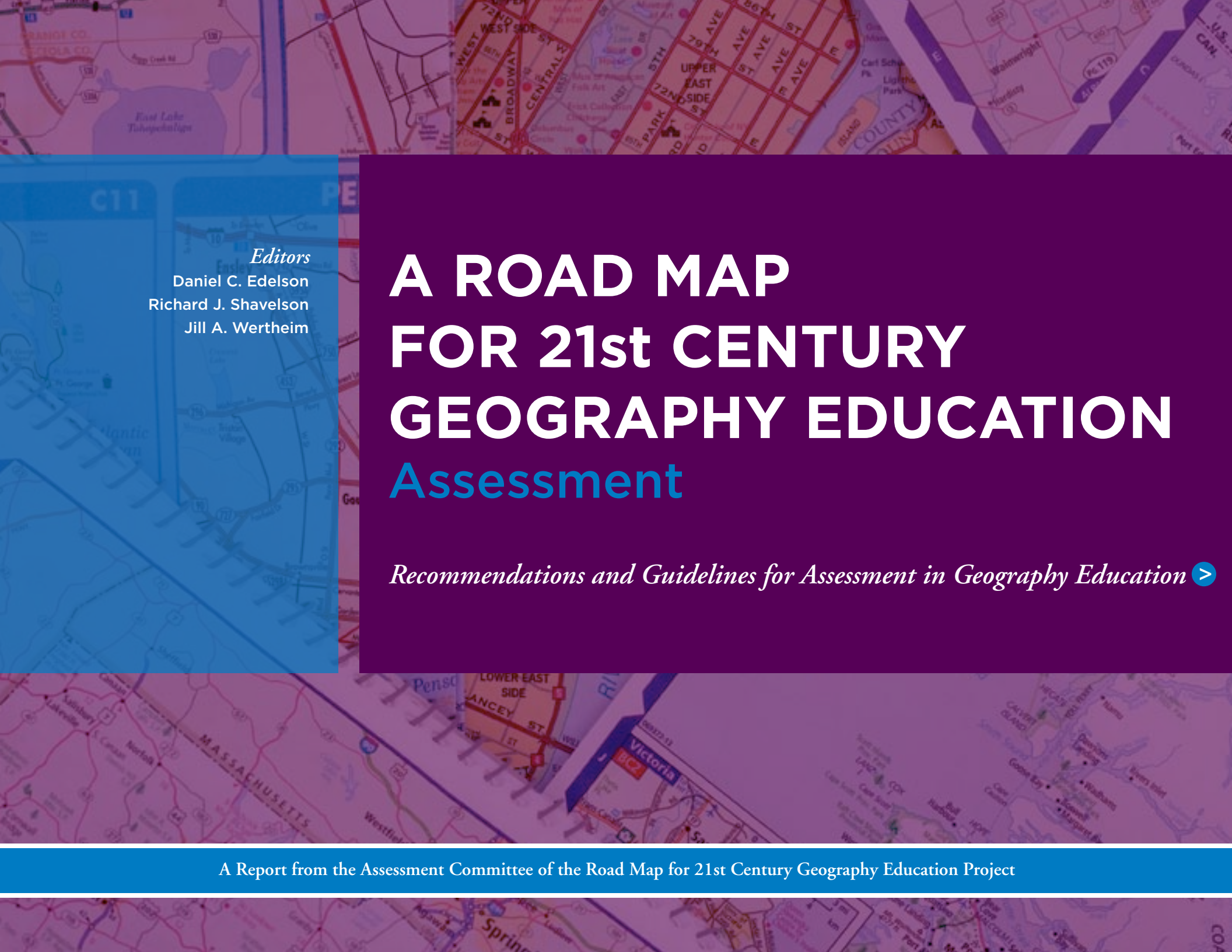


Editors
Daniel C. Edelson
Richard J. Shavelson
Jill A. Wertheim

A ROAD MAP FOR 21st CENTURY GEOGRAPHY EDUCATION

Assessment

Recommendations and Guidelines for Assessment in Geography Education

A Report from the Assessment Committee of the Road Map for 21st Century Geography Education Project

Road Map for 21st Century Geography Education Project

Assessment

Recommendations and Guidelines for Assessment in Geography Education

Editors

Daniel C. Edelson, Richard J. Shavelson, Jill A. Wertheim

National Geographic Society
Washington, DC

A Report from the Assessment Committee of the Road Map for 21st Century Geography Education Project

This report was created by the Road Map for 21st Century Geography Education Project.

Road Map for 21st Century Geography Education Project

Daniel C. Edelson, *Principal Investigator*
Virginia M. Pitts, *Project Director*

The Road Map for 21st Century Geography Education Project is a collaboration between the National Geographic Society, the Association of American Geographers, the National Council for Geographic Education, and the American Geographical Society. The views expressed in the report are those of the authors and do not necessarily reflect the views of these organizations.

The Road Map for 21st Century Geography Education Project has been supported in part by the National Science Foundation under Grant No. DRL-1049437. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Suggested citation:
Edelson, D. C., Shavelson, R. J., & Wertheim, J. A. (Eds.). (2013). *A road map for 21st century geography education: Assessment* (A report from the Assessment Committee of the Road Map for 21st Century Geography Education Project). Washington, DC: National Geographic Society.

Information about the project and copies of reports are available at http://natgeoed.org/roadmap.

The **National Geographic Society** is one of the world's largest nonprofit scientific and educational organizations. Founded in 1888 to "increase and diffuse geographic knowledge," the Society's mission is to inspire people to care about the planet. It reaches more than 400 million people worldwide each month through its official journal, *National Geographic*, and other magazines; National Geographic Channel; television documentaries; music; radio; films; books; DVDs; maps; exhibitions; live events; school publishing programs; interactive media; and merchandise. National Geographic has funded more than 10,000 scientific research, conservation and exploration projects and supports an education program promoting geographic literacy.

The **Association of American Geographers** (AAG) is a nonprofit scientific, research, and educational society founded in 1904. Its 11,000 members from more than 60 countries share interests in the theory, methods, and practice of geography (including GIScience, geographic education, and geographic technologies). The AAG pursues its mission through its many conferences, scholarly publications, research projects, educational programs, topical specialty groups, and its extensive international network of colleagues and organizational partnerships, which encompass professionals working across public, private, and academic sectors all around the world.

The **National Council for Geographic Education** (NCGE) works to enhance the quality, quantity, and status of geography teaching and learning in primary, secondary, university, and informal educational settings. It develops and promotes curricular materials and two journals, fosters best practices in pedagogy and geotechnology, connects educators through online communication and through its annual conference, supports research in geographic education, recognizes exceptional supporters and teachers of geography, and collaborates with other organizations that have similar goals.

The **American Geographical Society** is an organization of professional geographers and other devotees of geography who share a fascination with the subject and a recognition of its importance. Most Fellows of the Society are Americans, but among them have always been a significant number of Fellows from around the world. The Society encourages activities that expand geographical knowledge, and it has a well-earned reputation for presenting and interpreting that knowledge so that it can be understood and used not just by geographers but by others as well—especially policy makers. It is the oldest nationwide geographical organization in the United States. Its priorities and programs have constantly evolved with the times, but the Society's tradition of service to the U.S. government, business community, and nation-at-large has continued unchanged.

Assessment Committee

Daniel C. Edelson, Chair

Richard J. Shavelson, Co-Chair

Jill A. Wertheim, Research Director

Barbara Hildebrant, Educational Testing Service

Elizabeth R. Hinde, Arizona State University

Marianne Kenney, Denver Public Schools

Bob Kolvoord, James Madison University

David A. Lanegran, Macalester College

Jody Smothers Marcello, Sitka High School

Robert W. Morrill, Virginia Tech

Maria Ruiz-Primo, University of Colorado Denver

Peter Seixas, University of British Columbia

Project Steering Committee

Daniel C. Edelson, Principal Investigator and
Chair, Assessment Committee

Virginia M. Pitts, Project Director

Susan Heffron, Association of American Geographers and
Co-Chair, Geography Education Research Committee

Doug Richardson
Association of American Geographers

Jerome E. Dobson, American Geographical Society and
Study Director, Public Understanding and Values Study

Kristin Alvarez
National Council for Geographic Education

Joseph Kerski
National Council for Geographic Education

Robert Dulli
National Geographic Society

Sarah Witham Bednarz, Chair
Geography Education Research Committee

Emily M. Schell, Chair
Instructional Materials and Professional Development Committee

Kathleen J. Roth, Co-Chair
Instructional Materials and Professional Development Committee

Richard J. Shavelson, Co-Chair
Assessment Committee

Project Advisory Board

Gilbert M. Grosvenor, Chair
National Geographic Society

Harm de Blij, Michigan State University

Richard G. Boehm, Texas State University-San Marcos

Barbara Chow, The William and Flora Hewlett Foundation

Jack Dangermond, Esri

Michael F. Goodchild, University of California, Santa Barbara

Brian McClendon, Google

Alexander B. Murphy, University of Oregon

M. Duane Nellis, University of Idaho

Trevor Packer, The College Board

Lee Schwartz

Table of Contents

In the online version of this report, the icon on the lower left corner of the following pages will bring you back to this Table of Contents.

Executive Summary

Introduction

Student assessments typically are viewed simply as indicators of educational progress, but this report is based on the recognition that the utility of assessments can extend far beyond this role in education. For example, the results of student assessments can provide critical information for decision making in education policy and practice. In addition, what is being assessed and how it is assessed becomes a means to communicate goals and priorities to teachers, students, and other stakeholders in K–12 education.

This report explores how changes and improvements in assessment practices can support efforts to improve K–12 geography education. The report is one of three reports developed as part of the National Science Foundation-funded Road Map for 21st Century Geography Education Project, a collaboration of four national associations committed to improving geography education—the National Geographic Society, the Association of American Geographers, the American Geographical Society, and the National Council for Geographic Education. The other two project reports focus on geography education research and on instructional materials and professional development for geography.

This report begins by laying out a set of issues critical for the design of assessments that support instructional improvement and by reviewing current assessment frameworks and practices in K–12 geography education. The second half of the report contains a proposal for a new approach to assessment in geography that will enable assessment developers to address the critical issues in assessment design. As with the other Road Map Project reports, this one places a particular emphasis on how to move geography education toward a balance between developing geographic knowledge and learning to engage in geographic practices. Specifically, this report follows the balanced approach to geography education called for in the second edition of *Geography for Life: National Geography Standards* (Heffron & Downs, 2012), the national standards document that also was developed through a collaborative effort of the four Road Map Project partners.

Background

The four partners launched the Road Map Project because they share a concern that the poor state of geography education in America is a threat to our country's well-being, and by extension, to the well-being of the global community. The partners share the belief that geography education is essential for preparing the general population for careers, civic lives, and personal decision making in contemporary society. They also believe it is essential for the preparation of specialists capable of addressing critical societal issues in the areas of social welfare, economic stability, environmental health, and international relations. They fear that by neglecting geography education today, we are placing the welfare of future generations at risk.

While inspiring examples of highly effective geography education can be found in nearly every part of the United States, for the overwhelming majority of students, the amount of geography instruction they receive, the preparation of their teachers to teach geography, and the quality of their instructional materials are inadequate to prepare students for the demands of the modern world. Assessments of geographic concepts and skills expose the failure of our educational system in geography, confirming that a vast majority of American students are geographically illiterate. The 2010 National Assessment of Educational Progress (NAEP), known as "The Nation's Report Card," found that fewer than 30% of American students were proficient in geography, meaning that more than 70% of students at fourth, eighth, and 12th grades were unable to perform at the level that is expected for their grade (National Center for Education Statistics, 2011). At 12th grade, more than 30% of students scored below "basic," indicating that they had not mastered even foundational geographic concepts or skills.

The Value of Assessment for Improving Geography Education

This report takes the position that the primary purpose of educational assessment should be for making informed decisions. Because they typically regard assessment as a separate activity from instruction, educators, students, parents, and policy makers often overlook invaluable ways assessments can support and improve teaching and learning. The report describes four ways that assessment results can contribute to improvements in teaching and learning by providing evidence that guides critical decisions.

1. The results of assessments can inform teachers' instructional decisions. When assessments are integrated into instruction, they can improve its effectiveness by enabling the teaching and learning process to be tailored to students' specific needs.

2. The results of assessments can be used to inform decisions about students' academic programs. Assessments introduced at appropriate intervals can be used to measure a student's proficiency against benchmark goals for that student at that point in his or her academic career. The results of these assessments can be used to inform decisions about that student's academic program.
3. The results of assessments can be used to inform decisions about the function and effectiveness of educational programs. Aggregated results of student assessments can be used as part of program evaluation. Used in this way, assessments can inform decisions about program selection, program implementation, and other aspects of instruction. They also can be used in evaluations of the performance of classes, schools, and larger units that might reveal challenges that need to be addressed; likewise, they can be used to inform decisions about where to focus resources.
4. The results of assessments can be used to build a knowledge base for future decision making. Assessment results used for research enable examination of broader questions than those revealed by the performance of a specific student or program. They can be used to examine general questions about teaching and learning geography, such as what makes one approach more effective than another, or how students develop spatial learning skills. The results of these studies can inform efforts to improve education over longer time scales.

Considerations in the Design of Assessments

Designing accurate and useful assessments is extremely challenging. Four of the key decisions in the design of assessments are:

1. **Selection of goals:** What are the specific content and practices required for the competencies being assessed?
2. **Item characteristics:** What are the characteristics of the item that will be used to assess a competency (e.g., task type, response mode, scoring system)?
3. **Item quality:** How will the technical quality of the item be measured (e.g., validity, reliability, fairness)?
4. **Cost effectiveness:** How much time and resources are required to create, administer, and score the assessment?

In making these design decisions, assessment developers must carefully consider the *nature of the content and practices* to be assessed, the *context* in which they will be administered, the *population* whose competencies will be assessed, and the *purposes* for which the results will be used.

One way developers of assessments minimize the challenge of addressing these considerations is through assessment frameworks. An assessment framework plays the role of an outline in writing or a functional specification in engineering. Frameworks provide guidelines for making decisions in the development of an assessment.

Contemporary assessment frameworks use a two-dimensional framework to lay out the content and cognition targets for an assessment, their relative importance, and item characteristics. A comprehensive assessment framework also provides guidance on item quality and cost constraints.

Because of the role assessment frameworks can play in guiding the design of assessments, this report focuses on the development and dissemination of new assessment frameworks as a means to guide the development of high-quality assessments that evaluate 21st century knowledge and skills.

Assessment in Geography Today

To determine how well current assessment projects are aligned with the goals of geography education, as described in *Geography for Life*, this report examines the nature of existing assessment frameworks and current assessment practices in K–12 geography education.

Assessment Frameworks in Geography Education Today

There are currently three prominent assessment frameworks being used in K–12 geography education in the United States:

- **National Assessment of Educational Progress in Geography (1994, 2001, 2010).** The NAEP geography framework is the basis for assessments that are used in a national evaluation of geography education outcomes at grades 4, 8, and 12.
- **Advanced Placement Human Geography (2000).** The framework for Advanced Placement Human Geography (APHG) guides the design of the examination used by the College Board to determine if high school students who have completed an AP course in human geography have achieved a level of mastery equivalent to successful completion of an undergraduate course in the subject.

- **National Assessment of Educational Progress for Science (2008).** The NAEP Science framework is the basis for assessments that are used in a national evaluation of science education outcomes at grades 4, 8, and 12. It includes concepts and practices that are included in Geography for Life, such as, Earth processes, ecology, human-environment interaction, data analysis, and communication.

This report concludes that these three frameworks place too little emphasis on geographic practices to accurately assess students' mastery of the goals outlined in *Geography for Life,* although the NAEP Science framework serves as a model of how to assess other scientific practices.

Assessment Practices in Geography Today

This report includes the findings of a study, commissioned for the report, of existing K–12 geography assessments. The study was conducted to gather in-formation about how well current assessment practices reflect the goals of *Geography for Life,* and how well they implement the principles of effective assessment design described above.

The study found the content evaluated by current assessments is unevenly distributed across the goals described in *Geography for Life*. For example, 40% of all items across both large-scale and classroom assessments evaluated knowledge from only three out of the 16 content standards, and far fewer items assessed content from the *Environment and Society* category compared with any other content area. The study also found that geographic practices are not being widely assessed. Only 30% of large-scale geography assessment items required that students use any geographic practices at all. *Analyzing geographic information* was assessed in 21% of all large-scale items, but other geographic practices were rarely assessed (Figure 1).

The study found that assessments are largely failing to probe deep understanding. More than half of the large-scale assessment items required only declarative knowledge (e.g., *knowing that*), often at the level of recognizing a definition. Only 28% assessed students' procedural knowledge (e.g., *knowing how*), which includes reading and gathering information from maps, graphs, and texts. And, only 17% of geography items required schematic knowledge (e.g., *knowing why*), which includes explaining an unfamiliar context by drawing on general geographic principles or models.

Finally, the study revealed widespread problems with item quality. Of the items studied, 60% were judged to have problems that could impede students' ability to accurately represent what they know and what they can do with their geography knowledge.

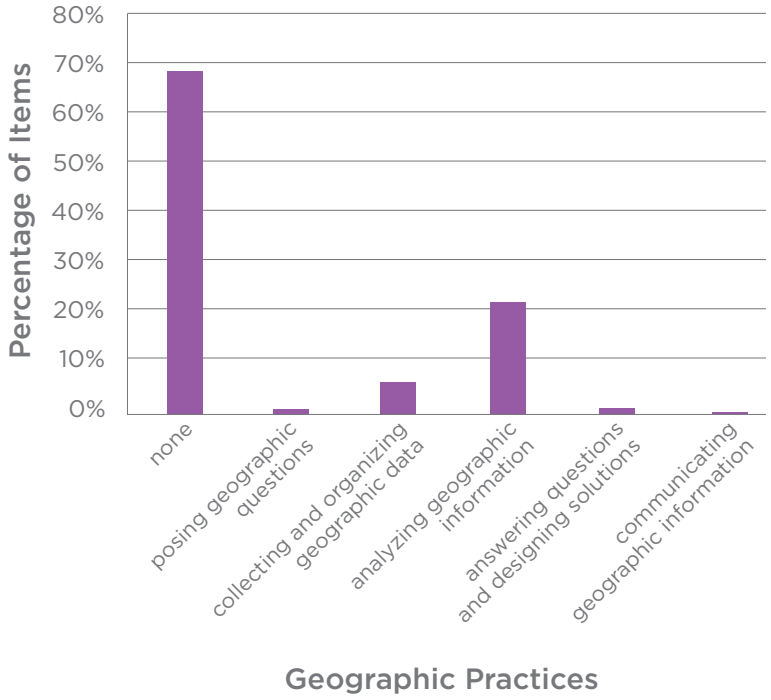
The report's review of current assessment practices reveals that both assessment frameworks and actual assessments do not reflect the balance between assessing what students know and their ability to apply their knowledge that is required to evaluate the development of 21st century geography competencies. Even within knowledge and practices, the review of assessment items reveals a large imbalance, as well.

A 21st Century Assessment Framework for the Geographical Sciences

This report introduces a new assessment framework to serve as a blueprint that guides the development of a new generation of geography assessments. Called a *21st Century Assessment Framework for the Geographical Sciences* (AFGS21), its goal is to support the design of assessments that are aligned with the goals of the national geography standards.

AFGS21 was designed to be a general assessment

Figure 1. Frequency Distribution of Large-Scale Geography Assessment Items That Target Each Geographic Practice



framework that would cover all of K–12 geography, with the idea that it will be a template for more specific assessment frameworks for specific contexts, audiences, and purposes. The report also lays out a process for creating specific assessment frameworks from AFGS21 and for using those frameworks to develop assessments.

The two dimensions of AFGS21 are designated as *content* and *cognition*. The categories in the content dimension are defined by the content standards in *Geography for Life*. The categories in the cognitive dimension are divided into *knowing and understanding* and *geographic practices.* The geographic practices, in turn, are divided into six categories. A central feature of the framework is a matrix that is used to blend the two dimensions systematically, articulating the specific performance expectations to be assessed. The contents of a cell within the matrix might describe a geographic concept that students would be expected to know or understand, or a cell might refer to the application of a geographic practice using a particular concept.

The report describes a process for developing specific assessment frameworks from AFGS21 that begins with defining the subset of content and cognition to be assessed, and the detailed process continues through the stage of specifying the desired distribution and characteristics of items. The assessments developed through this process and implemented by teachers, program and material developers, and researchers have the potential to be powerful tools for advancing the goals of geography education reform.

Preface

Background

This report is a product of the Road Map for 21st Century Geography Education Project. The Road Map Project has its origins in a directive from Congress in the 2010 budget appropriation for the National Science Foundation (NSF) that instructed the Foundation to "work with external partners with experience in geographic education to improve geography teaching, training, and research in our Nation's schools" (U.S. House of Representatives, 2009, p. 767).

In the spring of 2010, the National Geographic Society responded to this opportunity with a proposal to work with three other national organizations—the Association of American Geographers, the American Geographical Society, and the National Council for Geographic Education—to create a "road map" for efforts to improve geographic education. Building on three decades of collaboration, the partners argued that it was time to launch an initiative that would have large-scale impact across the United States over the course of the new decade. They proposed to undertake the Road Map Project to construct plans for the initiative. The project's goal would be to learn from the lessons of earlier educational improvement efforts in geography and other subjects to establish guidelines and set priorities for this new initiative.

Following extensive review, the National Science Foundation awarded a grant to support the Road Map Project in September 2010.

Project Structure

The Road Map Project was organized into four parallel efforts. Three efforts were headed by committees that were tasked with creating reports that review the current status of their area of expertise and establish recommendations for the future. The fourth effort—a study of public understanding and values—developed a survey, administered it, and analyzed its results.

As part of the proposal process, the partner organizations identified chairs and co-chairs for the three committees and a director of the study. Each of the partner organizations served as the administrative host for one of the project's four efforts and provided professional staff and administrative support for that effort. The four efforts, their hosts, and their chairs, as established by the partners, were:

- **Instructional Materials and Professional Development Committee**
  Administrative host:
  National Council for Geographic Education
  Chair: Emily M. Schell
  Co-chair: Kathleen J. Roth
- **Assessment Committee**
  Administrative host:
  National Geographic Society
  Chair: Daniel C. Edelson
  Co-chair: Richard J. Shavelson
- **Geography Education Research Committee**
  Administrative host:
  Association of American Geographers
  Chair: Sarah Witham Bednarz
  Co-chair: Susan Heffron
- **Study of Public Understanding and Values**
  Administrative host:
  American Geographical Society
  Director: Jerome E. Dobson

Once the project was funded, the project partners established a Steering Committee consisting of one representative of each of the partner organizations, the committee chairs and co-chairs, and the project director.

Report Development Process

Each report was created by a committee convened by the partners. It represents a consensus of the members of the committee. All three of the consensus reports created by the Road Map Project were developed following the same process, described below.

Committee Formation

The committees were recruited from lists of nominees and alternatives recommended by the Steering Committee. The committees were constructed to have representation from all of the following:

- academic experts in geographic education,
- academic geographers,
- academic experts in education in other areas of social studies and science,
- K–12 practitioners (teachers and administrators),
- experts in the specific foci of each committee (assessment, professional development, instructional materials development, educational research), and
- perspectives from outside the United States.

The Steering Committee wrote initial charges to the committees based on the goals of the original project proposal. Specifically, the Instructional Materials and Professional Development Committee was charged with making recommendations about the design of instructional materials and the education

of teachers. The Assessment Committee was charged with developing a framework for assessing progress toward geographic literacy across the progression from kindergarten through high school. The Geography Education Research Committee was charged with developing an agenda for educational research that would lay out questions about learning, teaching, and educational change that must be answered to maintain the effectiveness of geographic education into the future.

The chairs and co-chairs of the three committees, together with the project director and the committee research directors, formed a leadership team with the purpose of ensuring coordination and collaboration across their committees. The members of the leadership team maintained close communication with each other, but each committee was empowered to make its own decisions and to exercise independent editorial judgment over its own product. Each committee received substantial input and feedback from a variety of sources. However, they were not required to obtain approval for their products from their host organization, any of the other partner organizations, the National Science Foundation, or any other outside individual or organization. As a result, their reports reflect the opinions and judgment of their authors.

Research and Draft Phase

Each of the committees met for a kick-off meeting in Washington, DC, in January 2011. As part of the kick-off meeting, each committee reviewed and refined its charge. Once the committees were convened, they were given final editorial authority over their reports; all input from other sources was treated as advisory. The committees were each staffed by a research director with a doctorate in a related field and provided with a budget to seek input from outside experts. Each committee met face-to-face several times over the course of the research and writing process, in addition to conducting regular conversations via conference call. Information sharing among committees was facilitated through regular conversations among research directors and chairs/co-chairs.

All three committees collaborated on the organization of a workshop on geographic thinking in June 2011. The intent of this workshop, held in Washington, DC, was to bring together individuals who have insights into "expert" geographic thinking to address a series of questions related to the committees' interests in describing geographic literacy. Prior to the workshop, through a series of conference calls and online discussions, the committees identified a set of core questions they sought to have addressed at the workshop. Questions addressed at the workshop included

- "How do geographers reason about space?"
- "How do people develop spatial reasoning?"
- "How do professionals in geographic fields apply geography?"
- "How do geographers frame questions and problems differently compared with other fields?"
- "How do you train geographers?" and
- "What can we learn from how other disciplines have characterized skills, practices, and ways of thinking?"

The presenters at this workshop are listed on page 16.

Following the workshop, as part of their work on the reports, all three committees conducted additional research activities specific to their charge. Each of the three committees held face-to-face meetings in September 2011, January 2012, and April/May 2012 to work on draft-related tasks, resolve open issues, and plan work going forward.

Review and Comment Phase

Cross-committee review. First drafts of proposals were distributed to the other committees for review in January 2012. Each committee identified two to five representatives to review the other committees' reports. The intent of this review was to identify any cross-committee issues that needed to be addressed, and to give each committee input and feedback they might incorporate prior to public review. Each of the committees then met in person or via conference call to discuss the feedback and determine how to incorporate it into their next draft.

Public review. Revised drafts of the reports were made available for public comment in March 2012. The release of the draft reports was announced on the project website, and announcements about the public review were distributed to members of the Association of American Geographers, the American Geographical Society, the National Council for Geographic Education, and the National Geographic Alliance Network. All of the presenters at the June 2011 Geographic Thinking Workshop, as well as others who had contributed to the work of the different committees, were invited to comment as well.

Review Board. In an effort to ensure that the committees would obtain feedback from important constituencies for the reports, the leadership team reached out to organizations in related fields to help construct a formal Review Board for the reports. Eleven organizations were contacted and asked to nominate members or representatives of their organizations to review each of the reports. Of these, eight organizations nominated individuals. From this group, 15 individuals provided reviews of one or more reports each. The nominating organizations and the members of the Review Board are listed on page 15.

Final Preparation

Following the completion of the public and Review Board reviews, the committees carefully reviewed all of the comments received. The committee chairs, co-chairs, and directors met in Washington, DC, in April 2012, to discuss the themes that arose from the feedback and construct plans for how to address them. Each of the committees then met in late April/early May to discuss the feedback in greater detail, work on their responses, and finalize plans for completing the final drafts of the reports.

Final drafts of the reports were submitted for editing and layout in August 2012.

Dissemination

Following the publication of the reports, the four partner organizations will engage in a dissemination effort in order to bring the reports to the attention of their target audiences and to educate policy makers, funders, and front-line educators about the reports' findings and recommendations. They are being assisted in this effort by an Advisory Board. The members of the Advisory Board have reviewed and endorsed the reports and are committed to helping the partners achieve their dissemination goals. The Advisory Board members are listed on page 5.

Scope and Terminology

In a subject-specific educational project such as this, it is important to be explicit about the scope of the project. This is particularly important for geographic education, because there is so much confusion about the nature of geography and its relationship to the K–12 curriculum.

For this project, the partners chose to use the national standards document *Geography for Life: National Geography Standards,* Second Edition, as the scoping document (Heffron & Downs, 2012). *Geography for Life* lays out a scope for geographic education that cuts across the traditional boundaries of social studies and science in American schools, reflecting the fact that geography is concerned with both the physical world and the social world.

While the scope of geographic education as defined by *Geography for Life* is consistent with the way academic geographers define the field of geography, it is inconsistent with the way the term *geography* is used in most American schools and with the understanding of the term by many members of the general public. In most American schools and in the minds of the general public, the term *geography* refers to a set of basic map-reading skills; a collection of facts about place names and locations; and a body of information about people, places, and cultures around the world. Further, in American schools *geography* is a part of the social studies curriculum and is not recognized as including the substantial components of physical, life, and earth sciences included in *Geography for Life.*

For readers who are unfamiliar with the contents of *Geography for Life,* we encourage them to familiarize themselves with it prior to reading this report. We also recommend that all readers bear in mind that when the term geography appears in this report, it refers to the full range of knowledge, skills, and perspectives described in *Geography for Life*[1], not only those that are taught under the label of "geography" in schools today. In particular, the range includes elements of the social sciences, which typically are taught as part of the social studies curriculum in American schools, and elements of the physical, life, and earth sciences, which typically are taught as part of the science curriculum.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. DRL-1049437. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

[1] In the original proposal to NSF, the partners used the term geographical sciences to describe the project's scope. For the sake of readability, we decided to use geography in the project reports. This change in terminology does not reflect any shift in the focus of the initiative in the intervening time.

The Assessment Committee would like to acknowledge the contributions of many people to this report. Two of the original Committee members, Roger Downs and Sharif Shakrani, were unable to serve on the Committee through the entire project but made important contributions during the periods when they were able to participate. Four talented individuals formed the expert panel that contributed to the example assessment framework described in Chapter 5: Robert Austin (Utah Department of Education), James Dunn (University of Northern Colorado), Beth Ratway (American Institutes for Research), and Sandra Schmidt (Columbia Teacher's College). The Committee is very grateful to the presenters at the Workshop on Geographic Thinking, who helped shape our thinking on geographic practices. The Committee received very valuable feedback from the members of the Review Board, members of the other Road Map Project committees, and participants in the public review. The reports were improved immeasurably from this feedback. In addition, L. Karina Nabors assisted with data collection for the study on geography assessments, and Lafayette Cruise helped with many aspects of the report as an intern at National Geographic.

Finally, the chair and co-chair would like to acknowledge the insightful contributions of the entire Committee and express our gratitude for their patience, responsiveness to requests on short notice, and their hard work throughout the report development process. We would especially like to recognize Jill Wertheim, the Assessment Committee Research Director, who did a fantastic job of simultaneously conducting an original research study, coordinating a committee distributed across the United States, and writing and editing this report. Last, we would like to acknowledge the important role that the leadership team played in helping us to create this report and in working together to make the Road Map Project into a single, cohesive effort, rather than a set of independent committees. None of this would have been possible without the capable leadership and coordination of Virginia Pitts, the Road Map Project Director.

Road Map for 21st Century Geography Education Project

Review Board

The following organizations nominated reviewers to serve on the Review Board of the Road Map for 21st Century Geography Education Project:

- **American Association for the Advancement of Science** (AAAS)
- **American Federation of Teachers** (AFT)
- **American Geosciences Institute** (AGI)
- **Council of State Social Studies Specialists** (CS4)
- **National Board for Professional Teaching Standards** (NBPTS)
- **National Council for the Social Studies** (NCSS)
- **National Education Association** (NEA)
- **North American Association for Environmental Education** (NAAEE)

By participating in this review process, these organizations and individuals made an important contribution to the Road Map for 21st Century Geography Education Project. However, they were not asked to endorse the reports that they reviewed, so the participation of these organizations and individuals does not constitute an endorsement of the reports. While the members of the Review Board were nominated by organizations, they did not represent the views of their organizations in the review process.

The following individuals nominated by these organizations reviewed one or more of the Road Map Project Committee reports:

Assessment Committee Report

- **Ann Benbow** (AGI)
- **John Lee** (NCSS)
- **Glen MacDonald** (AAAS)
- **Lauren Mitterman** (NBPTS)
- **Sheryl Mobley-Brown** (AFT)
- **Dean Nakanishi** (NBPTS)
- **Alan Reid** (NAAEE)

Geography Education Research Committee Report

- **Fay Gore** (CS4)
- **Jackie Huntoon** (AGI)
- **John Lee** (NCSS)
- **Glen MacDonald** (AAAS)
- **Sheryl Mobley-Brown** (AFT)
- **Bora Simmons** (NAAEE)
- **Robin Wheeler** (NEA)

Instructional Materials and Professional Development Committee Report

- **John All** (AAAS)
- **Stephanie Hartman** (CS4)
- **John Lee** (NCSS)
- **Sheryl Mobley-Brown** (AFT)
- **Kevin O'Brien** (NBPTS)
- **Judith Wilson** (NEA)

Road Map for 21st Century Geography Education Project

Presenters

Workshop on Geographic Thinking

Washington, DC, June 16–17, 2011

The following invited speakers presented at a workshop on geographic thinking convened by all three committees of the Road Map for 21st Century Geography Education Project in June 2011:

Thomas Baerwald
National Science Foundation

Douglas Batson
National Geospatial-Intelligence Agency

Scott Bell
University of Saskatchewan

Sarah Brinegar
U.S. Department of Justice

Roger Downs
The Pennsylvania State University

Richard Duschl
The Pennsylvania State University

Carol Gersmehl
New York Geographic Alliance and
Renaissance Charter School

Phil Gersmehl
Michigan Geographic Alliance and
New York Center for Geographic Learning

Patricia Gober
Arizona State University

Susan Hanson
Clark University

Kim Kastens
Columbia University

Lynn Liben
The Pennsylvania State University

Janice Monk
University of Arizona

Daniel Montello
University of California, Santa Barbara

Alec Murphy
University of Oregon

Nora Newcombe
Temple University

Jeanette Rice
Rice Consulting, LLC

Peter Seixas
University of British Columbia

Chapter 1: Context and Goals for the Road Map for 21st Century Geography Education Project

The State of Geography Education in the United States

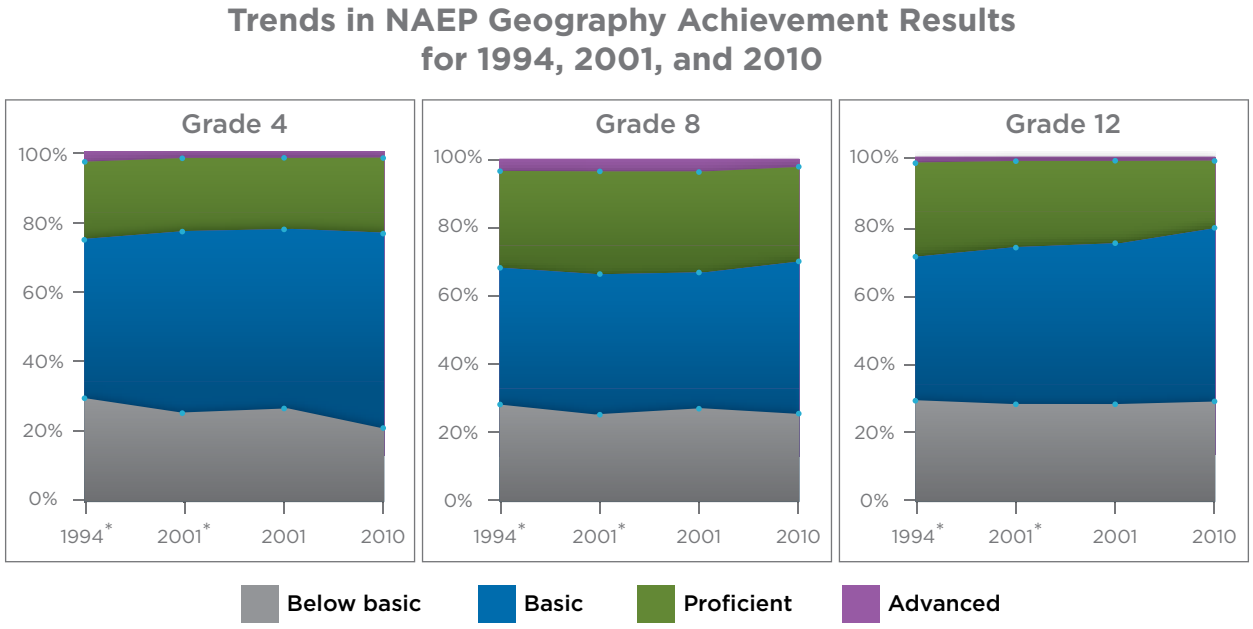
This report is one of three synthesis reports on geography education from the Road Map for 21st Century Geography Education Project. The Road Map Project has been a collaborative effort of four national organizations: the American Geographical Society (AGS), the Association of American Geographers (AAG), the National Council for Geographic Education (NCGE), and the National Geographic Society (NGS). These organizations share a concern that the dismal state of K–12 geography education across the United States is a threat to our country's well-being, and by extension, the well-being of the global community. The project partners share the belief that geography education is essential for preparing the general population for careers, civic lives, and personal decision making in contemporary society. It also is essential for the preparation of specialists capable of addressing critical societal issues in the areas of social welfare, economic stability, environmental health, and international relations. The Road Map Project partners fear that by neglecting geography education today, we are placing the welfare of future generations at risk.

While inspiring examples of highly effective geography education can be found in every part of the United States, the amount of geography instruction that the overwhelming majority of students receive, the preparation of their teachers to teach geography, and the quality of their instructional materials are inadequate to prepare students for the demands of the modern world.

Assessments of geographic concepts and skills confirm the failure of our educational system in geography, indicating that the overwhelming majority of American students are geographically illiterate. The 2010 National Assessment of Educational Progress (NAEP), known as "The Nation's Report Card,"(National Center for Education Statistics, 2011) found that fewer than 30% of American students were proficient in geography; more than 70% of students at fourth, eighth, and 12th grades were unable to perform at the level that is expected for their grade (NCES, 2011, Figure 1). At 12th grade, more than 30% of students scored below "basic," indicating that they had not mastered even foundational geographic concepts or skills.

From the NAEP results and other data, we conclude that an overwhelming majority of high school graduates are not prepared to do the ordinary geographic reasoning that is required of everyone in our society in the course of caring for themselves and for their families, making consequential decisions in the workplace, and participating in the democratic process. Furthermore, we conclude that more than 30% of high school students

Figure 1. Comparison of Results for Students in Grades 4, 8, and 12 on National Assessment of Educational Progress (NAEP) Geography Test in 1994, 2001, and 2010



*Test administrations in which accommodations were not permitted

Source: NCES, 2011

are so far behind that it is unlikely they will ever reach proficiency. To compare with textual literacy, this level of geographic illiteracy is analogous to having 70% of high school graduates unable to read a newspaper editorial and identify the assumptions, evidence, and causal connections in its argument.

The Importance of Geography Education

K–12 geography education is critical preparation for civic life and careers in the 21st century. It also is essential for postsecondary study in a wide range of fields, from marketing and environmental science, to international affairs and civil engineering.

Everyone in modern society faces personal decisions that require geographic reasoning. These decisions, such as where to live and how to travel from place to place, can have an enormous impact on one's life. We also must make decisions that have far-reaching consequences, such as which products to buy and how to dispose of them. While these decisions may seem insignificant, when they are multiplied by the number of people making them each day, they have enormous cultural, economic, and environmental repercussions for other people and places. Finally, in our democratic society, we all participate in societal decision making about public health, social welfare, environmental protection, and international affairs. In this era of such global challenges as ethnic and religious conflict, growing populations in poverty, increasing competition for limited natural resources, and degradation of the environment, it is essential that all members of society be prepared to make these decisions. Geography education helps prepare people for these tasks.

In addition, we need to provide young people with the opportunity to develop the understanding and interest to pursue the geography-dependent careers that are critical to our national interests. The Geo-Literacy Coalition, a consortium of businesses including Google, CH2M HILL, Esri, and the U.S. Geospatial Intelligence Foundation, had the following to say about the importance of geography education for our nation (National Geographic, 2011):

> [America's] inattention to [geography education] stands in contrast to the demand for geographically literate individuals in the workforce. There is substantial demand in both the public and private sectors for people who have the ability to interpret and analyze geographic information. The number of jobs for such analysts is growing rapidly, while the supply of Americans who can fill them is not. By not preparing young people for careers that depend on geographic reasoning, we are leaving ourselves vulnerable.
>
> In our global economy, the understanding and analytical skills developed through geography education are essential to make well-reasoned decisions about where to conduct business, how to conduct business in particular locations, and how to transport materials and goods from one location to another. Critical business choices such as where to build facilities, how to design a supply chain, and how to market to different cultures all require geographic reasoning.
>
> These skills are equally important for emergency preparedness, defense, intelligence, and diplomacy. In our government and military, we need individuals who understand the dynamics of specific locations well enough to prepare for and respond to emergencies. We need analysts who are able to track people and events around the world and put appropriate responses forward for decision-makers. We need people who are able to operate on the ground in every kind of foreign context and can read the cultural and physical landscape appropriately.

This Road Map Project is taking place against a backdrop in which many members of the global community are renewing their commitment to geography education. In Australia, a national curriculum is being introduced for the first time. In England, geography is a component of the recently introduced English Baccalaureate. In most of the world, geography holds a higher place in the K–12 curriculum than it does in the United States. In most countries, geography is required every year through age 16, in addition to history or other social studies subjects. In fact, the United States is almost unique in its treatment of geography as part of a single curriculum with history, government, and economics.

The Road Map Project partners believe that we, as a society, have a responsibility to prepare all young people for their personal needs and civic responsibilities, and we have a further responsibility to prepare sufficient numbers of young people for geography-dependent careers. We are not currently living up to those responsibilities, and we fear the consequences that our society will suffer if we continue to neglect geography education.

The Need for a "Road Map" for Geography Education

Over the past several decades, a small but dedicated community of geographers and educators has harbored concerns about the state of geography education and

has worked diligently to improve geography education. Their greatest success has been in establishing a firmer place for geography in K–12 education. The Elementary and Secondary Education Act (ESEA) of 2001 (January 8, 2002) recognized geography as a core academic subject, and all 50 states now have K–12 standards for geography. Geography has been included in the National Assessment of Educational Progress since 1994, and the College Board established an Advanced Placement exam for Human Geography in 2001.

However, these successes in improving the place of geography in the educational system have not been followed up with the levels of effort or resources necessary to bring about widespread improvement in the quality of instruction. As a result, educators and students who have had the good fortune of being impacted directly by the efforts of the geography education reform community have benefited enormously, but they represent a small minority. As measured by NAEP, there has been no broad improvement in students' learning of geography during the 17–year period of testing.

The project partners launched the Road Map Project with the goal of increasing the scale and accelerating the pace of efforts to improve geography education to meet our responsibility to prepare young people for the world they will inherit. The partners have two goals for this work:

- first and foremost, to make future efforts to improve geography education more strategic, focused, and coherent, so they can have greater and more enduring impact; and
- second, to provide a rationale for establishing requirements for geography education and allocating resources to improve geography education that accurately reflect its importance for our society.

This work targets the three audiences that are in the best position to effect improvement in our system of public education:

1. **Front-line professionals:** educators, teacher educators, developers, and researchers who directly influence instruction, assessment, and research;
2. **Policy makers:** individuals at national, state, and local levels who establish the goals and processes for public education; and
3. **Funders:** decision-makers in government and private organizations who provide the funding to support public education.

In planning the project, the partners identified five critical issues for improving geography education:

1. preparation and professional development of teachers,
2. instructional materials to support classroom instruction,
3. assessment of learning outcomes and instructional effectiveness,
4. research on teaching and learning, and
5. cultivation and maintenance of public support.

The partners divided these issues among four efforts, deciding to address the first four issues through synthesis reports to be developed by three committees of experts identified by the project partners:

The ***Instructional Materials and Professional Development Committee*** considered the current state of the instructional materials for teaching geography and the preservice and inservice education that teachers who are responsible for geography education receive. Based on this analysis and a review of the literature on the design of instructional material and the design of teacher professional development, the Committee formulated recommendations and guidelines for both instructional materials and professional development that will lead to improvements in instruction and in learning outcomes.

The ***Assessment Committee*** studied the current state of assessment in geography and reviewed its history. Based on their analysis of existing assessment practices and a review of the literature on assessment as a support for improving educational outcomes, the Committee formulated guidelines for developing assessment instruments and for conducting assessment that will lead to improvements in instruction and outcomes.

The ***Geography Education Research Committee*** reviewed the existing education and cognitive science research literature to identify gaps in our ability to answer significant questions about geography education based on research. Drawing on this analysis, the Committee formulated recommendations for research questions and approaches that will build a knowledge base to guide improvement efforts for geography education in the future.

For the final issue—developing and maintaining public support for geography education—the partners did not believe the existing knowledge base on public beliefs and attitudes about geography education would sup-

port the development of a synthesis report at this time. Instead, the partners initiated a pilot study of public beliefs and attitudes under the direction of the American Geographical Society.

Establishing a Destination: Goals for K–12 Geography Education

The value of a road map is that it enables you to select a route to your destination. Therefore, the first step in developing our Road Map for geography education was establishing a common destination. In education, destinations are expressed in terms of learning outcomes, so in the case of geography education, we will be able to say that we have reached our destination when our schools make it possible for all students to achieve the learning goals for geography that we have set for them.

Because the national geography standards were developed through an earlier collaboration of the project partners, they represent a logical choice of "destination." However, the members of the Road Map Project committees thought we should use this opportunity to consider alternatives as well. Therefore, as a collaborative effort across all three committees, we conducted an investigation into what it means to "do geography" in the 21st century and what that implies for the goals of K–12 geography education. The remainder of this chapter describes that process and its outcomes.

Establishing goals for geography education is no small challenge because geography is a broad field and it is constantly evolving. Fortunately, geographers and others have wrestled with this challenge for generations, and we were able to benefit from that work. Our investigation was guided by three criteria that we believe the goals for K–12 geography education should meet. Specifically, goals for geography education should:

1. reflect the essence of geography as defined by geographers;
2. convey the qualities of geography that capture its distinctive benefits as a subject of study; and
3. focus on the portions of geography that have the greatest value for students and society.

We approached the challenge of defining the goals for geography education from two perspectives—those of geographers and educators. To explore the perspective of geographers, we surveyed the existing literature on the nature of geography, and we convened current thinkers and practitioners for a workshop on "geographic thinking." At this workshop, we invited a wide variety of academic and practicing geographers, cognitive scientists, and individuals with other relevant perspectives to present on what it means "to think like a geographer" or "to do geography." To explore the perspective of geography educators, we examined the history of efforts to conceptualize geography education during the past half century. We summarize the findings of these investigations below.

Geographers on Geography

We started our investigation with a review of the ways that geographers have defined geography in recent decades. While there is great diversity of opinion among geographers about where the boundaries of geography lie, there is considerable consensus about its core. Geographers engage in a range of activities related to space, place, and the dynamic interactions of agents within and across spaces and places (Baerwald, 2010; National Research Council, 1997). As described in a recent National Research Council report (NRC, 2010), geography involves:

> documenting, analyzing, and explaining: 1) the location, organization, and character of physical and human phenomena on the surface of Earth; and 2) the interplay of arrangements and processes, near and far, human and environmental, that shape the evolving character of places, regions, and ecosystems (p. 10).

This report characterizes geography as being forward-thinking and essential to society for key issues including sustainability, economic stability, national security, and response to environmental change.

A consensus also has evolved in recent decades about the key themes of geography. Pattison (1964) identified geography's core as consisting of four "traditions," the spatial tradition, the area studies tradition, the human-land tradition, and earth science tradition. Taaffe (1974) identified three key organizers for geography: spatial organization, area studies, and human-land relationships. Contemporary geographers agree that the discipline focuses on a similar set of core ideas: spatiality, human-environment interaction, interconnections between places, and place-based and regional analysis (Abler, 1987; Baerwald, 2010).

Because geographers work on many of the same questions and problems as specialists in other fields, they have faced the challenge of differentiating geography from those fields. Susan Hanson confronted this challenge in a presidential address to the Association of American Geographers. In this presentation, Hanson

(2004) described the unique aspects of geography as "the geographic advantage," and she enumerated four aspects of this advantage:

1. Geography considers *the relationships between humans and environments.* Because of the traditional separation of social and physical sciences, other disciplines tend to focus on one or the other.
2. Geography *recognizes the importance of spatial variability.* Geography offers unique methodologies for investigating the way phenomena vary with location and explaining the place-dependency of processes.
3. Geography *considers the multiple and interlocking geographic scales at which processes operate.* Geography also offers unique techniques for studying phenomena and how they play out over multiple spatial scales.
4. Geography *integrates spatial and temporal analysis.* With its focus on spatial variability, geography offers unique techniques for integrating the analysis of variation over time with analysis of variation over space. Many other disciplines have focused on analysis of temporal variability without attention to the spatial dimension.

Evolving Conceptions of the Goals of Geography Education

In addition to looking at how geographers have characterized geography in recent decades, we also looked at the goals that geographers and educators have articulated for geography education over that same period. During the past 50 years, four efforts to conceptualize the goals of geography education have had nationwide influence. In our investigations, we looked both at the ways they characterized the goals of geography education and at the influence they had. We summarize what we learned in the paragraphs that follow. Across these efforts, we observed two important trends: (1) an increase over time in their richness and clarity, and (2) an ongoing struggle to present a balance between what it means to "understand" geography and what it means to "do" geography.

The High School Geography Project (1963 to 1971).

Today's efforts to improve geography education have their roots in the wave of educational reform initiatives that followed the Soviet Union's launch of the Sputnik satellite in 1957. One of these initiatives targeted geography education, and it set a tone that has influenced all subsequent geography education reform efforts. The NSF-funded High School Geography Project (HSGP) was an instructional materials development initiative with the goal of transforming high school geography (Association of American Geographers, 1966). In a reflection on the project, the project director said, "With little hesitation, teachers [who were consulted in the design of the HSGP] voiced the same litany of problems…dull textbooks, inadequately trained teachers, simple factual content… training in history not geography, lack of emphasis on geography in schools of education..." (Helburn, 1998, p. 212). HSGP attempted to address many of these concerns by creating instructional materials that engaged students and teachers in asking and answering geographic questions using data and simulations, and by building professional development opportunities around the curricula. Essentially, HSGP was an attempt to reconceptualize geography education as the integration of geography inquiry and geographic understanding.

In practice, the long-term impact of HSGP turned out to be more a result of its ideas than its implementation. The unconventional HSGP units entered a challenging implementation environment in the late 1960s and early 1970s. The objective was to create a dynamic, participatory learning environment in which students observed that geography is a conceptually rich and useful subject for daily life in their communities and the larger world. Although the units were favorably reviewed and supported with teacher training, they differed significantly from existing materials and teaching practices. Further, the learning outcomes that the inquiry-based units targeted could not be assessed using conventional testing. Consequently, the HSGP was not widely adopted in American high schools. However, the project did engage a community of academic geographers in K–12 education for the first time in more than a decade, and it introduced a concept of the goals and methods of geography education to a new generation of educators. These two impacts helped to lay the groundwork for the next wave of reform efforts in the early 1980s.

The Guidelines for Geographic Education (1984).

The next influential effort to reconceptualize geography education began in the early 1980s following the publication of *A Nation at Risk: The Imperative for Educational Reform* (National Commission on Excellence in Education, 1983), which, like the launch of Sputnik, triggered a wave of educational reform efforts across the curriculum. In 1984, a joint committee of the Association of American Geographers and the National Council for Geographic Education published the *Guidelines for Geographic Education,* which was designed to provide a clear, comprehensive set of national goals for K–12 geography education (Joint Committee on Geographic Education, 1984). The *Guidelines* established a concise framework for geography teaching that would be widely

adopted in schools, in teacher preparation programs, and among publishers of geography texts and curriculum materials. The *Guidelines* described geography as consisting of three basic elements:

1. a geographic perspective (spatial and ecological ways of viewing the world);
2. fundamental themes (Location, Place, Human Environment Interaction, Movement, and Region); and
3. core skills (asking geographic questions, acquiring geographic information, presenting geographic information, analyzing geographic information, and developing and testing geographic generalizations).

With these three elements, the *Guidelines* continued the effort begun with the HSGP to present a vision of geography that integrates knowing with being able to do.

Following the publication of the *Guidelines*, the Association of American Geographers, the American Geographical Society, the National Council for Geographic Education, and the National Geographic Society joined together to create the Geography Education National Implementation Project (GENIP), which aimed to translate the *Guidelines* into practice. During the ensuing five years, GENIP produced two additional documents to help educators to implement the *Guidelines:*

- *K–6 Geography: Themes, Key Ideas and Learning Opportunities* (Geography Education National Implementation Project, 1987), and
- *Geography in Grades 7–12: Themes, Key Ideas and Learning Opportunities* (Geography Education National Implementation Project, 1989).

These seminal publications extended the teaching examples in the *Guidelines*, and they were widely distributed, increasing the influence of the *Guidelines*.

The impact of the *Guidelines* was impressive. The publication was remarkably successful in achieving widespread awareness of the five fundamental themes. Educators and curriculum developers found the five themes to be memorable, relatively easy to understand, and easy to apply in teaching geography. Thus, the themes were widely integrated into school curriculum guidelines, preservice and inservice professional development, and instructional materials produced by publishers, school districts, and professional organizations through the concerted efforts of the nascent Geography Alliance network sponsored by the National Geographic Society. To this day, the five themes continue to influence geography education in many school settings and teacher preparation programs.

Unlike the content themes, however, the geographic perspectives and skills in the *Guidelines* received scant attention. They were largely overlooked in subsequent materials development and professional development efforts. While the five themes were consistent with the general focus on knowledge of the educational reform efforts of the 1980s, the perspectives and skills in the *Guidelines* were not. Like the inquiry-based elements of the HSGP, integrating these perspectives and skills into educational practices would have required a larger change than most educators were comfortable making, particularly because the reform efforts of the 1970s were widely criticized at that time for an excessive focus on "process" at the expense of "content."

The *Guidelines*, which had a much broader impact than the HSGP, led to a broad-based reconceptualization of the content of geography in mainstream education. However, its influence was largely limited to the conception of content in terms of the five themes it presented. The *Guidelines*' depiction of geography as an integration of content, perspectives, and skills was largely overlooked.

***Geography for Life: National Geography Standards* (1994).** The next major effort to articulate the goals of geography education began in response to federal legislation enacted in 1989. The Goals 2000: Educate America Act (1994) was passed in response to a renewed concern about the state of education in the United States. As a result of concerted efforts by the geography education community, geography was included as one of the five core subjects in the America 2000 reform plan. This recognition resulted in funding to create a national standards document for geography. (It was in this era that the term "standards" was introduced into the educational policy lexicon.)

With funding from the U.S. Department of Education, the National Endowment for Humanities, and the National Geographic Society, the four GENIP partners launched a standards-writing project. Over two years with extensive feedback and advice from a broad range of reviewers, advisory groups, and testimony at numerous public hearings, a diverse group of scholars and teachers created the first set of national standards for geography. In 1994, the product of this effort was published: *Geography for Life: National Geography Standards* (Geography Education Standards Project, 1994).

In contrast to the 26-page *Guidelines*, the 1994 edition of *Geography for Life* was 272 pages long. *Geography for Life* incorporated everything in the *Guidelines* in some form. For example, *Geography for Life* retained the *Guidelines*' three-part structure of perspectives, skills, and content. However, much was modified and added:

- The two geographic perspectives highlighted in the *Guidelines* were maintained in *Geography for Life*: spatial and ecological. They also were described in significantly greater detail than they had been in the *Guidelines*.
- The skills identified in *Geography for Life* are an elaboration of the skills described in the *Guidelines for Geographic Education*. They are: asking geographic questions, acquiring geographic information, organizing geographic information, analyzing geographic information, and answering geographic questions.
- Instead of the five themes discussed in the *Guidelines, Geography for Life* organized content around six essential elements (The World in Spatial Terms, Places and Regions, Physical Systems, Human Systems, Environment and Society, and The Uses of Geography). These essential elements were, in turn, made up of 18 content standards.

While *Geography for Life* took a large step toward presenting a picture of geography as integrating *knowing* and *doing* through its elaborate description of perspectives and skills, the authors were restricted by the constraints imposed on national standards documents at the time. Specifically, they were permitted only to use the term "standard" to label content objectives. For that reason, neither perspectives nor skills were described as standards in *Geography for Life*. However, the authors incorporated the application of geographic understanding into these content standards in two ways. First, two of the essential elements—*The World in Spatial Terms* and *The Uses of Geography*—describe the application of knowledge and understanding as content. For example, *The World in Spatial Terms* includes using maps and other geographic representations and technologies to report information from a spatial perspective; using mental maps to organize information about people, places, and environments in a spatial context; and analyzing the spatial organization of people, places, and environments on Earth's surface. *The Uses of Geography* element describes the application of geography to interpret the past and ways to apply geography to interpret the present and plan for the future. Second, for each content standard, the authors described what students should be able to do with that standard's content knowledge, implicitly reinforcing the importance of applying geographic knowledge.

Finally, *Geography for Life* helped to provide a well-rounded picture of modern geography by providing discussions of the nature of geographic inquiry and discussing why the study of geography is important. *Geography for Life* offered existential, ethical, intellectual, and practical reasons why individuals should learn geography, and the publication described how society benefits from having geographically informed citizens.

Like the *Guidelines for Geographic Education* a decade earlier, *Geography for Life* had a broad national impact on mainstream education. However *Geography for Life's* impact on classroom practice was largely indirect. Its direct impact was on educational policy. The publication's six essential elements were not as widely taken up by educators and curriculum developers as were the five themes. Even today, many textbooks and professional development programs still use the five themes as a central organizing scheme. On the other hand, *Geography for Life* has had an impact on educational policy that exceeds any other geography education document in the past 50 years. The release of *Geography for Life* provided impetus for all 50 states and the District of Columbia to establish state standards for geography, and it provided a model for them to follow. *Geography for Life's* content and structure were studied by the standards writers in every state, and its influence can be seen in nearly all of them.

As in previous documents, the balance between perspectives, skills, and knowledge that the authors of *Geography for Life* presented was not as influential as desired. Despite their prominence in *Geography for Life*, perspectives and skills are not nearly as well-represented in state standards as the content standards presented in the publication.

Geography for Life: National Geography Standards, Second Edition (2012). In 2007, the members of GENIP decided it was necessary to revise the national geography standards to reflect changes in the discipline of geography and in the world. The second edition of *Geography for Life: National Geography Standards* (Heffron & Downs, 2012) maintained the spatial and ecological perspectives and the 18 content standards of geography, and it extended and elaborated on the geographic skills section. Reflecting an important change in the world since 1994, it incorporates geospatial technologies for problem-solving into many of the standards. The writing team also completely revised the concepts and performance expectations throughout the

content standards based, in part, on new research in the learning and cognitive sciences. The new descriptions use consistent language for cognitive activities drawn from research in the learning sciences, and they reflect new understanding of developmental learning across the K–12 continuum.

The new edition continues to advance the notion that geography education should be framed around core ideas, many of which are applicable to peoples' daily lives, as well as personal and community decision making and problem solving. This edition makes the case that being an informed citizen requires knowing the content of geography *and* being able to use geographic reasoning and skills.

Choosing a Destination: *Geography for Life*

After careful review and consideration, all three committees agreed that the second edition of *Geography for Life* should serve as the "destination" for the Road Map Project, because it meets all three of the criteria we had established for the goals of geography education:

- ***Reflect the essence of geography as defined by geographers:*** In its presentation of the content standards, *Geography for Life* reflects the central elements that geographers have identified with geography.
- ***Convey the qualities of geography that capture its distinctive benefits as a subject of study:*** In its depiction of the perspectives and skills and its process-oriented content standards, *Geography for Life* captures the four components of the geographic advantage.
- ***Focus on the portions of geography that have the greatest value for students and society:*** In its focus on the scientific aspects of geography with practical applications, *Geography for Life* focuses on the portion of geography that the committees believe is most valuable for students to learn.[2] While *Geography for Life* does not capture the full diversity or richness encompassed by modern geography, the committees think it captures the subset that will be most valuable for students' personal, professional, and civic lives.

Describing the Destination: Effectiveness and Balance

Across the history of efforts to reconceptualize geography education summarized above, there has been an ongoing struggle to promote the multi-faceted nature of geography as perspectives, skills, and content, which is contrary to a tendency in the educational system to focus more narrowly on content. The multi-faceted view of geography presented by the second edition of *Geography for Life* contrasts with the stereotypical view of geography as being about facts, in particular, the locations and names of places. While this stereotype could not be more inaccurate as a description of the field of geography, it is distressingly accurate as a description of the geography education that American students experience.

If it is successful, the Road Map Project will change this reality over the next decade by increasing the reach and effectiveness of efforts to improve geography education. Each of the committee goals is designed to address a critical implementation issue: the preparation of teachers, the nature of instructional materials, the design and structure of assessments, and the research base to inform educational decision making. However, the success of all of these efforts hinges on the ability of individuals to communicate about the true nature of geography, including the geographic advantage, to key stakeholders. For that reason, we extended our consideration of the goals of geography education beyond what they should be to how they should be *expressed*. In doing so, we identified two important issues to address: (1) the need to present a view of the different aspects of geography that is balanced and integrated; and (2) the need to clarify what it means to "do geography."

A Balanced and Integrated View of Geography

The stereotypical view of geography as fact-based and descriptive has proven persistent, no doubt because the stereotype corresponds to the experience of most American students and teachers for generations. In practice, this "understanding gap" functions as a source of resistance to any efforts to change geography education. Making a significant change to geography education will require a change in the understanding of geography by all stakeholders. Introducing new concepts of subject matter has proven to be a difficult challenge in the American educational system, but this is an occasion where the geography education community has the opportunity to learn from the experiences of other disciplines. For example, the backlashes that have confronted both math and science education reform

[2] We characterize the geography presented by the second edition of *Geography for Life* as scientific because it employs methods of inquiry and standards of evidence that are associated with contemporary scientific practice. This subset of geography is sometimes referred to as the geographical sciences. By referring to this geography as scientific, we are describing its methods, not its content. *Geography for Life* reflects the consensus view of geographers that geography is concerned with both the social and physical worlds, and that it has a particular concern for the interactions between those worlds.

efforts teach us how important it is to present reform as a process of integrating traditional and new approaches, rather than as a replacement of traditional with new.

For that reason, it is essential that we present a balanced view of geography that recognizes the importance of learning the place names, locations, and terminology that have characterized geography education historically, along with understanding powerful geographic concepts, and being able to reason geographically. We must be careful not to present the new conception as being a rejection or abandonment of what has been valued traditionally, but rather as an enhancement that establishes a better balance. This lesson applies not only to stakeholders that have been untouched by earlier reform efforts, but also to those who have invested in those reforms. For example, educators who have embraced the richer conception of content presented by *Geography for Life* and its precursors should see a focus on geographic reasoning as an enhancement to their efforts, rather than as a replacement of them.

To help stakeholders understand the value of this multifaceted geography (and to motivate them to support it), it is essential that we communicate the limitations of the traditional focus of geography education on its own and the value of the additional components for learners. It is essential that we do so in terms that are meaningful to stakeholders (e.g., "college and career readiness" is the discourse of educational policy as this report is being prepared, as well as preparation for personal and civic life).

For pedagogical purposes, it also is important that we communicate the importance of integrating the different facets of geography in education, rather than teaching them separately. Educational research teaches us that it is ineffective to separate learning of facts, concepts, and reasoning because they need to be used together in practice. However, a traditional view, and one that would feel more comfortable to many stakeholders, would be that factual understanding should be taught first, followed by conceptual understanding, and then reasoning skills.

Therefore, it is essential that we present a view of geography education that integrates learning of facts, concepts, skills, and reasoning at all levels from K to 12.

Geographic Practices

In reviewing the history of geography education reform, we see that the aspect of geography that has been taken up the least in schools is the application of geography understanding to answer questions or to solve problems. Where the articulation of the five themes in the *Guidelines* led to a broader understanding of geography content among the educators who were reached by it, historically there has been no comparable broadening in the understanding of the practices of geography.

As a result, all three committees have paid special attention in their work to the question of how to ensure that "thinking geographically" and "doing geography" become integrated into classroom practices in the next generation of geography education reform. Over the course of our work, we identified terminology as an issue. *Geography for Life* uses the term *skills* to describe the activities that constitute the doing of geography. However, concerns were raised by how well the term *skills* describe the complex, goal-directed behaviors that constitute geographic practice. In the course of our research, we found an alternative in the science and mathematics education literature—the word *practice* has been adopted in recent years as a term for these kinds of activities we were trying to capture. In that literature, the term *practice* is used to describe the behaviors that comprise scientific inquiry and problem-solving. A scientific practice is a goal-directed set of actions that contribute to a scientific inquiry or problem-solving process. Some of the scientific practices identified in the National Research Council's recent *Framework for K–12 Science Education* are asking questions, defining problems, developing and using models, constructing explanations, and engaging in argument from evidence (NRC, 2012, p. 49). Practices are shared across disciplines, but they typically are conducted in different ways across different disciplines (NRC, 2011). In this respect, discipline-specific practices encode the perspectives of the discipline.

Working from the skills described in *Geography for Life*, we identified six categories of geographic practice. Each of these categories represents an aspect of geographic inquiry or problem-solving, and encompasses specific practices that, either independently or in combination, can achieve a reasoning goal (Table 1). More detailed descriptions of the practices, along with examples representing how they are used by practicing geographers, ordinary people, and classroom instructors, can be found throughout the three Road Map Project committee reports.

Because it suited their goals better, the Geography Education Research Committee condensed these six categories into a smaller set. The Committee combined acquiring, organizing, and analyzing geographic information into a single category, and also combined

answering questions and designing solutions with communicating geographic information. Thus, the Committee's three categories are:

1. Formulating geographic questions;
2. Acquiring, organizing, and analyzing geographic information; and
3. Explaining and communicating geographic patterns and processes.

Mapping a Bright Future

In this chapter, we have presented an overview of the rationale and goals for the Road Map for 21st Century Geography Education Project. The project is motivated by a concern for the current state of geography education and the slow progress partners and others have made in improving it. By identifying promising strategies in key areas, we aim to mobilize and focus resources in ways that will increase the magnitude and pace of improvement. The remaining chapters in this report provide an analysis of key issues for geography education, and offer recommendations for how to focus improvement efforts during the coming decade. In doing so, this report joins the other Road Map Project reports in laying out a path toward the destination described in *Geography for Life*—an integrated geography education that balances learning of knowledge, understanding, and practices.

Table 1: Geographic Practices[3]

Categories	Practices
Posing geographic questions	a. Identify problems or questions that can be addressed using geographic principles, models, and data; express problems and questions in geographic terms.
Acquiring geographic information	a. Identify geographic data that can help to answer a question or solve a problem.
	b. Collect data (including observations and measurements) about geographic phenomena, and/or gather existing data to help answer a question or solve a problem.
Organizing geographic information	a. Organize data and create representations of data to help solve a problem or answer a question.
Analyzing geographic information	a. Identify data analysis strategies that can be used to help solve a problem or answer a question.
	b. Find and describe spatial and temporal patterns in data, or find data that matches a pattern, to help solve a problem or answer a question.
	c. Construct an explanation or prediction for phenomena by comparing data to a model or theory.
Answering questions and designing solutions	a. Construct an answer to a question or a solution to a problem using geographic principles, models, and data.
	b. Evaluate one or more answers to a question or solutions to a problem using geographic principles, models, and data.
Communicating geographic information	a. Inform or persuade an audience using geographic principles, models, and data.

[3] While the categories and practices are listed sequentially in the table following a widely used model of inquiry and problem-solving, we make no assumption that they will or should be conducted in that order in practice.

A Road Map for Geography Assessments

This particular report focuses on assessment and how assessment can be used as a mechanism for improving geography education. The Leadership Team of the Road Map for 21st Century Geography Education Project identified assessment as a topic for a consensus report because of the central role that assessment plays in all aspects of education reform, and because of the need for a critical review of how assessments are being used to monitor and improve geography education. In this report, we examine the roles that assessments can play in advancing the goals of geography education and how existing frameworks and assessments are currently serving those roles. We present a framework for geography assessment that will contribute to the larger efforts to improve geography education over the next decade.

We begin our discussion in Chapter 2 with a description of the various roles that assessment can play in the improvement of education. In order for assessments to have the beneficial impact on educational practices and outcomes, they must be thoughtfully designed, implemented, scored, and interpreted. Chapter 2 describes some of the essential considerations for designing effective assessments, and how assessment frameworks can support the development of high-quality assessments that advance the goals of education reform.

In Chapter 3, we consider the current state of assessment in K–12 geography education. We begin with a discussion and comparison of existing frameworks for large-scale assessments in geography and closely related fields. We then report on a study we conducted of existing geography assessment items designed for classroom and for large-scale assessments. We analyze how well current assessment practices meet the goals outlined in the standards and frameworks, and how well they satisfy the criteria for effective assessments. Based on this analysis, we argue that there are weaknesses in current assessment practices that need to be addressed as part of any effort to improve geographic education, and that there is a clear opportunity for assessment to more effectively support improvement of geographic education.

In Chapter 4, we respond to the issues raised in the preceding chapters with a proposal for a new approach to assessment in geographic education. We present this approach in the form of a new framework for the design of assessments that we call the *21st Century Assessment Framework for the Geographical Sciences* (AFGS21), and we describe how this framework builds upon the existing frameworks to move toward the goals for geography education outlined in Chapter 1. We also describe how this general framework can be used as a blueprint for specific assessment frameworks targeted at specific learning objectives and learner populations.

In Chapter 5 we describe how AFGS21 is designed to be used by showing an example of a specific assessment framework. We describe the process of creating this framework from AFGS21 for an end of unit exam for a high school course. Chapter 3 also provides a detailed description of the steps taken to create this framework, and how to use the framework to develop assessments that meet the criteria for effective assessments laid out in Chapter 2.

Finally, Chapter 6 presents recommendations for specific steps toward the implementation of the new approach to assessment. This approach, combined with the recommendations in the other two Road Map Project reports, will lead to meaningful improvements in the effectiveness of geographic education over the next decade.

In Appendix A, readers can find additional details about the study of existing assessments described in Chapter 3. Appendix B has the full assessment framework described in Chapter 5.

Chapter 2: Improving Teaching and Learning Through Assessment

Assessments have always played an important role in education because of the information they can provide for decision making. Teachers use information from assessments to adjust their instruction. Schools and school systems use information from assessments to place students, evaluate programs, and allocate resources. Assessments play an important role in communication as well. At the classroom level, the form and content of an assessment informs students about what is expected of them. At the system level, the form and content of assessments serve to inform teachers, students, and other stakeholders about the priorities of the system.

Because of these multiple roles, assessment is increasingly being recognized as a point of intervention for reform of teaching and learning. For example, at the classroom level, assessments that reveal more information about what students think, know, and can do enable that teacher to make better decisions about instruction. At the school and system levels, efforts to change the focus of instruction can be supported by changes in the focus of assessment.

The premise of this report is that assessment can be used in these ways to improve teaching and learning in geography. In a nutshell, the question that the Road Map Project Assessment Committee has investigated is *how can assessment in geography support improvements in teaching and learning of geography?*

In this chapter, we begin our consideration of that question with a discussion of the different ways assessment data can be used in education. This discussion informs our recommendations about how assessments should be used to improve teaching and learning in geography. We continue with a discussion about the challenges of designing assessments to serve these purposes, which informs our recommendations about how to design assessments to serve our recommended uses.

Uses of Assessment Information in Education

Assessment is commonly viewed as a separate activity from instruction. Indeed, in the Committee's survey of geography assessments, fewer than 60% of geography curriculum units included assessments (Appendix A). The result of this separation is a missed opportunity for educators, students, parents, and policy makers to improve teaching and learning (Black & Wiliam, 1998, 2004a, b). By contrast, we believe that assessment must be integrally connected with teaching and learning (Figure 3). Each may move to the forefront at different points in time, but they must maintain a strong link to the others. In this report, we focus primarily on the assessment vertex of the teaching-learning-assessment triangle, where evidence from assessment informs and improves teaching and learning.

Figure 3. Teaching-Learning-Assessment Triangle

Different types of assessments provide different kinds of evidence. **Formative assessments** are used to monitor students' progress during instruction to inform the teacher of how to adjust instruction to suit their needs. Formative assessments can be used to identify gaps between a student's current level of mastery and the objectives for that student, and those gaps can be used for selecting targeted instructional approaches. **Summative assessments** are used following instruction to evaluate the outcomes of the instruction. They can be used to measure students' learning achievements and to inform programmatic changes to improve future instructional approaches.

Both formative and summative assessments provide valuable information for decision making. In this report, we focus on four primary uses of information from assessments:

1. instructional decision making,
2. measuring individual achievement,
3. evaluating programs, and
4. educational research.

Assessments That Inform Instructional Decision Making

Formative assessment is designed to inform ongoing teaching and learning. Formative assessments can take many forms, from formal tests to oral questions posed on-the-fly, but the key feature of these questions is that they are intended to make students' thinking apparent, so the teacher can monitor learning while instruction is

still taking place. Because formative assessments monitor students while a lesson or unit is under way, educators can use the results from the assessment to adjust instruction to either bring attention to the students' particular needs or to move past objectives that students have already mastered. Students can use the results of formative assessments to understand where they have made progress and where they need additional work. Effective instructional practices invariably involve ongoing, informal formative assessment in real time every day (Wiggins & McTighe, 2005); nevertheless, formative assessments may involve more formal assessment tools, which can be used at critical points in an instructional sequence.

Naturally, implementation of the information gathered from formative assessment is critical. If the information is not used, the purpose of formative assessment is not achieved. If feedback is provided to individuals or groups of students, it is important to make sure that it focuses on strategies that students can use to advance their learning. Feedback that focuses only on correctness often does not provide enough information to students about how to improve. For formative assessment to have the greatest impact on improving learning, students should be given the opportunity to use the feedback, and teachers should verify that it was used appropriately. Experts generally discourage the use of formal scoring or grading in formative assessment to maintain the focus on the learning process, rather than on achievement (Black & Wiliam, 1998).

Because formative assessments are used during instruction, consideration must be given to the amount of time they require and the degree to which they distract from other learning activities. The introduction of rapid assessment tools, such as classroom "clickers," that allow a teacher to collect immediate information about students' ideas with minimal disruption of the class, and computer-based assessments embedded in instructional activities that provide instantaneous feedback to students and point them toward useful lessons, represent promising innovations for integrating assessments more seamlessly into learning.

Formative assessments, even informal ones, must have clear goals for the information that the educator is trying to obtain. For these assessments to provide the necessary information to guide teaching and learning, they must enable the educator to identify what the students know, what they can do, and what they still have to learn. This requires that the assessment be based on a clearly defined, fine-grained target competency. That is, formative assessment has a narrower focus when compared with summative assessments. Each item focuses on a narrow scope of knowledge, practices, or competencies. Once a target competency, or construct, has been described explicitly, effective assessment strategies can be developed to identify where students are with respect to the target.

The central benefit of formative assessment for improving teaching and learning is that it provides information to teachers and students that can help tailor the teaching and learning process to the specific needs of the students. Whether these are questions to students in the course of a hands-on activity or quizzes at critical junctures during instruction, they can provide critical information about the progress of students and how to focus the remaining instructional time.

Assessment for Measuring Individual Achievement

A second purpose for assessment is to measure individual achievement. This is done through assessments that follow instruction. These summative assessments typically are used to measure students' competencies against the benchmark learning goals; the results are used to determine if students sufficiently met those goals and satisfied their requirements, or the results are used to rank students relative to each other.

Assessments of achievement have more stringent requirements for technical quality (i.e., validity and reliability) compared with formative assessments. These assessments typically take the form of tests that cover a broad span of learning objectives, which contrasts with the targeted questions on a specific learning objective typical of formative assessments. To characterize students' competency across a domain, assessments of achievement tend to rely on extrapolation from a student's performances based on a relatively small sample of the domain. If such a sample is to accurately represent a student's competency, the assessments should be held to stringent requirements for technical quality. For example, they must be designed to ensure that students interpret the items in the same way, and that their answers can be interpreted and evaluated accurately and consistently. Enforcing high standards for technical quality ensures that decisions are made based on valid, reliable, and fair information.

Assessments of achievement also have benefits for improving teaching and learning. However, rather than supporting short-term decision making within a unit of instruction as formative assessments do, they support

longer-term decisions about students' educational paths, such as whether they are prepared to move to higher levels, which courses and programs of study they should enroll in, and whether they should be admitted to selective educational programs or schools.

Assessment for Program Evaluation

Another use for summative assessments is program evaluation. Assessment information can be used to determine the effectiveness of materials, approaches, and educators. In the case of program evaluation, assessment information can be used to determine if some aspect of instruction is achieving its goals or if it should be modified in the future. For example, a school district might decide to revise or replace a set of instructional materials based on summative assessment data. Alternatively, an administrator might identify a group of teachers to participate in targeted professional development because assessments reveal that their students are struggling with the same set of skills.

As with the other purposes for assessment, assessments for program evaluation must be designed to serve the specific decision-making needs. Assessments for program evaluation must be sensitive to program effects, so they can collect accurate information about how well the program is achieving those outcomes and identify areas that require improvement. As in assessments of individual achievement, assessments for program evaluation must meet high standards for technical quality, providing information that is valid, reliable, and fair for making well-founded decisions.

Assessments for program evaluation also have great value for improving teaching and learning. Like assessments of achievement, they enable decision making over a longer time period than do formative assessments. They also enable improvement over a broader scale. They enable educators, administrators, and policy makers to make decisions about instruction based on one group of students that will impact subsequent groups of students.

Assessment for Educational Research

Scholarly or scientific inquiry in geography education often uses assessments to improve scientific understanding of teaching and learning. Assessments for research can provide insight into what students think, how they learn, and the efficacy of different approaches to instruction.

As with the other types of assessments, assessments for research must be reliable and valid for their intended uses, and they also require careful attention to how students interact with the assessment itself. Using assessments that do not measure the appropriate outcomes, because of poor technical quality or they are weakly aligned to the desired competencies, may lead to inaccurate conclusions. Therefore, for data from assessments to be useful for informing geographic research and development practices, technical rigor is of utmost importance.

Assessments for research contribute crucial information for improving teaching and learning. However, they generally have the longest time-delay between assessment and impact on students. Whereas classroom formative and summative assessments provide immediate information about specific students and program evaluation provides information about specific programs or educators, research provides information about students or educational approaches in general. Research findings require sufficient quantities and diversity of data to support generalizations, which can require multiple cycles of assessment. Once the research findings have been communicated, they must still be interpreted and applied to specific cases before they can improve the quality of teaching or learning. So, while research findings have the benefit of being general and of being supported by high standards of evidence, they have the longest delay between their administration and their impact on teaching and learning.

Considerations for the Design of Assessments

The design of sound assessments is a considerable challenge. In this section, we discuss important considerations for the design of assessments that can be relied upon to serve their purposes and improve teaching and learning.

Assessment development should always start with the following:

1. determining who the user of the assessment information will be (e.g., student, teacher, school and district administrator, state department of education);
2. identifying the population to be assessed (e.g., grade level, special needs, geographic distribution);
3. describing what information the user wants to obtain from the assessment (e.g., mental models, subject mastery, achievement); and
4. clarifying how the information will be used (e.g., teaching, course placement).

Assessment tasks must be carefully designed to elicit the content knowledge, practices, and cognitive processes that decision makers want to know about and that will provide the appropriate data.

Table 2. Characteristics of Assessment Design for Different Purposes

	Description	Context of Administration	Common Formats	Practical Considerations
Inform Instructional Decision Making	• Informs teaching • Informs learning	Classroom assessment	Informal and formal formative assessments embedded in instruction, such as: • Oral questioning • Observing • Informal evaluation (e.g., "stoplighting," thumbs up/down) • Constructed-response items • Selected-response questions	• Frequency of assessments • Evaluating performance and feeding back • Sampling practices (e.g., one student, some students, whole class) • Number of critical junctures or transition points in a unit • Time to administer and score an assessment
Measure Achievement	• Determines qualifications for advancement • Student accountability • Diagnose students' problems	Classroom or large-scale assessment (includes classroom, school district, state, national or international level)	• Selected-response items • Constructed-response items • Performance tasks • Interactive computer tasks	• Sampling of content and cognition to be measured • Length of the assessments • Time to administer the assessment • Logistics and delivery • Availability of technology
Evaluate Programs	• Monitoring for program improvement • Make comparisons between programs • Measure program effectiveness	Large-scale assessment (includes program, school or classroom reform, school district, state or national level)	• Selected-response items • Constructed-response items • Hands-on performance tasks • Interactive computer tasks • Interviews • Observations • Surveys	• Content and cognition to be measured • Length of the assessment • Time to administer the assessment • Logistics and delivery • Availability of technology
Research	Provides empirical basis for model building/testing or for policy or practice decisions about, for instance: • curriculum/program development, • teacher training • assessment • instructional program development	Small scale studies • Case studies • Ethnographies • Design studies • Quasi-experiments • Randomized experiments • Observational (longitudinal) Large-scale studies • Quasi-experiments • Randomized experiments • Observational (longitudinal and cross-sectional)	Different types of assessment can be used depending of the research being pursued, such as: • Interviews • Observations • Constructed and selected response items • Surveys	Depending on the type of study, sampling issues should be considered: • Representative of high and low achievement • Representative of different demographic groups • Random sample • Matched samples • Randomization to conditions • Selectivity bias

Pellegrino and colleagues developed a very useful three-component model for thinking about the design of assessments (National Research Council, 2001). The three components of the model are cognition, observation, and interpretation:

- **Cognition.** For responses to an assessment to be considered representative of what students know, each assessment item must be grounded in a model of how people learn that can be used to interpret what students think and know about the domain based on their responses.
- **Observation.** The assessment provides evidence for what the student can do (and by inference, what the student thinks—see below) while engaging in the specific assessment task.
- **Interpretation.** Translating the evidence from an assessment into a well-founded depiction of what the student knows about the domain requires more information than the assessment can provide. Therefore, the observations of what the student can do as demonstrated by his or her performance on the assessment must be applied to the model of cognition in order for the assessor to draw inferences about the student's competence in the domain.

Cognitive models can range from simple (e.g., a wrong answer means the student has an incomplete understanding of the domain) to complex (e.g., each answer, right or wrong, provides insight into the student's conception of the domain or "mental model"). Sophisticated cognitive models tend to have some empirical basis, often derived from a combination of cognitive psychology and education research.

The interrelationship among the cognitive model, observations, and interpretation shapes what (and how) information can be gleaned from an item. Together, they form the conceptual basis underlying an assessment. Decisions about the structural characteristics of assessment items, however, should be based on the user's purpose or conducting the assessment.

As discussed above, assessments can serve four main purposes: to assist teaching and learning, to measure individual achievement, to evaluate programs, and to conduct research. One type of assessment cannot fit all purposes (NRC, 2001). Therefore, some considerations are needed in determining the characteristics of an assessment to fit its purpose. In Table 2, we outline the common types of assessments used for each of the four purposes, how and when they tend to be used, and practical considerations for their design. We also provide a brief summary of the main characteristics, goals, and considerations for each assessment type.

Assessment items must be designed around the information the user wants to obtain and the interrelationship among cognition, observations, and interpretation that forms the theoretical foundation of the item, but there also are a number of decisions that must be made about the item's structure and delivery to ensure that the item serves the needs of the user most efficiently and effectively. The primary considerations include characteristics of the item format, item quality, and cost-effectiveness of delivery and scoring.

Item Characteristics

Assessments, regardless of purpose, are composed of three components: a task, a response, and an evaluation of the response. (The combination of task and response is what we refer to in this report as an "item.") The specification of these three components for a particular assessment determines the format of the assessment. The distinction between the components is important because, for example, a task that requires an extended constructed response, such as an open-ended question, is not necessarily an assessment, but it can become one if it is evaluated using a scoring system, even if it is an informal one. On the other hand, the task would not be considered an assessment if the other components are not employed; this is because there is no system established for evaluating responses. That is, an assessment is incompletely specified if it has only a task and response, because there must be interplay between the design of the response format and how the response will be interpreted.

1. **Task type.** Assessment tasks vary from asking for a definition, to solving a problem, to recommending a course of action. The task can be presented informally on-the-fly (e.g., a question posed during a lecture), or it can be formally prepared for class (e.g., test questions). Tasks can involve "hands-on" performances or presentations, or projects. Ideally, students are given opportunities to demonstrate their competencies with a variety of task types.
2. **Response mode.** Students' responses to questions or tasks can include selection from options (multiple-choice), constructed-response tasks that require writing a few sentences (short constructed answer) or a paragraph or more (extended constructed answer), oral response (e.g., presentations), performance of a task such as an experiment (performance

assessments), a portfolio of students' work, and so on. All of these modes can be used formatively in a classroom or for summative and program evaluation purposes. Selection of response mode depends in part on what is being tested. Some formats tend to tap different cognitive processes: some are particularly useful for assessing factual recall (e.g., multiple choice), while others are useful for observing students' abilities with reasoning through explanatory models (constructed response). Selection of response mode also is influenced by the amount of time that each response demands during an assessment, and the amount of resources required to develop and score it. For instance, an essay might provide extensive insight into a student's thinking about one question, but the amount of time it takes to complete imposes limits on the amount of other concepts and skills that can be evaluated during the assessment.

3. **Scoring system.** A scoring system converts students' responses to a numerical or other ordered scale. A scoring system is a form of "interpretation model," and the complexity of the interpretation depends on the nature of the task, the response format and the context of the assessment. For example, selected response items are generally scored as right or wrong (assuming there is only one correct answer). Scoring is even more complex with constructed-response items, since the interpretation model must account for all of the possible paths that students can follow to answer a question, and the scoring rubric must take into account all possible correct responses along each of those paths.

Item Quality

It is important to recognize two issues about assessment: (1) any assessment is only a sample of what students know and can do out of a large universe of what might be tested, and (2) any conclusions about what students know and can do require interpretation mediated by a process of reasoning from evidence obtained from the assessment; therefore, assessments are always imprecise to some degree (NRC, 2001). There are common principles that ensure the quality of assessments across all item types. The two central principles are:

1. **Measure what is intended.** An assessment should be designed based on a clear idea of what it intends to measure, or "the construct." The construct includes the content and/or the practices that are being assessed. Articulation of the construct gives focus to the assessment development process and provides a benchmark for verifying that the assessment will measure what it is intended to measure.

2. **Meet technical requirements.** Technical requirements ensure that the information gathered using the assessment is valid (i.e., the assessment tests the competencies it claims to test), reliable (i.e., results are consistent), and fair (i.e., the assessment does not favor a subset of the population being tested).

Cost Effectiveness

The consideration of cost effectiveness requires assessment developers to recognize that the practical reality that the ideal assessment may be too costly, time consuming, or logistically unwieldy to implement. Inevitably, compromises have to be made between what is measured and what we would like to measure. In selecting the type of assessment format, issues of time and cost should be considered, while holding true to the construct to be measured. Assessments designed to measure geographic practices, for example, are most likely to be problem-based performance tasks that entail multiple steps, making cost, time, and logistics considerable concerns. Students will likely be evaluated based on fewer items due to the time and financial costs associated with performance tasks. A mixture of performance tasks that provides the opportunity to present an in-depth view of students' thinking and doing, with multiple-choice items that sample foundational competencies, often satisfies the need to balance reliable and cost-effective measures.

The Role of Frameworks in the Design of Assessments

Because the creation of assessments that effectively implement all of the design criteria discussed above is so challenging, educators have developed tools to guide the assessment design and development process. One of the most powerful of these tools is the *assessment framework*. An assessment framework serves as a blueprint for assessment design and development, providing guidelines to assessment developers about how to construct an assessment. Assessment frameworks are generally designed around a specific purpose, body of knowledge or skills to be assessed, and population to be assessed. A framework can play an important role in maintaining coherent goals for different assessments. For example, a framework can be used to construct both formative and summative assessments for a course, ensuring that both

are monitoring the same learning objectives. An assessment framework also can play an important role in instructional design. In the approach commonly known as "backward design" (Wiggins & McTighe, 2005), instructional designers work backwards from assessment goals to design instruction and formative assessments.

An assessment framework plays multiple critical roles in the process of designing an assessment. A framework conveys what *knowledge, skills,* and *practices* should be assessed (and sometimes what should not) to those people responsible for creating a particular assessment, and a framework provides guidance about the form the assessment should take, including the types of items to use. An assessment framework provides a basis for validating the claim that the assessments accurately reflect the learning goals they are designed to assess, and a framework also provides a means for maintaining consistency among assessments developed by different people at different times and different places for different purposes.

Typically, the core of an assessment framework is a matrix organized around two dimensions. One dimension, often called the content dimension, describes the concepts and principles from the content domain to be covered by assessments. The other, often called the cognitive dimension, organizes behaviors (i.e., what students should be able to do). The names for the second dimension vary. For example, some assessment frameworks define the second dimension as "practices" (Table 3), and others define it as "cognition." (Table 4).[4]

Table 3. Framework Outline Using Science Content and Science Practices

		Science Content		
		Physical Science Content Statements	**Life Science Content Statements**	**Earth and Space Science Content Statements**
Science Practices	**Identifying science principles**	Performance expectation	Performance expectation	Performance expectation
	Using science principles	Performance expectation	Performance expectation	Performance expectation
	Using scientific inquiry	Performance expectation	Performance expectation	Performance expectation
	Using technological design	Performance expectation	Performance expectation	Performance expectation

Source: Reproduced from NAEP Science Framework (NAGB, 2008).
Note: This framework uses science content and science practices as the two dimensions of the framework.

Table 4. Framework Outline Using Geography Content and Geographic Cognition

		Content Dimension		
		Physical Science Content Statements	**Life Science Content Statements**	**Earth and Space Science Content Statements**
Cognitive Dimension	**Knowing**	Where is the world's largest tropical rain forest?	What mineral resources are often extracted by strip mining?	What factors stimulate human migrations?
	Understanding	Why are tropical rain forests located near the equator?	Explain the effects of strip mining and shaft mining on the landscape.	Explain the motivations of modern-day Mexicans and Cubans for immigrating to the United States.
	Applying*	Support the conclusion that tropical rain forests promote wide species variation.	How can both economic and environmental interest be reconciled in an area of strip mining?	Compare current settlement and employment patters of Cuban and Mexican immigrants in the United States.

Source: Reproduced from NAEP Geography (NAGB, 2010).
Note: This assessment framework crosses content with cognition for the two dimensions of the framework.
**Applying a range of higher-order thinking skills.*

[4] There is an important distinction between approaches to the cognitive dimension that is not always reflected in the naming convention. In some cases, the cognitive dimension focuses on cognition (mental behaviors that are not observable), and in others, the cognitive dimension focuses on external activities (observable behaviors). In some cases, the distinction is signaled by the names "cognition" versus "practices." Regardless of the naming convention, it is important that assessment developers be conscious of the difference between cognitive and external behaviors, and that they outline goals for both.

In the two dimensional matrix, a cell represents the combination of specific content with a specific cognitive behavior. Thus, the cells of the matrix inform test developers about what to assess. The contents of cells typically are called *performance expectations* (e.g., Tables 3 and 4).

In order for an assessment framework to guide assessment developers, it needs to include information about the characteristics of an assessment, including the nature of the items to be included. This information can include the types of items (e.g., multiple-choice, short answer, performance assessment), the distribution of those items across content and practice areas (i.e., the relative weight assigned to each performance expectation), and the levels of performance (e.g., below-basic, basic, proficient, advanced). As noted above, the specification of "item types" arises in recognition that assessment items tapping a student's ability to recall a fact or concept would be different from assessment items tapping a student's ability to design a well-controlled experiment or to use multiple documents to solve a problem. Put another way, different kinds of assessment items are needed to tap different performance expectations, and the assessment framework can organize the distribution of *how* items assess student abilities, in addition to *what* the items assess (see Table 5).

To illustrate how assessment frameworks can be used to guide assessment development, we provide three examples. In each of these examples, we show how content and cognitive dimensions can be combined to describe a performance expectation, then we show how a performance expectation and item characteristics together can provide the specifications for an item.

In these examples, the framework is based on the general assessment framework for geography presented in Chapter 4.

Example 1. The first example is for grades 3 to 5. The performance expectation derives from the combination of a content statement about sources of conflict and the geographic practice of analyzing patterns on maps to generate explanations or predictions. This performance expectation, when combined with the desired multiple-choice format represents a full specification of an item.

Level: Grades 3 to 5
Content: Conflicts arise when there is disagreement over the division, control, and management of Earth's surface. This content is from the second edition of *Geography for Life: National Geography Standards* (Heffron and Downs, 2012), (Standard 13, Theme 4, Grade 4).
Cognition: Analyze geographic patterns on a map or other representation.
Performance expectation: Analyze patterns on a map or other representation showing distribution of resources and borders or human settlements in order to explain or predict where disagreements over the division, control, and management of Earth's surface would occur.
Item characteristic: Answers are multiple-choice with map or other geographic representation.

Figure 4 shows Item 1, which provides an example of an item that fits these specifications. Item 1 displays a map of a fictitious region where natural resources are distributed across the territory of several nations, including one oil field that is bisected by the border between two nations, and the item asks students to identify a location that has the potential for conflict in that region.

Example 2. The next example is targeted at high school students. It shows how multiple performance expectations in a framework can be assessed together. In this case, the performance expectations derive from the intersection of a single content category with three cognitive categories. The content category concerns cooperation; and the cognitive categories deal with knowing and understanding, answering questions, and communicating. Below, we describe the three sets of item specifications that are covered by Item 2.

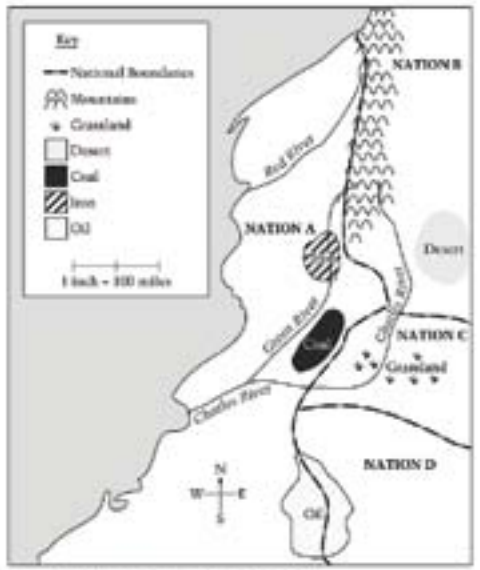
Table 5. Test-Item Weighting for Grade 12, Showing the Distribution of Items That Fall into Each Category for the Grade 12 NAEP Geography Test

Content Cognition	Space and Place	Environment and Society	Spatial Dynamics and Connections	Total
Knowing	9	9	12	30
Understanding	9	9	12	30
Applying*	12	12	16	40
Total	30	30	40	100

Source: NAEP Geography Framework (National Assessment Governing Board, 2010)

Figure 4. Item 1: Identifying Potential Conflict in the Region

Item 1



Which two nations are most likely to have a conflict over mineral resouces?

A. Nation A and Nation B
B. Nation A and Nation C
C. Nation A and Nation D
D. Nation C and Nation D

Source: 2001 NAEP Geography, Grade 4.
http://nces.ed.gov/nationsreportcard/itmrlsx/search.aspx?subject=geography

Figure 5. Item 2: Cooperation between Countries and Organizations

Item 2

You are a member of a **Washington, DC, Think Tank** that gives advice to the U.S. State Department. One of the issues you have been considering is the extent to which the United States needs to cooperate with other nations in order to be prosperous. Some members of the Think Tank argue that since the end of the Cold War, U.S. prosperity is much less dependent upon cooperation with other nations. Others, however, maintain that cooperation among nations, including the United States, is even more important today to promote prosperity. In which camp are you?

In the space below, write a memo to the other members of the Think Tank taking a position on the issue. Be sure to provide good reasons (evidence and examples) to justify your conclusions.

Source: Alternative Assessment in Geography, Grades 9-12.
http://my.ilstu.edu/~jabraun/socialstudies/assess/geo/assess/9-12-13.html

Level: Grades 9-12

Specification 1:

Content: Cooperation between countries and organizations may have lasting influences on past, present, and future global issues (*Geography for Life,* Standard 13, Theme 3, Grade 12).
Cognition: Explain a geographic principle or phenomenon.
Performance expectation: Explain how cooperation between countries and organizations may have lasting influences on past, present, and future global issues.

Specification 2:

Content: Cooperation between countries and organizations may have lasting influences on past, present, and future global issues (*Geography for Life,* Standard 13, Theme 3, Grade 12).
Cognition: Evaluate one or more answers to a question or solutions to a problem using geographic principles.
Performance expectation: Evaluate one or more answers to how cooperation between countries and organizations may have lasting influences on past, present, and future global issues.

Specification 3:

Content: Cooperation between countries and organizations may have lasting influences on past, present, and future global issues (*Geography for Life,* Standard 13, Theme 3, Grade 12).
Cognition: Communicate using geographic principles, models, and data to educate or persuade an audience.
Performance expectation: Persuade or inform an audience about how cooperation between countries and organizations may have lasting influences on past, present, and future global issues.

Item characteristic: The response takes the form of a constructed response/essay.

Item 2 (Figure 5) is an example of an item that could be used to assess these three performance expectations in a constructed response format. It asks students to use their understanding of human cooperation to evaluate the long-term impacts of two different approaches to cooperation for the United States, and to communicate their answers in the form of an argument made to a colleague.

Example 3. The third example item also is designed for high school students. This one focuses on physical geography content—how the amount of incoming solar energy varies across the Earth's surface—and on the cognitive practices of organizing geographic information.

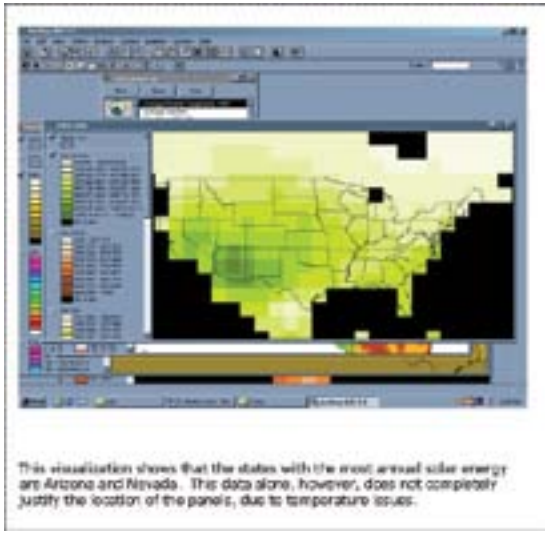
Content: Different regions receive different amounts of solar energy during a year, and the amount of solar energy a place receives depends on the latitude and cloud cover at that place.
Cognition: Sort data by spatial and other characteristics to answer a geographic question.
Performance expectation: Sort data by spatial and other characteristics to answer a question about the distribution of solar energy across a region.
Item characteristic: The response takes the form of a constructed response/representation.

Item 3 (Figure 6) shows how this performance expectation might be evaluated in a computer-based assessment in which students use geographic information systems (GIS) to organize geographic data in a way that prepares the student to answer a geographic question using the data.

Figure 6. Item 3: Solar Power Task

Item 3
In this task, students are asked to identify locations appropriate for solar power generation in the United States. One step students must take to complete this task involves sorting solar energy data for regions across the United States in such a way to help them complete the task. They must use a GIS program to organize the data to show which states will be able to generate the maximum amount of electricity over a year from their solar panels.

A screen shot showing an example of a student response:



Source: Adapted from Quellmalz et al. (2004).

Implications for Improving Geography Teaching and Learning

Efforts to improve geography education can be enhanced by attention to the design and use of assessments. Assessments for the purposes of informing instruction, measuring achievement, evaluating programs, and conducting research all can play important roles in improving teaching and learning of geography. However, assessments must be carefully designed to align with the learning goals, and they must be designed to be sensitive to the desired information and the context in which the information will be used. An assessment designed to help a teacher plan instruction for next week will likely look very different from an assessment created to evaluate the effectiveness of a newly introduced instructional program.

The design of assessments for geography should pay careful attention to each of the considerations described in this chapter. The characteristics of assessments, item quality, and cost-effectiveness must all be chosen with the assessment goals and purpose in mind to collect the desired information to assist geography teaching and learning.

Because the challenges of assessment design are considerable, the Committee has primarily emphasized developing assessment frameworks in our work. Assessment frameworks can make the task of creating valid, reliable, and useful assessments much more manageable, particularly when the target objectives are complex, such as reasoning and problem-solving skills.

The Committee believes the creation of carefully designed assessment frameworks for geography education could play an important role in improving geography teaching and learning. We argue that a set of assessment frameworks tailored to the goals of geography education would enable educators to develop assessments that would inform instruction, provide insight into program effectiveness, and build a research base that will ultimately lead to substantial advancement of those goals. If successful, these assessment frameworks will improve educators' abilities to create and use assessments that: (1) communicate the objectives of geography education to teachers, students, and other stakeholders; and (2) yield the information that educators and students need to make suitable decisions about teaching and learning.

In later chapters, we return to the challenges of designing assessments and the role of assessment frameworks, but first we turn to a discussion of the current state of assessment in geography education.

Chapter 3: The Current Context for K–12 Geography Assessment

In this chapter, we consider the context in which this report is responding: the current state of assessment in geography education. As explained in the previous chapter, well-designed assessments offer a mechanism for improving the effectiveness of education at scales from individual students and teachers to across districts and nationwide. Because assessment frameworks ease the challenge of designing effective assessments, the Committee has made the creation of assessment frameworks for geography education the focus of our work. As part of this work, we found it necessary to investigate two questions about the current practice of assessment in geography education: What is the nature of existing assessment frameworks for K–12 geography education? And, what is the nature of existing geography assessments? In this chapter, we present the results of these investigations, first with an overview of currently used assessment frameworks. Then we provide the results of a study we conducted of existing assessment practices in K–12 geography. We conclude the chapter with a discussion of the needs and opportunities revealed by this examination of recent history and current practices.

Overview of Current Geography Frameworks

As discussed in Chapter 2, assessment can serve as a means for improving geographic education. Our goal is to compare the actual practices of assessment with the goals for geographic education as laid out in the second edition of *Geography for Life: National Geography Standards* (Heffron & Downs, 2012). We turn now to a review of the most influential frameworks for assessment in geography, and discuss their relationship to the goals of geographic education established by *Geography for Life*. This comparison allows us to evaluate how well existing assessment frameworks support progress toward those goals, and to identify opportunities that are not currently being met by existing frameworks.

Frameworks for Large-Scale Assessments

The three assessment frameworks that are most relevant for K–12 geography in the United States are the National Assessment of Educational Progress (NAEP) assessment framework for geography, the Advanced Placement (AP) Human Geography framework, and the NAEP assessment framework for science. Each is the product of thoughtful efforts to articulate educational goals and each has been shaped by increasing attention to measurement of learning outcomes and accountability over that same period.

National Assessment of Educational Progress in Geography (1994, 2001, 2010). In the 1990s, Congress authorized the development of the broad-based NAEP geography assessment at grades 4, 8, and 12. To support the development of this assessment, a committee of professional geographers, educators, and administrators worked to create a framework that provided operational specifications for the geography assessments. This framework was heavily influenced by the *Guidelines for Geographic Education* (Joint Committee on Geographic Education, 1984), though the five themes were reduced to three categories of content. The NAEP geography framework is intended to cover the breadth of geography knowledge and skills that would result from good geographic instruction in the classroom.

The NAEP geography framework is organized along two dimensions (Table 4): a content dimension and a cognitive dimension. The content dimension is divided into three main areas: Space and Place, Environment and Society, and Spatial Dynamics and Connections. Each of these content categories is explored in detail within the framework, with specific objectives for each of the three grade levels. The emphasis on content is evident, and the relative lack of elaboration of the cognitive dimension is equally evident.

The cognitive dimension of the geography assessment is composed of three levels: Knowing, Understanding, and Applying (Table 4). Knowing (i.e., *What is it? Where is it?*) assesses students on their ability to answer questions by recall. Understanding (i.e., *Why is it there? How did it get there? What is its significance?*) assesses students' ability to provide explanations for phenomena. This cognitive level requires students to identify and explain geographic patterns and processes. Applying (i.e., *How can knowledge and understanding be used to solve geographic problems?*) requires a range of higher-order thinking skills. At this cognitive level, students are expected to hypothesize, apply geographic principles to new contexts, and form problem-solving models.

Together, the content and cognitive dimensions of the assessment form a matrix (Table 4) in which each con-

tent area is measured at each cognitive level. The cells of the matrix describe example performances that characterize the intersection of the two dimensions, and assessment developers are given information about how much each cell should be represented on an assessment (Table 5). The framework proceeds to "drill down" to delineate in some detail each of the broad content areas involved. However, it does not have a comparable clarification of expectations for the cognitive (practices) dimension.

Assessment development is guided by three Achievement Levels (ALs) that are used to evaluate performance at each grade level: Basic, Proficient, and Advanced. Basic indicates partial mastery of requisite knowledge and skills, Proficient represents solid academic performance for each grade assessed, and Advanced indicates superior performance.

Advanced Placement Human Geography (2000). Following the development of NAEP Geography, advocacy by the geography education community and the Road Map Project partner organizations led the College Board to develop an AP course on Human Geography. The Advanced Placement Human Geography course (APHG) is not patterned on previous standards such as the *Guidelines or Geography for Life,* but rather reflects a consensus of college faculty about the nature of modern human geography as taught in colleges in the United States. Unlike the NAEP geography assessment, the APHG exam covers only a portion of geography—human geography.

The two-part exam consists of a 60-minute multiple-choice section and a 75-minute open-response section. The first part focuses primarily on geography content, but the second part is intended to assess students' ability to synthesize their geographic knowledge using geographic skills and practices.

There is no traditional assessment framework for APHG available to the public. Instead, the teachers and students are provided with a detailed course description. While it takes the form of a course outline, rather than an explicit assessment framework, the description of the APHG course plays the role of an assessment framework in communicating to teachers and students what students are expected to master.

The APHG course outline divides content into seven sections:

1. Geography: Its Nature and Perspectives,
2. Population,
3. Cultural Patterns and Processes,
4. Political Organization of Space,
5. Agriculture and Rural Land Use,
6. Industrialization and Economic Development, and
7. Cities and Urban Land Use.

In the course outline, each of these sections is broken down into several subsections, each of which lists the topics to be covered in that subsection and the percentage of the AP exam that will focus on that topic.

The course outline also describes the following target practices and skills: use and think about maps and spatial data, understand and interpret the implications of associations among phenomena in places, recognize and interpret at different scales the relationships among patterns and processes, define regions and evaluate the regionalization process, and characterize and analyze changing interconnections among places.

The introduction of AP Human Geography has had a dramatic impact on the teaching of geography in American high schools. After several years of development, the course and test were launched in 2000-2001 with more than 3,200 tests administered. Ten years later, more than 83,000 students took the 2011 APHG exam. While the college-level exam questions cannot be directly used to assess students' ability to do high school geography, an increasing number of high school teachers model their course assessments on the constructed response and multiple-choice questions published by the College Board.

NAEP Science (2009). In 2008, the National Assessment Governing Board (NAGB) released a major revision to the NAEP Science framework titled *A Science Framework for the 2009 National Assessment of Educational Progress* (NAGB, 2008; Fu, Raizen, & Shavelson, 2009). The framework consists of a matrix (Table 3) with science practices on one dimension: identifying science principles, using science principles, using scientific inquiry, and using technological design. To provide more details, each practice is expanded into four general performance expectations. For example, one of the performance expectations for "Using scientific inquiry" is to "conduct scientific investigations using appropriate tools and techniques." On the other dimension are the three components of science content: physical science content statements, life science content statements, and Earth and space sciences content statements, and each of these areas of content is further elaborated into learning objectives for elementary, middle, and high school students.

The NAEP Science framework also includes examples of how a practice could be applied to a specific scientific

context to create performance expectations that integrate content and practices, as well as example items that demonstrate how the performance expectation could be probed. The description of the science practices and the examples of how they could be used to assess scientific reasoning place far greater emphasis on integrating scientific practices into NAEP Science assessments than the NAEP Geography framework does currently.

The framework's focus on practices is evident in the 2009 NAEP Science assessment, which includes a new set of hands-on interactive computer assessments to assess students' ability to perform the practices. These computer tasks provided students with a simulated environment to apply their scientific skills and practices, giving them a more open-ended environment than students had been provided in previous NAEP assessments. The rich, interactive environment allows students to test hypotheses and design and execute experiments, providing insight into students' ability to reason with scientific evidence. In practice, these items have revealed information about students that traditional content-focused multiple-choice and short-answer NAEP assessments had not. In a special report on the results of these hands-on interactive assessments, the National Center for Education Statistics (2012) showed that just over one-half of high school students reached the correct conclusions for the scientific investigations on the exam, but fewer than 30 percent of those students could justify their conclusions satisfactorily.

Current Geography Assessment Practices

To obtain a picture of current assessment practices in K–12 geography education, the Committee commissioned a study of existing assessments, both large-scale and classroom-level. The study was conducted by the Committee research director under the direction of the Committee co-chairs, and with assistance from Committee members.

The goal of the study was to shed light on which components of geography are being assessed currently in grades K–12 and how those components are being assessed. To conduct this study, we collected publicly released items from large-scale assessments and classroom assessments from commercially published and widely distributed non-commercial instructional materials. Our specific objectives were to: (1) compare what is being assessed with the goals for geography education described in *Geography for Life*, and (2) compare how geography is being assessed against the criteria for assessment discussed in the previous chapter. This analysis of current assessments provides empirical data for the recommendations we make in the chapters that follow.

Overview of Methods[5]

The study is based on an analysis of items from resources that were selected at random from our entire resource collection. The full resource collection included assessments that were labeled as geography assessments, as well as assessments from other areas of science and social studies that were judged to be likely to include items assessing geographic knowledge and practices. The resource collection included both large-scale and classroom-level assessments. The large-scale assessments included state, national, and international assessments. Classroom assessments were drawn from textbooks, geography units, and online item banks for teachers.

Once resources were selected from the collection, those that did not have items that test geographic knowledge aligned to any of the six essential elements from *Geography for Life* or to the geographic practices outlined in Chapter 1 were excluded from further analysis. When large-scale assessments were excluded, it was because they were general science or social studies assessments that did not contain any items with geographic content. When classroom resources were excluded, it usually was because they did not contain any assessment items at all. Out of 114 randomly selected resources in the collection, 79 of the resources were determined to have items that were testing geographic competencies, and these resources yielded 696 geography items for analysis.

The study categorized items according to: (1) *targeted ability*—what is being assessed, (2) *item characteristics*—how the item assesses student competencies, and (3) *confounding factors*—whether the item is likely to accurately reflect what students know. These three criteria were used as the basis for a taxonomy (Table 6) that was used to code each item.

Targeted ability describes the substance of what the item assesses. It includes the geographic content, geographic skills, and the type of cognition that the item requires. These criteria provide the core information about which competencies are being assessed. Content is classified according to the standards from *Geography for Life*, geographic skills are classified using categories adapted from *Geography for Life* by the Committee, and cognitive demand is classified according to the categories described by Li, Ruiz-Primo, & Shavelson (2006).

[5] See Appendix B for additional details about the methods for the study.

Table 6. Geography Item Coding Taxonomy

Coding categories	Subcategories
Targeted ability	1. Geographic content 2. Geographic skills and practices 3. Cognitive demand
Item characteristics	4. Assessment setting 5. Grade band 6. Item format 7. Type of representation
Confounding factors	8. Clarity 9. Content accuracy 10. Vulnerability to test-taking strategies

The *item characteristics* category captures the way in which the assessment task was presented to the students. Under item characteristics, we recorded information about the setting in which the assessment was used, the targeted grade level, how the item is structured, and what type of graphical representations appear in the stem or answer choices, if any. Coding for these characteristics enabled us to look for patterns in how goals are being assessed.

The *confounding factors* category includes problems with the way the item is written that make it an inaccurate (or less accurate) measure of what students know and can do. The confounding factors for which we collected information include whether the way the item is written could make it difficult for students to understand what was being asked and whether the language was unnecessarily complex. We also recorded whether there are ways in which the substance of the item was inaccurate, including if additional answer choices could be considered correct. Finally, we described ways that students could use non-geographic knowledge to evaluate the item, such as using logic to eliminate unlikely answer choices.

Findings

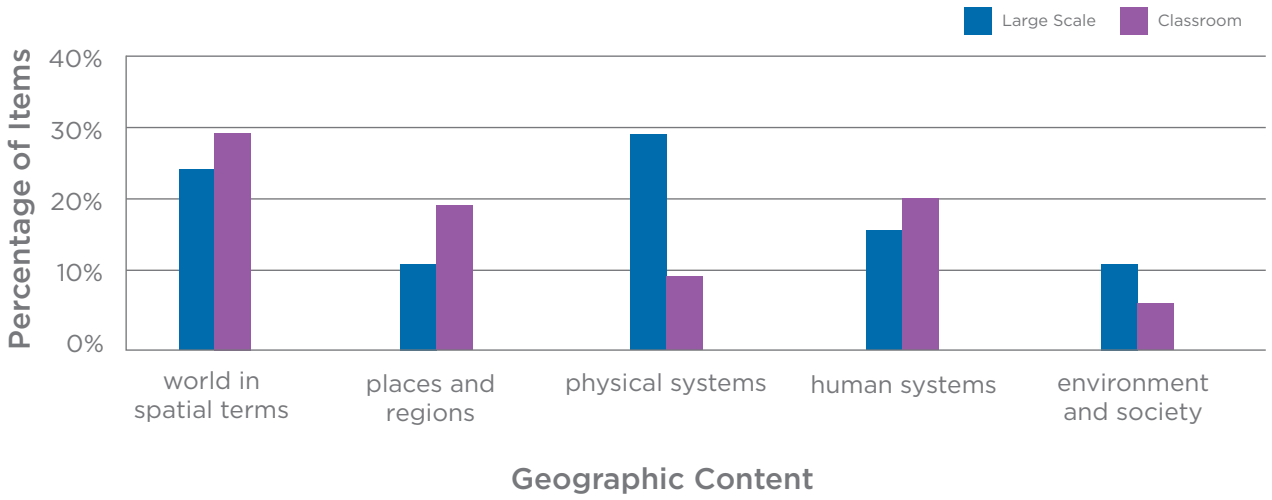
Geography Content Coverage

Our first analysis examined the distribution of content covered across the items. Our objective was to see if assessment items are aligned to the content in *Geography for Life.*

We found items assessing all five of the content-focused[6] essential elements in *Geography for Life.* This distribution confirms that at the coarsest level of analysis, current assessment practices are supportive of the goals for content coverage expressed by *Geography for Life.*

When we examined the relative frequency of specific essential elements, however, a pattern of uneven distribution emerged (Figure 7). Across all large-scale and classroom items, two categories, *Environment and Society* and *Places and Regions,* were assessed by substantially fewer items than the others. When we looked only at assessments that were explicitly labeled as geography assessments (excluding science and other social studies assessments), we observed that the three middle categories (see Figure 7) are assessed more evenly, but *The World in Spatial Terms* is even more heavily emphasized relative to the others, and *Environment and Society* slightly less.

Figure 7. Frequency Distribution of Assessment Items in Large-Scale Assessments and Classroom Assessments for the Content-Focused Essential Elements from Geography for Life



[6] There are six essential elements in Geography for Life, but the sixth is explicitly about the application of geographic knowledge, so we do not consider it to be content-focused.

When we compared the content distribution in large-scale assessments with classroom assessments, different patterns of emphasis emerged (Figure 7). Most notably, items aligned to *Physical Systems* are overrepresented in large-scale tests and underrepresented in classroom assessments. In addition, while *Environment and Society* is underrepresented overall, that content area is assessed even less frequently in classroom assessments than in large-scale tests.

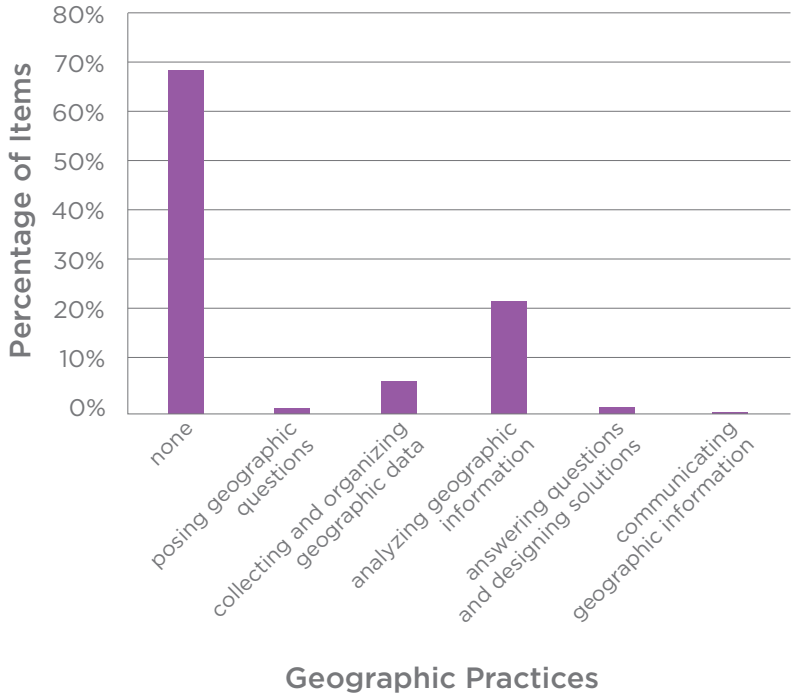
The large number of assessments targeting *Physical Systems* in large-scale tests is likely a result of that category's overlap with earth science, a topic that more frequently appears in large-scale tests compared with other geography content. However, this content area rarely appears in geography assessments at the classroom level. Moreover, we find the underrepresentation of items aligned to *Environment and Society* in classroom assessments worrisome, given the mounting importance for students to understand humans' interactions with their environment. It is possible that this topic is assessed infrequently because it spans the boundary between social studies and science, and demands competence with both disciplines. Rather than being addressed in complementary ways by each discipline, it seems that the category is being overlooked by both.

When we considered coverage of specific standards within the sixteen essential elements, we found the distribution to be even more uneven. We found that 40% of all items evaluated knowledge from only three standards. These three standards focus on the properties of maps, Earth's physical processes, and the characteristics of specific places on Earth. Certainly, not all standards demand the same amount of attention, but this uneven distribution is cause for concern if it signals to teachers that some standards are unimportant, or if it indicates that educators are not receiving information they should have about the status of their students' learning in certain areas.

Geographic Skills and Practices[7]

As discussed in Chapter 1, since the 1960s, major efforts to communicate the goals of geographic education have argued for the importance of geographic skills and practices, but these arguments have had limited impact on the classroom. Not surprisingly then, our study reveals that only 30% of geography assessment items required that students use any geographic practices at all (Figure 8). A clear pattern emerged from this analysis, showing that most of those items assess the same geographic practice, *analyzing geographic information. Analyzing geographic information* is assessed in 21% of large-scale items, but the other geographic practices rarely appear in any assessment items at all. One might expect that items asking students to apply their knowledge to a geographic question or problem might be reserved for high school assessments, but we find that absence of geographic practices from assessments to be consistent across grade levels. In fact, middle school items have the highest percentage of items that require a geographic practice (40%). The heavy emphasis on analyzing information may reflect the fact that it is regarded as the easiest practice to assess in a multiple-choice or short-answer format. For example, the most common type of analysis in this item involved students observing a graphic, such as a photograph or a map, and characterizing the pattern that they see. See Figure 9 for an example of an item that probes students' ability to look for patterns by comparing settlements and features of the physical

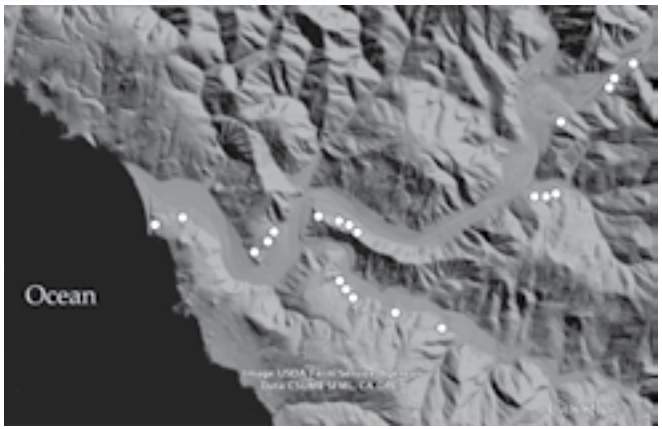
Figure 8. Frequency Distribution of Large-scale Geography Assessment Items that Target Each Geographic Practice



[7] In the analysis for geographic skills and practices and all subsequent analyses, we report only data for large-scale assessments because of the challenges of conducting random sampling with classroom materials. See Appendix A for more information.

Figure 9. An Example of Middle School Social Studies Item that Has a Representation (Map) and Probes a Geographic Practice (Analysis)

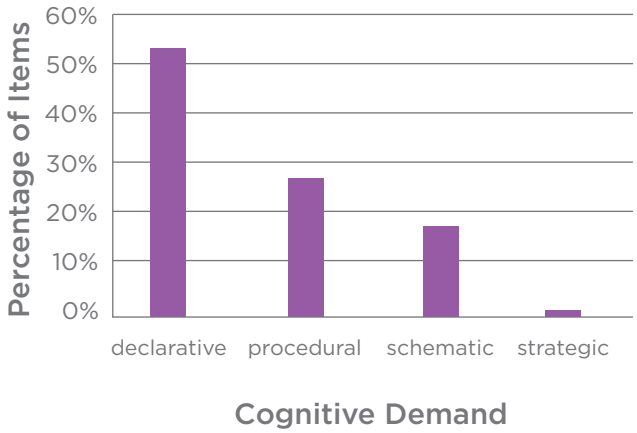
The map below shows a coastal region with a river winding through the mountains and meeting the coastline with a delta. Early agricultural sites found in this region are indicated with white dots.



According to this map, most early agricultural sites in this region developed near the

A. mountains.
B. rivers.
C. coastlines.
D. deltas.

Figure 10. Frequency Distribution of the Cognitive Demand Required for Large-Scale Geography Items



environment. Other practices, such as data collection, perhaps seem more difficult to evaluate in a format that easy to score.

Our findings indicate that geographic practices are not being assessed in a way that conveys their importance. As a result, current assessments are providing an incomplete picture of students' proficiency with geography.

Cognitive Demand

To get a sense of the depth of conceptual understanding being assessed, we analyzed items for the cognitive demands required to answer the questions (Figure 10; see Table A4 for a description of cognitive demands). We found that almost 54% of the items tapped students' declarative knowledge (i.e., *knowing that*), often at the level of recognizing a definition. More than 27% assessed students' procedural knowledge (i.e., *knowing how*), which includes reading and gathering in-formation from maps, graphs, and texts. Only 17% of geography items require schematic knowledge (i.e., *knowing why*), which includes explaining an unfamiliar context by drawing on general geographic principles or models. Virtually none of the items directly tapped strategic knowledge (i.e., *knowing how, when, and where knowledge applies*). This last finding was anticipated, because an item requiring schematic knowledge often would require probing competencies in multiple content areas and practices, which makes assessment design much more complex. Both factual knowledge and the ability to read maps and graphs are important elements of geography literacy. The minimal demand for higher-order thinking and reasoning with geographic concepts and the overall imbalanced pattern of assessment across cognitive demands, however, requires increased attention to support assessing knowledge and skills that better reflect the reasoning we want students to be prepared to do with their geographic knowledge.

Item Characteristics

The dominant item format used in both large-scale and classroom geography items is selected response, with 68% of all items using some form of multiple-choice format (Figure 11). Classroom assessments offer more diverse opportunities for students to represent their knowledge; 47% of classroom items use short answer or essay formats, though few classroom materials include scoring rubrics, which limits their utility (and assessment developers would not consider open-response items questions without a rubric). Fewer than 3% of all items ask students to make or modify a representation, a format that holds promising opportunities for evaluating students' geographic reasoning.

We also looked at the use of representations that students must refer to in order to answer a question. Representations appeared in 59% of the items, with maps being the most common (Figure 12). However,

we found that only 16% of the questions with maps required students to process the information on the map, as does the item in Figure 9. On the other hand, 84% asked students to read information directly from the map, indicating that map reading is well represented in geography assessments, but reasoning with geospatial representations is not.

Confounding Factors

We found that 60% of the items displayed apparent problems that could impede students' ability to reflect their knowledge. One of the most common problems found in our analysis, particularly for large-scale tests, was ambiguous questions and/or answers (14%). Even more common is that at least one answer choice is implausible (15%). The concern with any confounding factors such as these, is that they raise doubts as to whether the results of those geography assessments are valid indicators of what students know or are able to do.

Implications for Improving Geography Education

Taken together, the discussions in this chapter present a compelling picture of the needs and opportunities for assessment in geography education.

Needs

Our review of current assessment practices exposes weaknesses in the way geography is being assessed across the curriculum. These weaknesses include:

- **Coverage of content areas is uneven.** Most content standards are underrepresented in assessments. This trend signals to educators and learners that this

Figure 11. Frequency Distribution of Item Formats for Classroom and Large Scale Geography Items

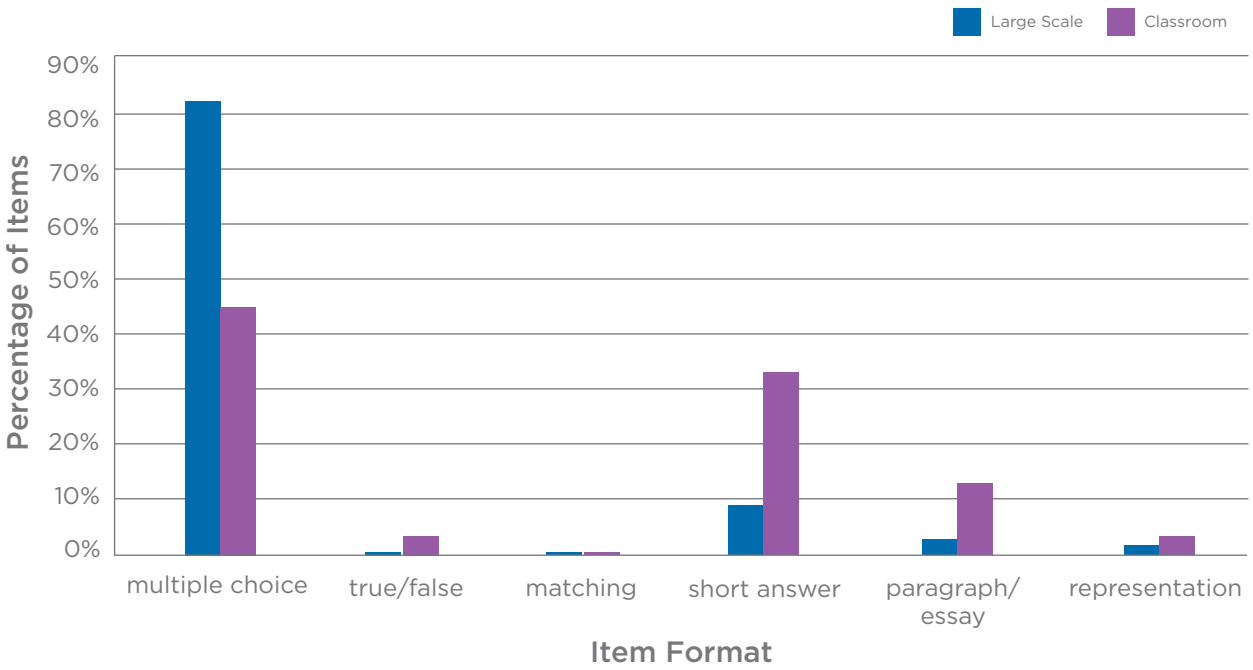
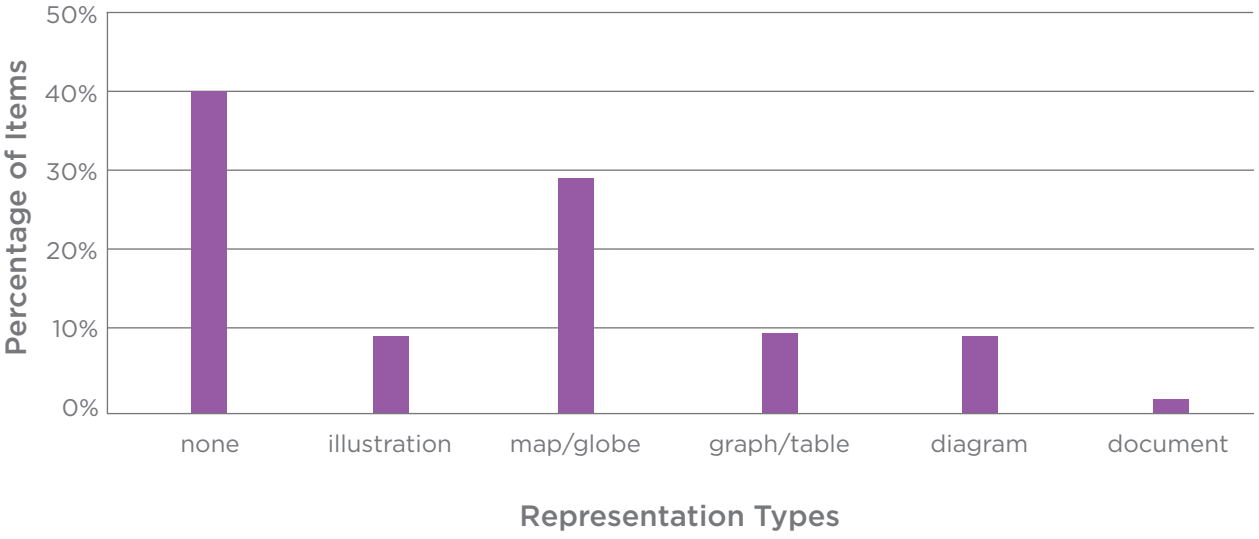


Figure 12. Frequency Distribution of Types of Representations that Appear as Part of a Large Scale Geography Assessment Item



content is less important and it indicates that educators are not being provided with information about student learning in these areas.

- **Assessments tend to focus on less-demanding cognitive tasks.** The overwhelming majority of assessment items in our survey focus on recalling or recognizing geographic facts and simple procedural knowledge, such as using a map legend to retrieve information from the map. Students rarely are asked to apply their geographic knowledge to a novel context, explain or justify their reasoning, or evaluate the strength of an argument. The result is that these assessments reinforce a conception of geography as being about facts and simple recall; they also fail to provide educators with an accurate picture of students' ability to reason, analyze, or communicate about questions and problems with their geographic knowledge.
- **Technical quality can be improved.** A substantial number of items in our survey have structural or content flaws that threaten their validity as measures of student understanding and ability. If an item is difficult to understand, has inaccurate content, or allows students to narrow down answer choices through test-taking strategies, then it does not provide the information about students' understanding and abilities that educators need to make good instructional decisions.

We believe that addressing these weaknesses is an essential step toward improving geography education overall, and our recommendations in the chapters that follow are intended to help educators to do so.

Opportunities for Improvement

During the course of nearly 50 years, geographers and educators have produced increasingly explicit descriptions of the goals of geography education, culminating in the second edition of *Geography for Life*. This vision includes expectations for students by grade band and example performances for knowledge and understanding, perspectives, and skills. While educators and curriculum developers have adopted the general goals for knowledge and understanding, it appears that they have not adopted the goals for skills and practices to the same degree. With the exception of the 2009 NAEP science framework, our review of existing frameworks for large-scale assessment indicates that these assessments are acting to maintain the existing focus on content, rather than signaling the importance of applying content knowledge to exploring geographic questions, problems, and phenomena. This bias toward the lower-level *knowing and understanding* portion of cognition is clearly reflected in the assessments that are being used to evaluate students' geographic literacy.

Geography for Life describes a vision of geographic education in which students use spatial analytical skills to reason about people, places, and connections among them, but we find that assessments present a much narrower view of expectations for geographic literacy. Therefore, we see an enormous opportunity to use assessment as a mechanism to shift geography education closer to the version of geographic literacy captured by *Geography for Life*. Assessments that increase attention to the roles of geographic reasoning and problem solving, that draw on the full range of geographic concepts, and that challenge students to manipulate their knowledge with higher cognitive demands could provide a mechanism to communicate these priorities to teachers and students, and such assessments would provide tools to help educators monitor learning. In the chapters that follow, we describe an approach to assessment that uses a framework organized around these priorities, specifically designed to capitalize on these opportunities for improvement.

Chapter 4: A 21st Century Assessment Framework for the Geographical Sciences

Chapter 3 revealed that existing geography assessments do not reflect the priorities established by the *Geography for Life: National Geography Standards* (Heffron & Downs, 2012). The same three (out of 16) content standards are assessed in 40% of the items, competency with geographic practices is rarely assessed, and structural problems with items are common. Here, we present an assessment framework that we believe will help assessment developers create assessments that support the overall goals of geography education by providing a structure to define, organize, and systematically select the assessment goals. We call this framework the *21st Century Assessment Framework for the Geographical Sciences* (AFGS21).

Our goal is that this assessment framework serves as a blueprint to support the development of a new generation of assessment frameworks, and they will generate assessments with higher fidelity to the goals for geographic literacy and the principles of assessment development. In this chapter, we describe the process of using the AFGS21 to create specific frameworks.

A 21st Century Assessment Framework for the Geographical Sciences

AFGS21, like most contemporary assessment frameworks, is based on a 2-dimensional matrix with a "content" dimension and a "cognitive" dimension (Figure 13). The content dimension is made up of geographic concepts and principles, and the cognitive dimension is made up of knowing, understanding, and geographic practices. The cells in the body of the matrix are designed to hold performance expectations, which describe the application of a cognitive behavior to a geographic concept or principle.

AFGS21 is designed to support assessment of critical knowledge and practices for geography. Specifically, it is designed to:

1. support the development of high-quality assessments that are matched to the goals of the unit, course, standard or program being assessed; and
2. facilitate the development of assessments that demand integrated knowledge, understanding, skills, and practices.

AFGS21 has been designed as a *general* assessment framework, meaning that it is not intended to be used directly to generate *specific* assessments. Rather, it is designed to serve as a blueprint that developers can follow for developing specific assessment frameworks for specific purposes. By using AFGS21 as a guide, developers of assessments can achieve the objectives that AFGS21 was designed to achieve, applied to their assessment context.

A "specific assessment framework" is used by an assessment developer to create instruments tailored to the particular set of goals for the population being assessed (see Chapter 5 for an example of a specific assessment framework). The process of using the framework to create assessments does not change, but the contents of each dimension and, therefore, the constructs developed from them, will vary. The framework dimensions are specified based on the goals for the population that will be assessed. The goals will change in scope, detail, and sophistication. For example, an elementary school geography course that focuses specifically on human geography might assess only content within Human Systems, and would eliminate the other content areas from the framework. Similarly, this course might target only a subset of the geographic practices, so the practices that students would not be expected to use would be eliminated from the framework. Because of its concentration on one content area, the course might require far more detailed understanding of that content than a broader course, so there could be more content and cognition goals specified under Human Systems than included in a framework for another course. In addition, the description of expectations for competency for each of the content and cognition goals would be much less sophisticated than in a framework of the same scope for a comparable high school course.

The Content Dimension in AFGS21

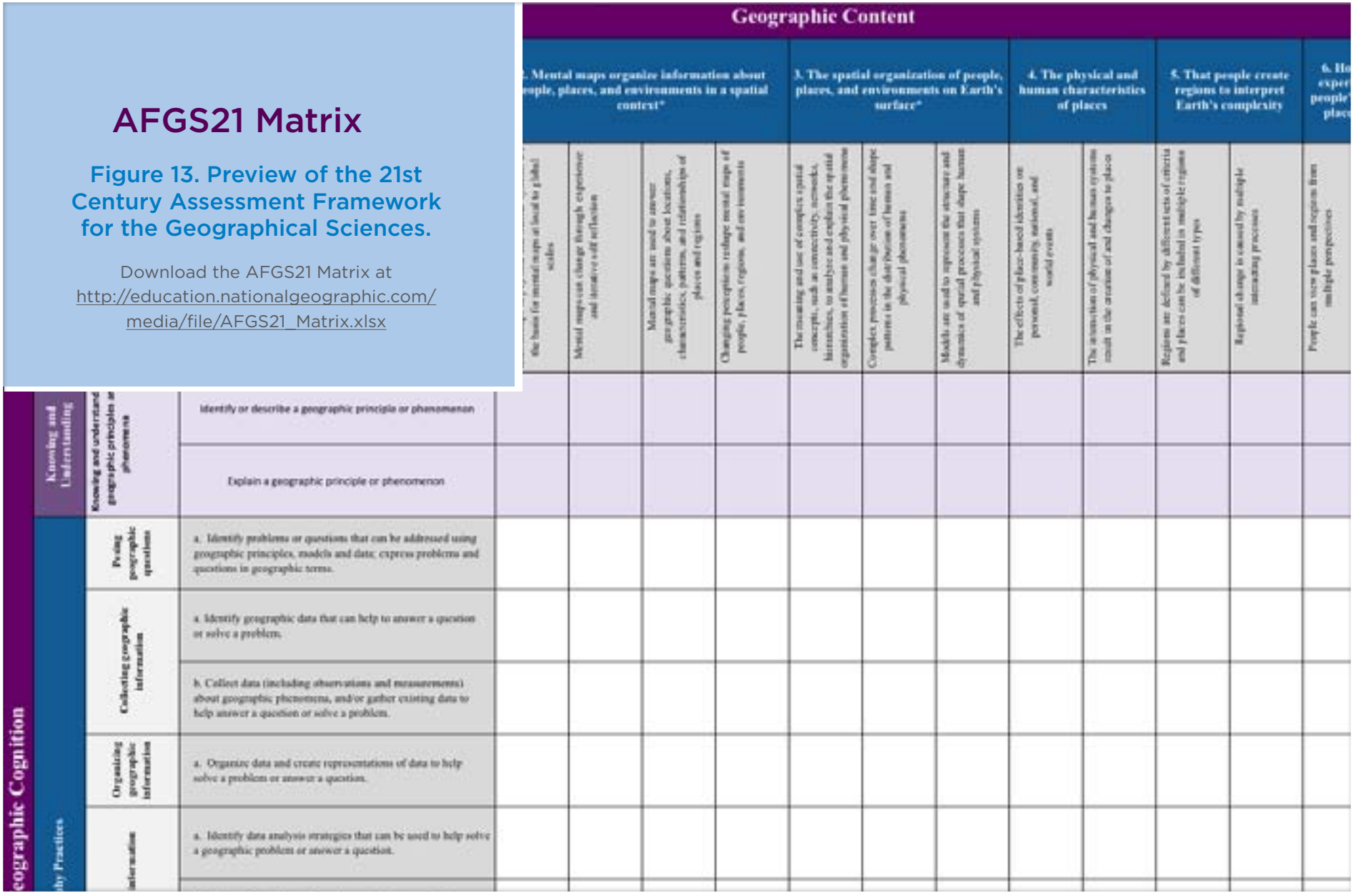
The content dimension of AFGS21 enumerates and describes the geography concepts and principles to be assessed. In AFGS21, the content dimension is segmented into 16 categories corresponding to 16 of the 18 standards from *Geography for Life*.[8] Within

[8] Of the 18 standards, two (Standards 17 and 18) focus on activities. They have been incorporated into the practices portion of the framework's cognitive dimension.

AFGS21 Matrix

Figure 13. Preview of the 21st Century Assessment Framework for the Geographical Sciences.

Download the AFGS21 Matrix at http://education.nationalgeographic.com/media/file/AFGS21_Matrix.xlsx



each of these categories, there are subcategories for the specific content statements that *Geography for Life* provides for each of the standards. Depending on the standard, there are between one and four subcategories per standard.

Because the framework describes the culmination of the K–12 curriculum, we used the content statements for grades 9-12 to define the content subcategories, but one could construct versions of AFGS21 using elementary school or middle school content statements from *Geography for Life*.

The Cognitive Dimension in AFGS21

The cognitive dimension of AFGS21 describes expectations for performing geographic thinking and reasoning. AFGS21 divides cognition into two major categories:

1. knowing and understanding geographic principles and phenomena, and
2. geography practices.

Knowing and Understanding in AFGS21

The first category in AFGS21's cognitive dimension is "knowing and understanding." From an assessment perspective, knowing and understanding corresponds to the ability to identify, describe, and explain facts and concepts. The ultimate goal of geography education is to cultivate the ability to perform complex reasoning, but knowing geographic facts and understanding geographic concepts is critical for that reasoning. This means that assessments must probe knowing and understanding to accurately characterize students' progress toward geographic literacy.

The Categories of Geographic Practices in AFGS21

While it is important for students to master the foundational facts and concepts of geography, the geographic advantage comes from the ability to apply this understanding to answer meaningful questions and solve meaningful problems. The central goal of 21st century geography education is to prepare students to be able to reason through questions and problems that they will encounter in their personal, professional, and civic lives.

The geography practices in AFGS21 describe the kinds of activities students should be pre-pared to perform in order to reason through geographic questions and problems. We divide these practices into six categories, each of which encompasses a set of more specific practices (Table 7).

While these practices are presented as a linear sequence, this should not be interpreted as indicating that they necessarily proceed in this order in practice, or that they should. A geography investigation might start with any one of these practices, and it may even require only one of them. Alternatively, it might move iteratively back and forth between one or more steps. As Figure 14 indicates, the requirements of the specific question or problem and its context should determine which practices are necessary and in what order they should be performed.

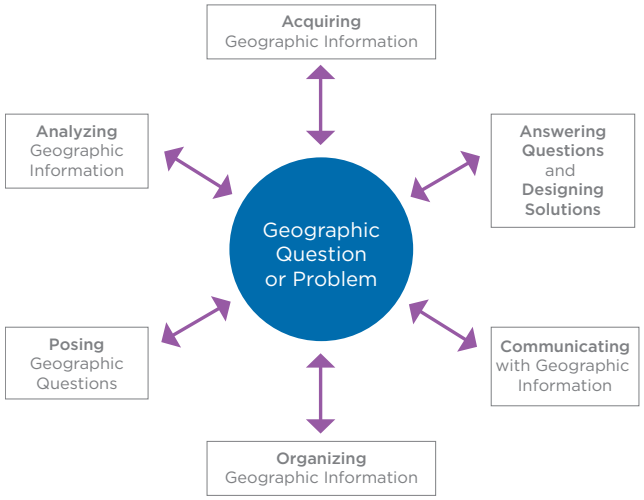
Table 7. The Six Categories of Geographic Practices in AFGS21 and the Geographic Practices That Fall Within Those Categories

Practice Category	Geographic Practices
1. **Posing geographic questions**	a. Identify problems or questions that can be addressed using geographic principles, models, and data; express problems and questions in geographic terms.
2. **Acquiring geographic information**	a. Identify geographic data that can help to answer a question or solve a problem.
	b. Collect data (incl. observations or measurements) about geographic phenomena, and/or gather existing data to help answer a question or solve a problem.
3. **Organizing geographic information**	a. Organize data and create representations of data to help solve a problem or answer a question.
4. **Analyzing geographic information**	a. Identify data analysis strategies that can be used to help solve a geographic problem or answer a question.
	b. Find and describe spatial and temporal patterns in data, or find data that matches a pattern, to help answer a question or solve a problem.
	c. Construct an explanation or prediction for phenomena by comparing data to a principle or model.
5. **Answering questions and designing solutions**	a. Construct an answer to a question or a solution to a problem using geographic principles, models, and data.
	b. Evaluate one or more answers to a question or solutions to a problem using geographic principles, models, and data.
6. **Communicating with geographic information**	a. Communicate using geographic data, principles, and models to educate or persuade an audience.

We also note that the term *data* in the descriptions of the specific practices includes both qualitative and quantitative data.

Five categories of geographic practices were derived from the geographic skills described in *Geography for Life*, with one addition, *communicating with geographic information*. This addition makes the framework consistent with the categories of scientific practices described in the National Assessment of Educational Progress (NAEP) 2009 Science framework and the NRC (2012) *Framework for Science Education*. The categories of practices speak directly to the nature of geography, but they also have a clear relationship to the model of inquiry that underlies recent assessment and standards frameworks for science. The relationship among the categories of practices in AFGS21 and the practices as they have been described in recent science and geography education standards and framework documents is illustrated in Table 8.

Figure 14. A Graphical Representation of the Relationship Between the Geographic Practice Categories and the Geographic Question or Problem Driving the Inquiry Process



The Geographic Practices in AFGS21

The specific practices in AFGS21 describe the ways that geographic knowledge and understanding can be applied to answer questions and solve problems. While these practices have analogues across the natural and social sciences and engineering, they are distinctively geographic and, as such, contribute to the geographic advantage. For example:

- **Posing and answering geographic questions.** The data that geographers work with often overlap with those from other fields; for example, historians also consider people and places, economists look at patterns within and among human settlements, and ecologists study living and physical systems. Geographic questions differ from other types of questions in that they may look at the same data, but from a spatial perspective. The posing of a geographic question might involve looking at the same data the economist is looking at, but the spatial focus might inspire a question that leads the geographic investigation to incorporate different variables, analyses, and ultimately, to contribute different information to understanding a phenomenon.
- **Collecting, organizing, and creating visualizations of geographic data.** As with any scientific discipline, geography is focused on understanding phenomena. In geography, these are phenomena at or near the surface of the Earth, and involve humans and/or their environment. Geographic data are integral to geographic inquiry because they provide information about the current or historical state of the world that can be used to understand people, places, and the connections between them. A key aspect of geographic data is their association with a specific location; for data to be geographic, they must have a spatial component. Given the association with a location, geographic patterns are particularly visual when displayed across a map, and these visualizations are useful for analyses as well as for communication.
- **Applying geographic theories and models to data.** The use of geographic principles models, theories, and data might be characterized as the "geographic perspective." Geographic principles, theories, and models are components of geographic content that, when appropriately applied to data, have explanatory and predictive power. These principles focus primarily on spatial relationships between people and places, and they can be used to make sense of patterns in geographic data, answer questions, and construct solutions.

In Table 9, we present some examples of specific geographic practices from AFGS21. These tasks can be used as part of instruction and assessment.

Using AFGS21 to Develop Specific Assessment Frameworks

Earlier, we explained that, as a general assessment framework, AFGS21 is designed to serve as a blueprint for the development of targeted assessment frameworks for specific purposes. The Committee's intent is that

AFGS21 will give rise to a variety of specific frameworks for assessments that measure students' abilities to engage in complex reasoning and geographic practices. Our objective is to serve a wide range of assessment developers, including:

- committees developing nationwide or statewide assessments to inform policy and practice,
- state and district committees developing assessments for school-level accountability,
- classroom resource (e.g., textbooks and instructional materials) developers creating end-of-unit exams for use by teachers,
- classroom resource developers creating formative assessments for teachers to use to guide and inform instruction,
- researchers designing instruments to evaluate the efficacy of one or more educational curricula or programs,
- district or school committees developing end-of-course exams to assess individual student achievement, and
- classroom teachers developing, refining, and aligning courses to revised state standards using the "backwards design" approach to instructional design (see Wiggins & McTighe, 2005).

Each of these audiences will bring different constraints and goals to the use of AFGS21, and their particular purpose will lead them to create different frameworks. The benefit of AFGS21 is that offers each user a process for developing frameworks and a structure that will enable them to create frameworks that remain faithful to the core content and practices of AFGS21.

The development of a specific geography assessment framework from AFGS21 is a 4-step process:

1. specification of content goals,
2. specification of cognition goals,
3. identification and clarification of performance expectations, and
4. specification of assessment characteristics.

Table 8. Comparison Between Geographic Practices in the AFGS21 and Geography for Life, and Scientific Practices as They Have Been Conceptualized in Recent Standards and Framework Documents

21st Century Framework for the Geographical Sciences (2012)	*Geography for Life*, 2nd Edition (2012)	NRC Framework for Science Education (2012)	College Board Science Standards for Success (2009)	ACT College Readiness Standards™ (2010)
Posing geographic questions	Asking geographic questions	Asking questions (science) defining problems (engineering)	Scientific questions	Scientific investigation
Acquiring geographic information	Acquiring geographic information	Planning and carrying out investigations	Generation of evidence	
Organizing geographic information	Organizing geographic information	Analyzing and interpreting data	Data analysis	Interpretation of data
Analyzing geographic information	Analyzing geographic information	Developing and using models	Quantitative applications	
Answering questions and designing solutions	Answering geographic questions	Constructing explanations (for science) and designing solutions (for engineering)	Evidence-based explanations and models	Evaluation of models, inferences, and experimental results
		Engaging in argument from evidence		
Communicating with geographic information		Obtaining, evaluating, and communicating information		

The primary product of this process is an assessment matrix with content descriptors across the columns, cognition descriptors down the rows, and performance expectations in the cells. The process also results in a set of specifications for assessment developers on how to generate assessments from the matrix, and some drafts of items as models to guide item development.

Step 1: Specification of Content

The first step in developing a specific assessment framework from AFGS21 is the identification and description of the content goals for assessment. The framework developer[9] must consider the purpose of the assessment to generate explicit objectives for the content. In many cases, the objectives at the coarsest level will be pre-determined. For example, the development of a statewide assessment typically starts with state standards, and the development of an end-of-course exam typically starts with course objectives and a course outline. However, these objectives may be articulated in a way that combines content with cognition. For example, the Virginia state standards include the following standard for world geography: The student will analyze past and present trends in human migration and cultural interaction as they are influenced by social, economic, political, and environmental factors (Virginia Department of Education, 2008). In this case, the objective would need to be revised to extract the content goal. In addition, objectives generally are expressed too vaguely, or are too broadly conceptual for use in developing assessments, so they must be "unpacked" to describe explicit goals (example in Chapter 5). This unpacking process produces concrete descriptions of the knowledge needed to have proficiency with the content objective, and these descriptions indicate the boundaries and the level of sophistication at which students may be assessed on that knowledge.

The product of Step 1 is a list of content knowledge and concepts that can be used to label the columns in an assessment framework.

Step 2: Specification of Cognition

The second step involves considering the goals for the cognitive dimension: specifying what students should know and understand, and the geographic practices they should be able to perform. For example, taking the term "analyze" from the Virginia standard above, the assessment developer must consider: What would be a satisfactory analysis of trends for the purpose of the assessment? Should students be able to perform statistical calculations? Or are observations and descriptions of patterns sufficient? Are there certain kinds of trends in migrations or cultural interactions that students should be able to analyze?

The process of elaborating content objectives is a familiar one to those involved with designing assessments. The process of specifying the first level of geographic cognition, *knowing and understanding*, also has been explored in detail. By contrast, elaborating geographic practices is much less common. Descriptions of practices to be assessed as detailed as we advocate here are even more rare, so we provide a thorough explanation of how the AFGS21 can serve as a model for completing this step.

Framework developers first need to consider the purpose of their framework to determine which general practices they need to probe. It is not necessary that a framework should include each of the geographic practices, but it is important that the decision to include or exclude practices from an assessment framework be made deliberately. The framework developers must ask: What does the targeted population of students need to be able to do with their geographic knowledge? What kinds of geographic questions and problems should they be able to investigate? What geographic practices would be required for those investigations, and which of those practices should students be assessed on? Based on the answers to these questions, they should select the required rows of the general AFGS21 framework to incorporate into their specific framework.

The first level of geographic practices in the AFGS21, the categories of practices, is described at a general, abstract level. To target precisely the practices that are necessary and sufficient for investigating the geographic questions and problems of interest, the practices must be described more explicitly. For that reason, the Committee has developed a set of prompts to assist framework developers in generating a detailed set of geographic practices (Table 10).

These prompts pose questions about which specific geographic skills, practices, questions, and data are appropriate at different stages of learning, and the amount of structure that a learner at that stage might require. The prompts about structure are based on the model of a learner's zone of proximal development (Vygotsky, 1978), in other words, the range of tasks that a learner is capable of doing with a certain amount of external

[9] In this report we distinguish creators of assessment frameworks from creators of assessments. Framework developers create assessment frameworks. Assessment developers use assessment frameworks to build assessments to be administered to students.

Table 9. Examples of Kinds of Tasks That Can Be Used to Demonstrate Competency with Geographic Practices

Practices		Example Tasks
Posing geographic questions	a. Identify problems or questions that can be addressed using geographic principles, models, and data; express problems and questions in geographic terms.	• Describe geographic problems and questions that can be addressed using a observations, measurements, geographic concepts, and models • Recognize whether or not questions and problems are scientifically investigable using geographic information. • State questions and problems using geographic terminology and ways of thinking
Acquiring geographic information	a. Identify geographic data that can help to answer a question or solve a problem.	• Plan how to collect data via both primary and secondary sources • Determine which methods of data collection are most appropriate in addressing a given question or problem
	b. Collect data (incl. observations or measurements) about geographic phenomena, and/or gather existing data to help answer a question or solve a problem.	• Describe the ways in which and the places where data can be collected • Collect original data as well as data from secondary sources
Organizing geographic information	a. Organize data and create representations of data to help solve a problem or answer a question	• Recognize/describe ways to organize or represent data • Organize data or create an original representation • Explain why a type of representation will or will not help in addressing a question • Compare the effectiveness of organizational structures or representations
Analyzing geographic information	a. Identify data analysis strategies that can be used to help solve a geographic problem or answer a question.	• Identify and describe the methods with which data can be analyzed • Describe the qualitative and quantitative analyses and appropriate strategies for solving the problem • Generate or select parameters for querying data to find a location • Explain which parameters are/aren't helpful in solving the problem
	b. Find and describe spatial and temporal patterns in data, or find data that matches a pattern, to help answer a question or solve a problem.	• Comparing, classifying, finding, and describing relationships, associations and trends in data • Characterize the data patterns (i.e. linearity, clustering, dispersion, etc.) • Extrapolate from a pattern in data in order to infer how the pattern extends in (i.e. the past, present, other geographic regions) • Apply parameters to a data set and describe how well the data for different places match the desired criteria • Look at data to indicate places with matching attributes
	c. Construct an explanation or prediction for phenomena by comparing data to a principle or model.	• Develop explanations for phenomena by comparing relationships and patterns in data to geographic principles and models • Compare data as prompted by a given principle, theory or model. • Describe the extent to which data fit the generalizations, theories and models
Answering questions and designing solutions	a. Construct an answer to a question or a solution to a problem using geographic principles, models, and data.	• Answer a question/solve a problem using an analysis of the relevant data to justify the answer • Explain how an answer is supported by a given data representation depicting a pattern or a trend
	b. Evaluate one or more answers to a question or solutions to a problem using geographic principles, models, and data.	• Evaluate a response to a geographic question, a solution to a geographic problem, or an opinion about a geographic issue • Compare an argument to data to determine the extent to which data supports it • Describe information they would use to evaluate the argument • Compare several arguments and explain why they are/aren't supported by the data • Discuss whether given data is enough to support a given argument
Communicating with geographic information	a. Communicate using geographic data, principles, and models to educate or persuade an audience.	• Address an audience by writing about, making a representation of, or orally presenting an argument or solution to a geographic problem • Describe the information they would include based on their audience • Compare different examples of communication and infer who they think the audience and the examples' effectiveness for the given audience

Table 10. The Geographic Practices with Prompts for Creating a Specific Framework

Practices		Example Tasks
Posing geographic questions	a. Identify problems or questions that can be addressed using geographic principles, models, and data; express problems and questions in geographic terms.	• What specific kinds of problems and questions are appropriate for students at this stage? • How much support and structure do students at this stage require to formulate a question or problem?
Acquiring geographic information	a. Identify geographic data that can help to answer a question or solve a problem.	• What data are appropriate for students at this stage?
	b. Collect data (incl. observations or measurements) about geographic phenomena, and/or gather existing data to help answer a question or solve a problem.	• What forms of observation and measurement techniques are appropriate for students at this stage? • Are students capable of designing a data collection protocol independently, or do they require structure and support? • What sampling strategies are appropriate for students at this stage? • What kinds of archives for existing data are appropriate for students at this stage? • How much structure and support do students need to locate and assemble data from existing archives?
Organizing geographic information	a. Organize data and create representations of data to help solve a problem or answer a question.	• Which types of maps and other data representations are students prepared to construct and interpret? • How refined should the data be before students are asked to organize it? • Should students be able to choose how to organize the data, or are they just expected to be able to organize it in a given way?
Analyzing geographic information	a. Identify data analysis strategies that can be used to help solve a geographic problem or answer a question.	• What data analysis techniques should students at this stage know about and be able to perform?
	b. Find and describe spatial and temporal patterns in data, or find data that matches a pattern, to help answer a question or solve a problem.	• Which kinds of patterns should students be able to identify? • What forms and representations of data should students be able to read for patterns? • Which processes for identifying patterns in data are appropriate? • How much structure do they require for conducting the identification/description process?
	c. Construct an explanation or prediction for phenomena by comparing data to a principle or model.	• What kinds or forms of data should students be able to compare against explanations or predictions? • Are there certain inconsistencies between explanations for a phenomenon and data for that phenomenon that students should be able to identify? • How structured should this comparison be for students?
Answering questions and designing solutions	a. Construct an answer to a question or a solution to a problem using geographic principles, models, and data.	• Are there specific kinds of questions and problems should students be able to respond to? • How structured should the response be; should students be expected to use the data and generate a response? Should they be given a response and asked to support it with data?
	b. Evaluate one or more answers to a question or solutions to a problem using geographic principles, models, and data.	• In what ways should students be able to judge how well evidence supports an argument? • How much structure should they be given in their evaluation and comparisons of alternative arguments?
Communicating with geographic information	a. Communicate using geographic data, principles, and models to educate or persuade an audience.	• In what formats should students be expected to communicate? • Would students be judged on how well they communicate as well as the appropriateness of the mode of communication that they choose or the quality of the supporting information? • Should students be able to select the appropriate supporting information or should that be given to them? • Are there certain audiences students should be prepared to communicate an argument or solution to?

structure or support. For example, at one moment in a process of learning and development, a learner might be able to interpret spatial patterns on a map with the assistance of a sequence of prompting questions. At a later stage, students might be expected to complete the same task without any external support or structure. So, the framework developer must consider the stage of learning that is appropriate to expect for the population of students that will be assessed.

To develop an assessment framework tailored for a specific purpose, framework developers may use these prompts and their understanding of the audience to create specific descriptions of practices to be assessed. Framework developers might find a different set of prompts more suitable for describing the expectations for their particular population, so they should not be limited to this list of questions, but they could consider it a model for creating their own prompts.

The product of Step 2 is a list of practices that are used to label the rows in the assessment matrix.

Step 3: Identification and Elaboration of Performance Expectations

In the third step, framework developers use the intersections of the two dimensions to systematically define a set of performance expectations that describe the application of practices to content.

The description of performance expectations is created by filling in the cells in the body of the assessment matrix with a statement that combines the content from the column and the cognition from the row. When the cells are all filled in, they encompass the full range of the performances that could be assessed across the domain. Each of the cells is filled with at least one assessable description of what students should know and be able to do. Depending on the breadth of content in the column and the range of cognition in the row, there can be a large number of possible performance expectations in each cell. The decision about how comprehensive the cell should be depends on the purpose of the assessment. A framework that will be used to generate formative and summative assessments, for example, might require several performance expectations for each cell. Performance expectations always should be reviewed to ensure that they are clear, concrete, and specific enough to provide sufficient guidance for assessment developers.

Once the performance expectations have been written, framework developers should provide guidance to assessment developers on how to select performance expectations to use in developing items. In most cases, there will be more performance expectations in a framework than can be practically assessed, so framework developers should indicate to item writers how to sample across the matrix in a manner that would reflect their priorities. This can be communicated by attaching a weight to each performance expectation or section of the matrix to signal how often items that probe those cells should be represented, as in Table 5. Alternatively, framework developers can simply select which cells are the priorities for the item developers, which is the approach used in the example framework in Chapter 5. In that example, specific cells on the matrix are highlighted for developers to indicate that the performance expectations in those cells should be assessed.

The product of Step 3 is a completed assessment matrix for the specific assessment framework together with guidance on how to sample from the performance expectations in the cells.

Step 4: Specification of Assessment Characteristics

In the fourth step, for each performance expectation framework developers specify types of items, including cognitive demands and item formats. The cognitive demands (see details in Chapter 3) describe the degree to which students are expected to manipulate information to complete an item. For example, the framework developers could designate that an item probing a certain cell should draw on declarative knowledge, whereas an item probing another performance should test students' procedural knowledge. There could be more than one item probing the same performance expectation, but each item tests students' competency at a different cognitive level. Most importantly, there should be a range of cognitive demands specified across the framework matrix.

Likewise, framework developers should consider the nature of the performance expectation and cognitive demand to designate an item format for each cell that will be assessed to ensure that students have a variety of opportunities to demonstrate their competencies. In deciding item formats, the following factors should be considered (see Table 2): the purpose of the assessment, the construct measured, practical constraints that the assessment developers will be operating under, the context in which the assessment will be administered and scored, the student population that will be assessed, and the way the assessment information is to be used. These factors will guide the specifications in the development process to ensure that a variety of formats are being used across the matrix, and ideally, that students are given the opportunity to be assessed on any performance using multiple forms.

In addition to specifying item types, we recommend that framework developers also provide sample items to assessment developers to ensure that they are communicating their intent. These samples should illustrate the approach that framework developers recommend, including format, cognitive demand, and other characteristics. Providing samples helps to give assessment developers an understanding of the kind of information that the framework developers want the assessments to collect about the students.

The product of Step 4 is a specification of the kinds of items that assessment developers should use in the construction of the assessment, as well as prototype items that fit these specifications.

Implications for Improving Geography Education

As a general framework, the primary role for AFGS21 is to communicate what should be assessed and how an assessment can be built around those objectives. Its central message is that the goal of geography education is a balance of the main components in the cognitive dimension: knowing, understanding, and engaging in practices.

Beyond its communicative function, the framework has a practical role: to support the creation of frameworks targeted to specific purposes, which in turn, will guide the design of high-quality assessments that advance the goals of geography education. When used in this way, the framework serves as a blueprint for frameworks that can serve these purposes in a concrete and practical way. These specific frameworks will convey to assessment developers the importance of balance among knowing, understanding, and engaging in practices and the integration of practices with content that are embodied in AFGS21's structure. To illustrate what this might look like in practice, we provide an example of the process of developing a specific framework using AFGS21 in Chapter 5.

Chapter 5: An Example Geography Assessment Framework

In the previous chapter, we introduced a general assessment framework for geography, the *21st Century Assessment Framework for the Geographical Sciences* (AFGS21). We describe it as a general framework because it is too broad for practical use; it encompasses the entire set of content standards from the second edition of *Geography for Life: National Geography Standards* (Heffron & Downs, 2012) and all of the geographic practices. The purpose of AFGS21 is to serve as a blueprint or template for specific assessment frameworks with narrower scope of content and practices. In Chapter 4, we also described a procedure for creating specific frameworks for a targeted set of learning objectives and audiences from the general AFGS21 framework. In this chapter, we give an example of that process and its products.[10]

To generate this example, the Committee convened a working group with the charge of creating an assessment framework from AFGS21 for a specific purpose and context. The process was set up to follow as closely as possible the process that a working group from a school district or state agency would follow. In the remainder of the chapter, we describe the entire process from determining the composition of the working group, to creating the assessment matrix, to creating sample assessment items.

Assembling the Framework Development Team

The construction of an assessment framework requires a team with expertise in the content, assessment, teaching, and in some cases, policy of geography (Shavelson et al., 2008). Assessment experts often can manage the technical aspects behind what needs to be in a framework, but capturing the essential learning goals requires an understanding of the discipline that only a content expert can bring. In turn, a content expert can ensure accuracy to the discipline, but an educator also must be involved to check that expectations are reasonable and appropriate for the students. Finally, for large-scale assessments district or state administrators are required to bring an understanding of how the assessments will be used within the educational system.

In the development of the assessment framework presented here, we assembled a team consisting of experts in geography and assessment, classroom geography teachers, and district and state-level administrators. The team worked together to outline goals, create a framework, and describe specifications for item developers. The team was led by the Assessment Committee's research director.

Preparing the Team

The framework development team was given the task of developing an assessment framework for a six-week high school unit on human settlement. They were asked to undertake this task as if the framework was to be used for developing assessments to be used in several different settings across a district or a state. To make the task concrete, they were told that the unit would cover a cluster of high school social studies standards from Virginia, so their assessment framework should be designed around assessing those standards. They also were encouraged to integrate assessment of Virginia standards for English and mathematics if appropriate.

Prior to the development team beginning their work, we clarified how the assessment framework they were being asked to create might be used by a school or district. They were told to approach their task as if their framework was going to be used as part of an explicit effort to move instruction beyond "knowing and understanding" to include geographic practices. So, we asked them to think about their assessment framework as part of a strategy for communicating the importance of teaching geographic content and how that content can be used to answer questions and solve problems using geographic practices.

We also told the team that the framework and the performance expectations in the framework should support the creation of both formative and summative assessments for teachers and, potentially, the design of large-scale assessments for districts or states. As part of developing this framework, we asked them to create sample assessments. Specifically, we requested examples of formative assessments for use during project-based instruction and examples of summative assessments that could be used as part of an end-of-unit or end-of-year assessment of student's overall understanding of content and practices from the unit.

[10] This chapter includes excerpts of the example framework and drafts of assessments derived from the framework. See Appendix B for the entire example framework and drafts of formative and summative assessments designed to be models for assessment developers to follow.

Finally, the development team was told that this example was not intended to serve as a practical assessment framework—in part because that would require more extensive review, and in part because we would not have time to fill in all of the columns. Instead, it would be an illustration of the process of generating a specific assessment framework based on the generic framework, AFGS21.

The specific standards that the development team was given to work from were from the Virginia 12th grade Social Studies Standards (Virginia Department of Education, 2008):

Standards for World Geography (WG)

WG.5 The student will compare and contrast the distribution, growth rates, and characteristics of human population in terms of settlement patterns and the location of natural and capital resources.

WG.7 The student will identify types of natural, human, and capital resources and explain their significance by

a) showing their influence on patterns of economic activity and land use, and

b) evaluating perspectives and consequences regarding the use of resources.

WG.9 The student will analyze the global patterns and networks of economic interdependence by

a) identifying factors, including comparative advantage, that influence economic activities and trade;

b) describing ways that economic and social interactions have changed over time and

c) mapping, describing, and evaluating the formation of economic unions.

WG.11 The student will analyze the patterns of urban development by

a) applying the concepts of site and situation to major cities in each region,

b) explaining how the functions of towns and cities have changed over time, and

c) describing the unique influence of urban areas and some challenges they face.

Skills for World Geography

WG.1 The student will use maps, globes, satellite images, photographs, or diagrams to

a) obtain geographical information about the world's countries, cities, and environments;

b) apply the concepts of location, scale, map projection, or orientation;

c) develop and refine mental maps of world regions;

d) create and compare political, physical, and thematic maps and

e) analyze and explain how different cultures use maps and other visual images to reflect their own interests and ambitions.

WG.12 The student will apply geography to interpret the past, understand the present, and plan for the future by

a) using geographic knowledge, skills, and perspectives to analyze problems and make decisions and

b) relating current events to the physical and human characteristics of places and regions.

Standards for 12th Grade English

12.1 The student will make a five- to ten-minute formal oral presentation.

a) Choose the purpose of the presentation: to defend a position, to entertain an audience, or to explain information.

b) Use a well-structured narrative or logical argument.

c) Use details, illustrations, statistics, comparisons, and analogies to support purposes.

d) Use visual aids or technology to support presentation.

e) Use grammatically correct language, including vocabulary appropriate to the topic, audience, and purpose.

12.2 The student will evaluate formal presentations.

a) Critique relationships among purpose, audience, and content of presentations.

b) Critique effectiveness of presentations.

12.4 The student will read and analyze a variety of informational materials, including electronic resources.

a) Identify formats common to new publications and information resources.

b) Recognize and apply specialized informational vocabulary.

c) Evaluate a product based on analysis of the accompanying warranty and instruction manual.

d) Evaluate the quality of informational and technical materials.

12.8 The student will write documented research papers.

a) Identify and understand the ethical issues of research and documentation.

b) Evaluate the accuracy and usefulness of information.

c) Synthesize information to support the thesis.

d) Present information in a logical manner.

e) Cite sources of information, using a standard method of documentation, such as that of the Modern Language Association (MLA) or the American Psychological Association (APA).

f) Edit copies for correct use of language, spelling, punctuation, and capitalization.

g) Proofread final copy and prepare document for publication or submission.

Developing the Framework

The framework development process took place in two parts: matrix development and specification of assessment characteristics.

Part 1. Developing the Assessment Matrix

Step 1: Selection and description of content categories. The first step in creating an assessment matrix from AFGS21 is selecting the content to be covered and categorizing it into the columns that organize the content dimension. Drawing on the team's understanding of the domain and experience with teaching these topics to high school students, the team proposed that the framework should include the six categories of content as shown in Figure 15. Because time was limited, the team was asked to select only four of the six themes to flesh out for the example framework. The categories they selected are indicated with an asterisk.

With the content categories defined, the next step is to elaborate them in the form of statements that describe in more detail what students will need to know and understand. Because these are descriptions of content, not cognition, these content statements take the form of principles, concepts, and models. To identify the content statements for their framework, the team posed questions about what concepts, principles, and models are essential for the kind of reasoning about human settlements that students might encounter around them or in the news. For example, the team articulated the concept in Figure 16 to describe that set of relationships that students should know and understand to be able to address questions and problems around resource distribution.

Step 2: Selection and description of cognition categories. Once the content categories have been defined in explicit statements, the next step is to select and define the categories of cognition to be assessed. Unless the assessment framework is being used exclusively to develop items that require practices, then the "knowing" and "understanding" categories always are included in a framework. For the example framework, the development team considered all 12 of the geographic practices under the six practice categories, and they described in detail what they thought high school students should be able to do with geographic data by the end of a six-week unit on human settlement. These practices are not an exhaustive list of everything that could be assessed; they represent what the group, based on their expertise, decided were high priorities for geographic reasoning

Figure 15. Content Categories for a High School Unit on Human Settlement

***1. Resource distribution**
2. Safety and security
***3. People design places**
***4. Systems of interconnections**
***5. Characteristics of people and places**
6. Urban change

Figure 16. An Example of a Content Statement Under the Resource Distribution Content Category for the Example Framework

The resources needed to fulfill human needs are unevenly distributed across space and over time, which influences the location, size, and spatial organization of settlements, and requires movement of resources and people between places.

Figure 17. An Example of a Set of Geographic Practices Specified for the Cognition Dimension of the Example Framework

4b. *Find and describe spatial and temporal patterns in data, or find data that match a pattern, to help answer a question or solve a problem.*

Students find patterns in data including clusters and nodes, edges and boundaries, population pyramids, anomalies to patterns and arrangements, gradation, size, hierarchy, sequences, and order across space.

around the topic of human settlement. For example, Figure 17 shows that for practice 4b under "Analyze Geographic Information," the development team specified the following set of spatial and temporal patterns that high school students might look for to answer geographic questions and to solve problems around human settlements.

The process of specifying the content and cognition dimensions is necessarily iterative; as the team moved to the next step and started considering specific phenomena and what students would need to do with that phenomena to answer geographic questions and solve problems, they continually returned to the descriptions of content and cognition to make sure that all of the knowledge, skills, and practices students would need were represented. Underlying all decisions were questions about what is appropriate for high school students, what is reasonable for a six-week unit, and what are realistic expectations in terms of resources available to public school classrooms.

Step 3: Description of performance expectations—know and understand statements. Once the two dimensions of the matrix have been laid out, the next step is to begin describing performance expectations. The team started specifying performance expectations by focusing on the first two rows, the *Knowing and Understanding* statements. They began with the concepts and conceptual models that had been described for the content dimension in Step 1, and they translated its components into assessable *knowing* statements. The team considered everything students would need to know to recognize, identify, define, or describe to show that they know each concept and how they function. The team went through the same process with the *understanding* category. Again, they started with the concepts and translated their components into assessable statements where students have to describe a complex system of relationships or explain how or why a phenomenon occurs. For example, Figure 18 shows the performance expectations describing what students should be able to *identify or describe* to have the knowledge they will need for resource distribution.

The performance expectations in the *Knowing and Understanding* rows will be used as the objectives for assessing students' competence with the content. The performances in these two rows require only content knowledge, whereas all performances in the rest of the matrix ask students to apply elements of this content to geographic questions and problems using geographic practices.

Step 4: Description of performance expectations—geographic practices. To describe performance expectations for the *Geographic Practices* rows, the team crossed the conceptual model in the content dimension with the practice in the cognition dimension; in the cell where they intersect the team described something a student could be expected to do that would show that the student could use both the content and the practice to answer a geographic question or solve a problem. The team used the specific *Knowing and Understanding* statements and specific skills and practices to generate these concrete descriptions. Team members also brainstormed phenomena that they would like students to have experience reasoning through, and they used that as a context to help identify appropriate, realistic, and engaging performances. For example, under Resource Distribution on the content dimension and *Construct an Explanation or Prediction* in the practices dimension, they describe performance expectations (Figure 19).

Figure 18. Performance Expectations for the Example Framework Aligned with Resource Distribution (Content) and Knowing and Understanding (Cognition)

Identify or describe a geographic principle or phenomenon

- Identify factors that influence resource distribution.
- Identify SEPE (social, economic, political, environmental) factors that influence how and where people settle.
- Give examples of resources that are unevenly distributed.
- Describe ways that settlements use resources.
- Describe ways that settlements acquire resources.
- Identify resources needed by settlements.
- Identify relationships between settlements and the environment.
- Identify ways that the human capital impacts a settlement.
- Describe decisions that settlements make about resources.
- Describe ways that resource distribution can change over time.
- Identify reasons why access to resources can be limited.

The development team articulated the performance expectations in such a way that any performance could be assessed with a variety of different contexts. Therefore, any performance could be assessed multiple

times in multiple ways, providing deep insight into the students' competence with each objective. Furthermore, multiple items aligned to the same performance expectation could require different cognitive demands, or could use different item formats, each providing distinct information about a students' ability to reason around that content and practice.

Part 2. Specifying Assessment Characteristics

Once the development team had filled in all of the performance expectation cells (see the finished framework in Appendix B), the team turned its attention toward characteristics of items for the assessment. They were given two different purposes for assessment: (1) one was to be a formative assessment, designed to elicit student thinking with each of the selected performances, such that the teacher could diagnose if each student was developing the desired competency, and if not, where the student was getting confused; and (2) the other assessment was to be a summative assessment of student achievement, designed to sample student competencies at a broader scale across the whole matrix. These items would evaluate how well the student mastered the performance expectations. Although the items would span a wide range of difficulty so the teacher can capture students' who lie on the upper and lower ends of the spectrum of mastery, the items would not need to provide detailed information about the students' incorrect thinking. The team split into two groups, one for each assessment.

Step 5a: Development of a project-based learning and performance assessment. The formative assessment group described a specific problem that a high school class might investigate to illustrate how the framework could be used to structure project-based learning and performance assessments around the performance expectations described in the framework. The project is designed to provide students with experience applying their knowledge of concepts around resource distribution using some of the geographic practices, such as identifying questions that can be addressed using geographic principles, models and data, and expressing those questions in geographic terms. The group then designed prompts for the teacher to use to check students' progress toward competency with the performance expectation throughout the project, and a scoring rubric to evaluate their responses. Figure 20 shows a task given to students as part of the project in which students propose a geographic question that they could investigate in an effort to solve a geographic problem. The activity provides information to help teachers evaluate how well their students are mastering this practice (see the rubric illustrated in Table 11).

Figure 19. Performance Expectations for the Example Framework for Resource Distribution (Content) and Construct an Explanation or Prediction (Cognition)

- Compare trends of changes to resource demands by a settlement that has been studied to the resource demands of a settlement with similar characteristics to make a prediction about how demands may change in the future.
- Evaluate how well a pattern of settlement could be explained by dependence on a resource using models of how different populations depend on resources based on economic factors, social factors, cultural factors, or environmental factors.
- Evaluate implications of different decisions on how to distribute access to a resource by comparing each option to an appropriate model of how allowing or limiting access to that resource affects a population, such as comparing how allowing open enrollment to a school has affected the students, the community around the school, and communities in the region.

Step 5b: Development of an assessment for achievement. The achievement assessment group evaluated the entire matrix, and discussed priorities for a 45-minute long end-of-unit test. The group decided that Practices 1 (*posing questions*) and 6 (*communicating*) would be more effectively assessed through embedded performance assessments, such as an end-of-project evaluation of students' products, so those rows were eliminated from consideration for their assessment. The group then selected a set of cells from the matrix to use for item development, at least two from each row and each column to ensure that each part of the content and cognition dimension was being assessed in more than one way (Table 12).

The group then created a smaller matrix consisting only of the cells they aimed to target with this test, and the group added item development recommendations about item characteristics (Table 12). This distribution ensures that the test will cover the range of item formats and cognitive demand across the practices and content, allowing students opportunities to demonstrate their competencies in different ways and at different levels. A

Table 11. Scoring Rubric for Formative Assessments Embedded in a Project Aligned with the Example Framework

Skill	Developing	Proficient	Advanced
Posing geographic questions	Limited participation in brainstorming of possible geographic questions	Actively participates in brainstorming of possible geographic questions	Takes leadership role in brainstorming of possible geographic questions
	Poses questions related to the overarching project that may not be geographic, investigable, or relevant	Poses geographic questions related to the overarching project some of which are investigable or relevant	Poses multiple, investigable, relevant geographic questions related to the overarching project
Acquiring geographic information	Loosely connects the data to the question and assesses the quality of the data for the question (obvious gaps are present)	Connects the data to the questions and assesses the quality of the data for the question (minor gaps may exist)	Thoroughly connects the data to the questions, and assesses the quality of the data for the question (no gaps are present)
	May not identify parts of questions that are addressed by existing data; specifies some data, not necessarily geographic, to address those questions	Identifies some parts of questions that are not addressed by existing data; specifies geographic data to address those questions	Identifies all parts of questions that are not addressed by existing data; specifies a variety of geographic data and sources to address those questions
Organizing geographic information	Identifies existing theories with which to construct an explanation (may not be appropriate or geographic)	Identifies appropriate existing geographic theories with which to construct an explanation	Identifies a range of appropriate existing geographic theories with which to construct an explanation
	Constructs a limited explanation; may not fully compare interpretation of the data to existing theories or models	Constructs an explanation that compares interpretation of the data to existing theories or models	Constructs an explanation that fully compares interpretation of the data to the range of possible existing theories or models
	Evaluates how well the theory or model explains the data (obvious gaps are present)	Evaluates how well the theory or model explains the data (minor gaps may exist)	Evaluates how well the theory or model explains the data (no gaps are present)
Analyzing geographic information	May or may not determine if the initial geographic guiding question needs to be refined or revisited and, if needed, partially updates the question and analysis	Determines if the initial geographic guiding question needs to be refined or revisited and, if needed, updates the question and analysis	Determines if the initial geographic guiding question needs to be refined or revisited and, if needed, thoroughly updates the question and analysis
	Constructs answers to the question that are consistent with the data. Uses the explanations constructed to support the answers (obvious gaps are present)	Constructs answers to the question that are consistent with the data; uses the explanations constructed to support the answers (some gaps may exist)	Constructs thorough answers to the question that are consistent with the data; uses the explanations constructed to support the answers (no gaps exist)
	Uses the answers to propose a partial solution to the overarching geographic problem	Uses the answers to propose a solution to the overarching geographic problem.	Uses the answers to propose multiple solutions to the overarching geographic problem.
Answering questions and designing solutions	Partially identifies the level of detail and types of representations appropriate for the intended audience (obvious gaps are present)	Identifies the level of detail and types of representations appropriate for the intended audience (some gaps may exist)	Completely identifies the level of detail and types of representations appropriate for the intended audience (no gaps exist)
	Produces and delivers an incomplete presentation communicating possible solutions to the overarching problem	Produces and delivers an effective and geographically supported presentation communicating possible solutions to the overarching problem (minor gaps may exist)	Produces and delivers a convincing and geographically sound presentation communicating possible solutions to the overarching problem
Communicating with geographic information	Partially answers questions from the audience by referring to data, analysis and constructed representations	Answers questions from the audience by referring to data, analysis, and constructed representations	Comprehensively answers questions from the audience by referring to data, analysis, and constructed representations

Figure 20. An Excerpt from a Project Aligned to the Example Assessment Framework with Embedded Formative Assessment

The project is based on the following scenario: The County Council in Washington, County, Maine, notices a decrease in tax revenue and loss of downtown businesses, such as the cafe and the drugstore. They are concerned about how to maintain services that they provide.

Your task: You have been contracted as a consultant to the Council to analyze theproblem and propose a solution.

Project stage 1 (pose geographic question):

Assignment:
(1) Brainstorm issues and develop questions related to these issues.
(2) Pose some questions that you want to research.

Resources available:
- table showing tax revenue over time
- newspaper article describing decline in central business district (cafe, drugstore, barber shop, etc.)
- Teacher has on-demand resources about the case: population, demographics, natural resources trends, employment

Sample student work
Advanced:
Are we losing population?
Is it because of economic opportunities?
Is it natural resources related?
Is it short-term or long-term issue?

Developing:
Where is Maine?
What is a tax collector?
What do they do for work in Maine?

Rubric for formative evaluation:
See Figure 21.

Instructor Note: Consider how to provide feedback to the students based on your formative assessment. For example, in this stage you might have a class discussion of the questions, or you might have each group present their questions to the class andreceive feedback.

test with items that sample the domain with this scope of content, practices, cognitive demands, and item formats also provides teachers a depth of information that can be used evaluate how well students mastered the topic and that can be used to pinpoint areas of weakness.

Step 6: Development of sample item outlines. The group recognized that the framework is a result of extensive conversations about how the high school students could be assessed on their development of knowledge and skills for this unit. The framework captures the essence of these discussions, but the group also wanted to ensure they were conveying their vision of how the items based on this framework might look. To that end, the group brainstormed several item outlines based on the framework specifications that item developers could use as guides for realizing the group's conceptualization of items (Figure 21). These descriptions are just conceptions of what an item aligned to the framework might look like. Although the descriptions are incomplete and they have not been validated, they serve as another type of recommendation to help item developers interpret the intent of the assessment framework experts.

Table 12. Matrix with Specifications for Item Developers for the Distribution of Items for a High School End-of-Unit Exam on Human Settlement

	Resource Distribution	People Design Places	Systems of Interconnections	Characteristics of People and Places
Knowing and understanding	selected response	selected response		
Posing geographic questions				
Acquiring geographic information			selected response	constructed response
Organizing geographic information	constructed response			selected response
Analyzing geographic information		constructed response	selected response	constructed response
Answering questions and designing solutions	constructed response		selected response	
Communicating using geographic information				

Figure 21. A Cluster of Three Item Outlines (Items 1-3) Created as Examples for Assessment Developers of the Types of Items to be Developed from the Framework

Item 1

Content: The nature of settlements is influenced by the interaction of characteristics of populations and physical environments.

Cognition: Identify geographic data that can help to answer a question or solve a problem.

Cognitive demand: strategic

Item format: constructed response-short answer

Performance expectation: Identify resources and data that would help answer a question about impact of policy decisions made for a settlement on how the population interacts with the physical environment.

A city has a problem that people are not recycling plastic bottles, and the bottles are ending up in the rivers and lakes. The city officials are trying to determine if they should institute a deposit fee, so they have asked your class to determine if a refundable bottle deposit would be an effective strategy for increasing recycling rates. Use the two maps below to analyze how effective deposit fees have been in other regions.

1. One map shows the distribution of deposit fees for plastic bottles.

2. The other map shows the incidence of plastic bottles in watersheds (landfill). The map needs to make watersheds clear so that students can identify a relationship between rivers and density of bottles on the landscape.

Is there any other information you would need to perform this analysis?

Item 2

Content: The nature of settlements is influenced by the interaction of characteristics of populations and physical environments.

Cognition: Find and describe spatial and temporal patterns in data, or find data that match a pattern, to help answer a question or solve a problem.

Cognitive demand: schematic

Item format: selected response

Performance expectation: Describe patterns or relationships shown between the maps using concepts of density and gradient.

Compare the two maps. Are there patterns to where most bottles are found and the fewest bottles are found across the region? Where the bottles are, are there any patterns to where they appear within the landscape?

- Students should note where the highest density of bottles occur relative to areas where deposit fees are charged.

- Students should note where the highest density of bottles occur relative to watershed features.

Item 3

Content: The nature of settlements is influenced by the interaction of characteristics of populations and physical environments.

Cognition: Construct an explanation or prediction for phenomena by comparing data to a model or theory.

Cognitive demand: schematic

Item format: constructed response

Performance expectation: Make a prediction about how policy decisions made for a settlement will impact the population and landscape using observations about their relationship.

Justify whether a bottle deposit fee would be an effective strategy for increasing recycling rates using the patterns observed in the maps.

Conclusion

The matrix, assessment specifications, project outline, and items in this chapter comprise a specific assessment framework. The description of the process illustrates how a framework for a specific purpose can be created from the AFGS21. In addition, this example shows how AFGS21 can be used to align formative and summative assessment goals, where formative assessments monitor students' progress toward the goals and summative assessments evaluate how well students achieved those goals by the end of the instruction. Finally, the example shows how the framework matrix can be used to ensure that assessment developers probe the full range of targeted objectives, and that assessments provide learners with a variety of opportunities to demonstrate mastery or reveal weaknesses for each objective.

Chapter 6: Recommendations

Over the course of this report, we have built a case for the role that assessment can play in improving teaching and learning of geography. We also have presented evidence that current assessments of geography, by and large, do not capture the information needed to improve geography teaching and learning.

In this chapter, we present a set of specific recommendations for actions that will lead to the development of assessments that will support instructional improvement in geographical sciences. These recommendations are targeted at educators, policy makers, and funders.

A Focus on Assessment in Efforts to Improve Geography Education

The Committee believes that a focus on assessment, formative and summative, should be integral to any efforts to improve geographic education. According to our research, few classroom materials for geography are accompanied by high-quality assessments to assist teachers and students in monitoring learning. In addition, in the 2010-2011 academic year, fewer than one-third of states required students to take a high school exit exam or end-of-course assessment that assessed geographic knowledge or practices (Unpublished Geography Education National Implementation Project report, 2012). The National Assessment of Educational Progress (NAEP) assesses geography for elementary, middle, and high school students across the United States, but these tests have been occurring at seven- to nine-year intervals. Well-designed, properly timed assessments can provide every stakeholder in the system, from students to teachers to policy makers, with crucial information for improving teaching and learning.

Therefore, the Committee recommends:

1. Current state and national efforts to evaluate the status of geographic literacy should be continued, and additional resources should be allocated to improve and expand them. In particular, we recommend that the NAEP Geography assessment should be conducted at intervals consistent with other subjects of critical importance for workforce preparation and national security, such as science (which is conducted every four to five years) and math (conducted every two years).

A New Approach to the Assessment of Geography

To be useful for instructional improvement, assessments should capture the behaviors that the instruction is designed to cultivate. In the case of geography, that includes the ability to answer questions and solve problems by reasoning with geographic evidence. Our research indicates that existing geography assessments do not reflect the importance of these skills. As a result, existing assessments evaluate performances that are not representative of the objectives of geography education. Further, current assessments do not signal to educators and learners what those objectives are in the way that assessments should.

Therefore, the Committee recommends:

2. Assessments for geography should reflect the ways that geographic knowledge and skills are used in the world, including using evidence-based reasoning, problem solving, and communication.

The Committee has developed the *21st Century Assessment Framework for the Geographical Sciences* (AFGS21) to be used as a blueprint for this approach to assessment, and we have provided guidelines for using AFGS21 to create assessment frameworks for specific contexts and purposes.

Therefore, the Committee recommends:

3. The *21st Century Assessment Framework for the Geographical Sciences* should be used as the basis for designing the next generation of assessment frameworks and assessments for geography.

Because large-scale tests influence assessments conducted at the district, school, and classroom levels, it is particularly important that the tests represent the kind of 21st century geographic thinking on which we expect students to be assessed.

Therefore, the Committee recommends:

4. The NAEP assessment framework for geography should be revised to reflect this 21st century approach to assessment of geographic thinking.

The Committee considers the version of AFGS21 in this report to serve as a starting point, not a final

product. This general assessment framework should undergo continuing research and development to ensure that it can be used successfully to guide the development of assessments of the nature and quality that we are advocating in this report.

Therefore, the Committee recommends:

5. AFGS21 should be a focus of ongoing research and development in framework and assessment development for geography that will lead to continuous improvements in assessment of K–12 geography knowledge and skills.
6. A program of research should be initiated immediately to study learning progressions in geographic practices over the K–12 timeline and to study techniques for assessing mastery of geographic practices at all levels. The results of this research should be used to inform the refinement of the articulation of expectations for content and cognition in AFGS21.

Shared Frameworks and Assessments

The geography education reform community is small and has few resources. Therefore, these resources must be applied strategically and efficiently.

Therefore, the Committee recommends:

7. The field should develop a set of shared frameworks and assessments to serve assessment needs across a broad range of instructional improvement efforts. These should be addressed to areas of high need and broad applicability. For example: facility with map interpretation and analysis at an elementary level; key issues in human geography at elementary, middle, and high school levels; human-environmental interactions at middle and high school levels; environmental dynamics at elementary and middle school levels; and facility with geospatial technologies at a high school level.

Capacity Building

The degree to which the implementation of the recommendations in this report is successful will depend on the field's capacity to execute them. Specifically, we need professionals with the necessary geography expertise and assessment development training and experience to develop effective assessment frameworks, items, and instruments for geography. We need educators who have been trained in the best practices for incorporating assessments and their results to inform teaching and learning in the classroom.[11] Finally, we need educators and policy makers who have the expertise to make well-founded decisions based on the results of these assessments. Although expertise with creating assessments and integrating them into classroom practice is not abundant in the broad education community, it is vastly under-represented in the small and historically under-resourced geographic education community.

Therefore, the Committee recommends:

8. The field should invest in training and professional development programs to cultivate the assessment professionals that the geography education community needs, and to prepare teachers and policy makers to use and learn from a new generation of assessments.

Knowledge Base

A lack of research has impeded the Committee's efforts to develop empirically based guidelines for determining what geographic content and practices students at different grade levels should be assessed on and how they should be assessed.[12]

Therefore, the Committee recommends:

9. A substantial investment should be made in the development of research and experience based on: (1) the nature of understanding, reasoning, and learning geography to inform assessment design; and (2) techniques for assessing understanding and reasoning in geography.

[11] Recommendation 2 in the Road Map Project report on Instructional Materials and Professional Development also relates to the preparation of teachers for the use of assessments (Schell, Roth, & Mohan, 2013).

[12] The Road Map Project report on Geography Education Research also contains recommendations about the need for research that can inform the design and use of assessments (Bednarz, Heffron, & Huynh, 2013).

Appendix A: Additional Details About the Assessment Study

In this appendix, we provide additional details about the methods used in the study of current assessment practices described in Chapter 3. Specifically, we describe how items were selected for study, how they were coded, and how the codes were analyzed.

Item Selection

The process of selecting items for analysis consisted of a collection phase, a sampling phase, and a filtering phase.

In the collection phase, we created a database of geographic education materials that were likely to have assessment items, including all of the most recently released state, national, and international (English-speaking countries only) assessments. Our database also includes all classroom geography units that were accessible through Social Studies School Service, Amazon.com, the National Council for Geography Education, the College Board, Social Studies Central, and the Library of Congress. In many states, geography is not tested separately from other subjects, so we included both the science and social studies tests in those instances.

We randomly selected resources from the database separately by resource type, but we found that almost 60% of the classroom units selected had no assessments that fit our coding criteria, which resulted in under-sampling in that category overall and a significant bias toward middle school (Table A1). In addition, textbooks had so many items that they would be overrepresented in the study if we sampled them using the same method, and there was little variation between textbooks. So instead of sampling several textbooks, we chose one of the most widely used textbooks, and coded all of the items in each of the randomly selected chapters.

The methodological differences between sampling across the large-scale and classroom assessment categories led us to report the results separately. Selection of a representative sample was far more straightforward for large-scale assessments, so for the purpose of this report we focus the results mainly on those items, though we show results of classroom assessments where we find those results to be particularly relevant.

In the filtering phase, we eliminated items that did not assess geography by determining if the items assessed content or practices from any of the six essential elements in *Geography for Life: National Geography Standards* (Heffron & Downs, 2012).

The number of resources selected includes the number selected through random sampling of the list of resources (Table A2). The number of resources coded eliminates the resources with no geography items and also eliminates the resources included in assessments that were not available to the public.

Item Coding

Items selected for the study were coded according to our taxonomy for geographic assessments (Table 6), which describes item characteristics, targeted abilities, and confounding factors. In the sections below, we describe each of these coding categories and the codes that were used.

Table A1. Number of Items Coded from Geography Resources by Grade Band

	Large Scale	Classroom	Total
Elementary school	109	18	127
Middle school	131	214	345
High school	172	49	221
Total	412	281	693

Table A2. Distribution of Resources Sampled for the Study

	Large Scale	Classroom Resources
Number of resources	154	69
Number of resources selected	88	29
Number of resources coded	69	12

Item Characteristics

The **assessment setting** describes whether the item is designed as a proximal or distal evaluation of student learning. This category classifies items intended for a large-scale setting, including state, national, and international tests, or in a classroom setting, which includes items designed to accompany a geography unit or a ge-

ography textbook. Program assessments are identified in the taxonomy as an additional setting, but items from this setting were not included in this pilot study.

The **grade band** indicates whether the item is targeted to students in elementary, middle, or high school. In rare cases items are intended for more than one grade band, so we included those items with the lowest targeted grade.

Item format records how the question was asked, specifically how structured the question and responses are. Highly constrained response formats include multiple-choice items, which are characterized by whether the question is written with a complete or incomplete stem (e.g., fill-in-the-blank), in addition to the less common true/false and matching formats. Open-ended item structures include short answer, which is limited to words and phrases or partial sentences, whereas the paragraph and essays category includes multiple-sentence responses. An additional open-ended item format that is occasionally used in geography assessments asks students to construct a representation.

Type of representation characterizes the representation(s) used in the item so that we can collect data on what information, other than text, students are asked to evaluate as part of the item. For items that have a representation, this category records if it is an illustration or photograph, map or globe, graph or table, diagram, or document.

Targeted Ability

These categories describe what is being assessed by the item, including the content, practices, and how students are being asked to use the content and practices.

Table A3. Essential Elements and Standards from Geography for Life (2nd ed.)

The National Geography Standards: *Geography for Life* (2nd ed.)	
The World in Spatial Terms	1. How to use maps and other geographic representations, geospatial technologies, and spatial thinking to understand and communicate information 2. How to use mental maps to organize information about people, places, and environments in a spatial context 3. How to analyze the spatial organization of people, places, and environments on Earth's surface
Places and Regions	4. The physical and human characteristics of places 5. That people create regions to interpret Earth's complexity 6. How culture and experience influence people's perceptions of places and regions
Physical Systems	7. The physical processes that shape the patterns of Earth's surface 8. The characteristics and spatial distribution of ecosystems and biomes on Earth's surface
Human Systems	9. The characteristics, distribution, and migration of human populations on Earth's surface 10. The characteristics, distribution, and complexity of Earth's cultural mosaics 11. The patterns and networks of economic interdependence on Earth's surface 12. The processes, patterns, and functions of human settlement 13. How the forces of cooperation and conflict among people influence the division and control of Earth's surface
Environment and Society	14. How human actions modify the physical environment 15. How physical systems affect human systems 16. The changes that occur in the meaning, use, distribution, and importance of resources

Note: The sixth element is not included because it involves geographic practices, which are described in the cognitive dimension.

Geographic content describes the geographic content knowledge assessed by the item (Table A3). These categories are taken directly from the K–12 national standards recommendations *Geography for Life*. Each of these categories is discussed in detail in *Geography for Life*. The five categories included in the study are the essential elements that target geography content knowledge. There is a sixth essential element, *The Uses of Geography,* which we consider to be outside of the content domain.

Geographic practices scored for the study are the geographic practices as described in AFGS21 (see Chapter 1), with the exception of *collecting and organizing geographic information,* which was considered one practice category for this study, but has been split into two categories in AFGS21. These categories are explained in depth in Chapter 4. If an item assesses a practice, but the practice is not specifically geographic, such as reading a bar graph, the item was scored as requiring no geographic practices.

The **cognitive demands** category characterizes the degree to which students are asked to process information as shown in Table A4 (Li, Ruiz-Primo, & Shavelson, 2006; Ruiz-Primo, Shavelson, Hamilton, & Klein, 2002). Items that ask students to respond with declarative knowledge impose the lowest cognitive demand, while strategic knowledge requires the highest cognitive level by asking students to understand a topic at a sophisticated level such that they know how and when to apply practices and content to a new problem in a new setting.

Confounding Factors

An item might be intended to assess a student's ability across a given knowledge or skill domain, but if the item is poorly written, we cannot be certain that the student's performance on that item accurately reflects his or her ability. For example, an implausible answer choice can reduce the number of options, increasing the probability that the student could choose the correct answer without using the targeted knowledge or skill (i.e., false positive). Similarly, poor wording might confuse a student, causing him to select an incorrect answer choice even though he knew the correct answer (i.e., false negative).

Clarity includes descriptors of problems that students might have with reading and interpreting the item. This category includes problems such as unclear questions and/or answers, or representations that are unclear. Some items also are written with an unnecessarily high reading load because it has too many words or it has words that are more technical or sophisticated than is necessary. Items also are scored for insufficient information, which includes representations not being labeled well enough, the context being unfamiliar to most students, or not having enough supporting background to make sense.

Content accuracy describes problems with the substance of the item. This category records whether items have more than one answer choice that could be correct, no answer choice is correct, there is a conceptual error, and if the representation is unnecessary for evaluating the item.

Resistance to test-taking strategies is a category that describes ways that students can select or reject at least one answer choice using reasoning other than the targeted knowledge or skill. This category includes criteria that indicate if one answer choice is significantly longer or shorter than the others, or if some answer choices stand out because they seem more technical than the others. This category also includes criteria that indicate if an answer choice seems unlikely to be true, or if there is a grammatical mismatch between the stem and at least one answer choice; both of these criteria make an answer choice less plausible. An additional criterion records if an item has at least two answer choices that say or mean the same thing, allowing students to rule out both using logic instead of the targeted geographic skill or knowledge.

Table A4. The Categories of Cognitive Demands with Examples of Tasks That Typically Draw on These Demands

Cognitive Demands	Tasks
Declarative	• Identify, describe, classify • Recognize a true statement • Find relevant information in a text • Make a comparison between two variables
Procedural	• Read a map • Collect information from a table, graph, map, or representation • Plot data on a table, graph, map, or representation
Schematic	• Make a comparison across multiple variables • Explain or predict a phenomenon using a general principle or model • Perform evidence-based reasoning • Recognize and compare patterns
Strategic	• Find a solution to a problem by devising an appropriate method of approach

Analysis of Coding Reliability

All items were coded across the geographic assessment taxonomy by one researcher. A second researcher coded 10% of the items to establish inter-rater agreement, and to test for consistency of coding between independent raters. Percentage agreement ranged from 71-93% (mean 83%), and the kappa coefficient ranged from

showing fair to substantial agreement (0.249-0.654; $p < 0.001$, 95% CI). Reported findings were calculated using the first rater's results.

An area where inter-rater agreement was low was in coding for content category. Coders frequently found it difficult to assign items to a specific standard in the *Geography for Life* standards. The standards are presented in the form of broad concepts and principles, and the examples of performances under each standard cover only a small subset of the concepts and principles encompassed by a standard. When faced with an item that requires students to have very context-specific factual recall of geographic information, it was difficult to discern if we could reasonably expect that knowledge to be covered under any of the broadly defined standards. This dilemma occurred frequently throughout the scoring process, and we felt that we did not have enough information about expectations for each standard to reliably code the items. We coded according to our best estimation of which standard best fit the knowledge needed, and when we could not make that determination, we coded the item under the closest general standard heading. We scored items as "no content" only when the knowledge needed could not be considered geographic.

Appendix B: Example Assessment Framework

An excerpt from an example of an assessment framework for a high school unit on human settlement is shown below.

Download the full Example Assessment Framework at http://education.nationalgeographic.com/media/file/Example_Assessment_Framework_2.xlsx

		The resources needed to fulfill human needs are unevenly distributed across space and over time, which influences the location, size, and spatial organization of settlements, and requires movement of resources and people between places.	People design places and settlement patterns in ways that facilitate their access to resources, goods, and services.	Humans create interconnected sy differing characteristics of places and ideas among them.
Geographic Practices		**Performance Expectations**	**Performance Expectations**	**Performance E**
Know and understand	Identify or describe a geographic principle or phenomenon.	• Identify factors that influence resource distribution. • Identify SEPE factors that influence how and where people settle. • Give examples of resources that are unevenly distributed • Describe ways that settlements use resources. • Describe ways that settlements acquire resources. • Identify resources needed by settlements. • Identify relationships between settlements and the environment. • Identify ways that the human capital impacts a settlement • Describe decisions that settlements make about resources • Describe ways that resource distribution can change over time. • Identify reasons why access to resources can be limited.	• Identify/describe kinds of decisions humans make regarding the organization of space (zoning laws, aesthetics) • Identify/describe the kinds of rules that governments make that regulate how and where people can move within and among settlements. • Identify/describe a city's hinterland. • Identify/describe that different populations within a settlement have differential access to resources, goods, and services. • Identify constraints on the urban planning process in areas based on population change and limited resources. (megacities, declining cities; water shortages) • Identify/describe transportation and communication systems created to link settlements and their needed resources. • Identify/describe transportation and communication systems that form land use patterns within settlements. • Identify/describe how landscape tastes and cultural preferences influence the design of places. • Know that certain resources are particularly influential for the design of places such as access to water, energy, air quality.	• Describe influences of topographic f transportation and communication be • Describe transportation and commu within and between settlements. • Describe what spatial hierarchies ar and number of functions). • Describe political systems that contr interaction among settlements. • Describe the nature of flows and int settlements (movement of goods, pe
	Explain a geographic principle or phenomenon.	• Explain why settlements depend on [specific] resources, and why those dependencies change over time and space. • Explain why some settlements have and others do not have access to certain [specific] resources, and why access can change. • Explain why settlements occur around certain natural and capital resources (include reasons relevant across time and space). • Explain why changes in resources (amount, value, access) affect settlements.	• Describe how settlements are dependent upon their region beyond their settlement (hinterland) (resources, goods, services) [Von Thunen Model of city and agricultural land use; Christaller's Central Place Theory]. • Explain advantages and disadvantages between the planned design and unplanned design of places. • Identify/describe how transportation and communication systems form land use patterns within and between settlements. • Explain how transportation costs influence land use patterns within and among settlements. • Explain how the physical environment influences the design of places and	• Explain the hierarchical patterns of population and range of a good or ser • Explain the pattern of transportatio differential costs between modes of t • Explain how patterns of trade emer • Explain the pattern of political powe formal and functional region and bou • Explain Riley's Law

Appendix C: Assessment Committee Member and Staff Biographies

Daniel C. Edelson

Committee Chair

Daniel C. Edelson is vice president for Education, National Geographic Society, and executive director, National Geographic Education Foundation. As a curriculum designer, software developer, and educational researcher, Dr. Edelson has dedicated his career to improving young people's understanding of the world they live in and their role in determining its future. In his position as vice president for Education, he oversees National Geographic's outreach to educators and its efforts to improve geographic and geoscience education in the United States and abroad. This work includes the creation of educational materials for learners of all ages, the delivery of professional development for educators, the implementation of public engagement programs, advocacy on behalf of geographic education in policy discussions, and grantmaking to support geographic literacy initiatives throughout the United States and Canada. He has written extensively on motivation, classroom teaching and learning, educational technology, and teacher professional development. He is an author of more than 50 papers in journals, edited books, and conference proceedings, including *The Cambridge Handbook of the Learning Sciences, The International Handbook on Science Education, Journal of the Learning Sciences, Journal of Research on Science Teaching,* and *The Science Teacher.*

Richard J. Shavelson

Committee Co-Chair

Richard J. Shavelson is chief scientist and partner at SK Partners, LLC, and the Emeritus Margaret Jacks Professor of Education and Professor of Psychology (by courtesy), and former I. James Quillen Dean of the School of Education at Stanford University, and senior fellow in the Woods Institute for the Environment. He served as president of the American Educational Research Association, and he is a fellow of four professional associations and a Humboldt Fellow (Germany). His work includes assessment of undergraduates' learning, including the Collegiate Learning Assessment, accountability in higher education, assessment of science achievement, the enhancement of minorities' performance in organic chemistry, and the role of mental models of climate change on sustainability decisions and behavior. His publications include *Statistical Reasoning for the Behavioral Sciences, Generalizability Theory: A Primer* (with Noreen Webb), *Scientific Research in Education* (with Lisa Towne), and *Assessing College Learning Responsibly: Accountability in a New Era.*

Jill Wertheim

Committee Research Director

Jill Wertheim is program manager for evaluation and assessment in National Geographic Society's division of Education Programs. Dr. Wertheim has diverse experience in scientific research, assessment framework development, and assessment research and development. Her experience includes working for four years at AAAS Project 2061 as a research associate performing research on learning around earth science topics, research on assessment in earth science, as well as developing classroom assessment items and instructional materials, and evaluating items, instructional materials, and standards. Her scientific research includes studies on using fossils to uncover evolutionary patterns in South American mammals.

Barbara Hildebrant

Committee Member

Barbara Hildebrant is senior director, College Board Programs, Educational Testing Service (ETS). She brings more than 13 years of experience in assessment development to the Road Map Project. Her professional expertise includes academic program review, higher education, and testing program management. Her relevant experience includes developing and reviewing curriculum framework and form assembly specifications, developing tests, and managing scoring activities. Her academic areas of expertise are anthropology, geography, and world history. Prior to joining ETS, she taught geography and anthropology classes at Rutgers University. In 2008, she received the Association of American Geographers (AAG) Gilbert Grosvenor Honors for Geographic Education in recognition of her exceptional record of leadership in advancing geography education both at the university level and in K–12 schools. She earned her PhD in Geography from Rutgers University, her MA in Anthropology from the University of Washington at Seattle, and her BA in Anthropology from Drew University.

Elizabeth R. Hinde

Committee Member

Elizabeth R. Hinde is associate professor and director of the Division of Teacher Preparation at Arizona State University's Mary Lou Fulton Teachers College. Prior to her career in higher education, Dr. Hinde taught elementary school for 20 years. She is the author of numerous articles concerning social studies education and curriculum integration, and she has been recognized nationally for her work in curriculum development. She was research director of the Arizona Geographic Alliance's GeoLiteracy and GeoLiteracy for English Language Learners programs, and she was a member of the curriculum development team of iCivics.org. In 2005, Dr. Hinde received the National Council for Geographic Education's Distinguished Teaching Award, and she also received the 2010 Geography Excellence in Media Award. She is past-president of the Arizona Council for the Social Studies, teacher consultant with the Arizona Geographic Alliance, past-chair of the National Council for the Social Studies Steering Committee, and is active in numerous state and national professional organizations.

Marianne Kenney

Committee Member

Marianne Kenney is a social studies specialist/evaluator at Denver Public Schools. Ms. Kenney serves in the Teacher Effectiveness Unit funded by a grant from the Bill and Melinda Gates Foundation. She has been a senior consultant at McREL, the region's federally funded research laboratory, and the state social studies content specialist for the Colorado Department of Education. She has been directly involved in the development of standards at the national and state levels. She has directed many summer institutes to assist districts in designing both district and classroom performance assessments, as well as units of instruction that align with national, state, and/or district standards. Ms. Kenney has been on the advisory committee for the 1988, 1994, and 2001 National Assessment of Educational Progress Geography tests, and she has worked with the Council of Chief State School Officers to collaboratively develop state test items. She is an experienced presenter at workshops and conferences and has published numerous articles for education journals. Ms. Kenney taught high school geography for 17 years.

Bob Kolvoord

Committee Member

Bob Kolvoord is interim dean, College of Integrated Science and Engineering, James Madison University. Dr. Kolvoord also serves as the interim head of Engineering and as a professor of Integrated Science and Technology. His research interests focus on the use of data visualization and geospatial technologies in the K–16 classroom and the development of student spatial thinking skills through the use of these technologies. He is a co-creator of the Geospatial Semester program (http://www.isat.jmu.edu/geospatialsemester), and with Kathryn Keranen is the co-author of the award-winning *Making Spatial Decisions Using GIS* (1st and 2nd editions) and *Making Spatial Decisions Using Remote Sensing* (forthcoming). He was the recipient of a 2010 Commonwealth of Virginia Outstanding Faculty Award for teaching with technology.

David A. Lanegran

Committee Member

David A. Lanegran is John S. Holl Professor of Geography, Macalester College, St. Paul Minnesota. Dr. Lanegran is an urban geographer specializing in urban issues related to city planning, urban development, and historic preservation. He has conducted extensive comparative research on urban planning around the world and has published and spoken widely on urban, agricultural, and cultural geography. He has been a driving force in Macalester's nationally recognized programs to improve the teaching in geography in elementary and secondary schools. He is coordinator of the Minnesota Alliance for Geographic Education, and the chief reader of the Advanced Placement Human Geography exam. He has directed 70 summer institutes for geography teachers. He has received several honors, including Macalester's Jefferson Award for Teaching Excellence; The George J. Miller Award, from the National Council for Geographic Education; and the Gilbert Grosvenor Honors for Geographic Education from the Association of American Geographers. His publications include 12 books and more than 60 articles and chapters in professional publications.

Jody Smothers Marcello

Committee Member

Jody Smothers Marcello is a teacher at Sitka High School, Sitka, Alaska. Dr. Smothers Marcello is a National Board Certified teacher and educator; she teaches AP Human Geography (APHG), global issues, history, and English. She has received multiple teaching awards at the state and national levels, including from the National Council for Geographic Education (NCGE) and National Council for the Social Studies (NCSS). She has served as NCGE president, on the NCSS Board, on the National Board for Professional Teaching Standards Social Studies-History standards committee, and as a table leader for the APHG College Board reading. She has authored multiple articles and curricula, including recent publications for the College Board and NCGE. She was a teacher contributor to the 2011 book, *The American Public School Teacher: Past, Present, and Future,* she is a contributor to the Association of American Geographers' Center for Global Geography Education. She edits *The Geography Teacher* journal. She received BS and MEd degrees from Texas A&M University.

Robert W. Morrill

Committee Member

Robert W. Morrill is professor emeritus, Geography Virginia Tech, 1973-2003. Dr. Morrill serves as co-coordinator, Virginia Geographic Alliance. He has received human geography and geography education grants from the National Science Foundation, National Geographic Society, Commonwealth of Virginia, and Foundation for the Improvement of Post Secondary Education. Dr. Morrill led study abroad programs and teacher institutes to Switzerland, Italy, Germany, The Netherlands, Canada, Ecuador, Brazil, and New Zealand. He was a Fulbright Research Fellow (1986), University of Turku, Finland. A primary author for *Guidelines for Geographic Education* (1984) and *Geography for Life: National Standards in Geography* (1994), Writing Committee member for the *Geography Framework for the National Assessment for Educational Progress.* Dr. Morrill's work has been published in geography and education journals, curriculum monographs, atlases, and national geography education reports. In 1989, he served as president of the National Council for Geographic Education (NCGE). In 2007 he was awarded the NCGE George Miller Award for geography education contributions, and in 2012 he received the Association of American Geographers Gilbert Grosvenor Honors for Geographic Education.

Maria Ruiz-Primo

Committee Member

Maria Ruiz-Primo is an associate professor at the School of Education and Human Development, University of Colorado Denver. Her work focuses on two strands: assessment of students learning at both large-scale and classroom level, and the study of teachers' assessment practices. Her publications reflect these two strands: developing and evaluating different strategies to assess students' learning, such as concept maps and students' science notebooks, and studying teachers informal and formal formative assessment practices, such as the use of assessment conversations and embedded assessments. Her recent work focuses on the development and evaluation of assessments that are instructionally sensitive and assessment instruments of formative assessment practices in the classroom.

Peter Seixas

Committee Member

Peter Seixas is professor and Canada Research Chair in the Department of Curriculum and Pedagogy at the University of British Columbia. Dr. Seixas is director of the Centre for the Study of Historical Consciousness, and a member of the Royal Society of Canada. He taught high school social studies in Vancouver over the course of 15 years and earned a PhD in history from the University of California at Los Angeles. He is the author of numerous articles on history education in Canadian and international journals, editor *Theorizing Historical Consciousness* (University of Toronto Press, 2004), and co-editor, with Peter Stearns and Sam Wineburg, of *Knowing, Teaching and Learning History: National and International Perspectives* (NYU Press, 2000). He is director of the pan-Canadian Historical Thinking Project (www.historicalthinking.ca), which aims to promote critical historical literacy through provincial history curricula, textbooks, assessments, and professional development. His current research aims toward the development of robust assessments of historical thinking.

References

Abler, R. F. (1987). What shall we say? To whom shall we speak? *Annals of the Association of American Geographers, 77*(4), 511–524.

Association of American Geographers. (1966). *High School Geography Project: Geography in an Urban Age.* New York, NY: Macmillan.

Baerwald, T. J. (2010). Prospects for geography as an interdisciplinary discipline. *Annals of the Association of American Geographers, 100*(3), 493–501.

Bednarz, S. W., Heffron, S., & Huynh, N. T. (Eds.). (2013). *A road map for 21st century geography education: Geography education research* (A report from the Geography Education Research Committee of the Road Map for 21st Century Geography Education Project). Washington, DC: Association of American Geographers. Retrieved from www.natgeoed.org/roadmap

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7–74.

Black, P., and D. Wiliam. (2004a). Classroom assessment is not (necessarily) formative assessment (and vice-versa). *Yearbook of the National Society for the Study of Education, 103*(2): 183–188.

Black, P., and D. Wiliam. (2004b). The formative purpose: Assessment must first promote learning. *Yearbook of the National Society for the Study of Education, 103*(2): 20–50.

College Board. (2009). *Science college board standards for college success™. Science.* New York, NY: Author.

Fu, A. C., Raizen, S. A., & Shavelson, R. J. (2009). The nation's report card: A vision of large-scale science assessment. *Science, 326,* 1637–1638.

Geography Education National Implementation Project. (1987). *K-6 geography: Themes, key ideas and learning opportunities.* Washington, DC.

Geography Education National Implementation Project. (1989). *Geography in grades 7-12: Themes, key ideas and learning opportunities.* Washington, DC.

Geography Education National Implementation Project. (2012). State of Geography Education in the United States. Unpublished report.

Geography Education Standards Project (1994). *Geography for life: National geography standards 1994.* Washington, DC: National Geographic Research and Exploration.

Goals 2000: Educate America Act. (1994). Public Law No. 103–227.

Hanson, Susan. (2004). Who are "we"? An important question for geography's future. *Annals of the Association of American Geographers, 94*(4): 715–722.

Heffron, S. G., & Downs, R. M. (Eds.). (2012). *Geography for life: National geography standards* (2nd ed.). Washington, DC: National Council for Geographic Education.

Helburn, N. (1998). The High School Geography Project: A retrospective view, *Social Studies, 89.*

Joint Committee on Geographic Education. (1984). *Guidelines for Geographic Education.* Washington DC: Association of American Geographers and National Council for Geographic Education.

Li, M., Ruiz-Primo, M. A., & Shavelson, R. J. (2006). Towards a science achievement framework: The case of TIMSS 1999. In S. J. Howie & T. Plomp (Eds.), *Contexts of learning mathematics and science: Lessons learned from TIMSS* (pp. 291–311). New York, NY: Routledge.

National Assessment Governing Board. (2008). Science framework for the 2009 National Assessment of Educational Progress (NAEP). Retrieved from http://www.nagb.org/publications/frameworks.html

National Assessment Governing Board. (2010). *Geography framework for the 2010 National Assessment of Educational Progress (NAEP).* Retrieved from http://www.nagb.org/publications/frameworks.html

National Center for Education Statistics. (2011). *The nation's report card: Geography 2010.* Washington, DC: Institute of Education Sciences, U.S. Department of Education.

National Center for Education Statistics. (2012). Science in action: Hands-on and interactive computer tasks from the 2009 Science Assessment. Retrieved from http://nces.ed.gov/nationsreportcard/pubs/main2009/2012468.asp

National Commission on Excellence in Education. (1983). *A nation at risk. The imperative for education reform.* Washington, DC: U.S. Government Printing Office.

National Geographic Society. (2011). Geo-literacy coalition responds to 2010 National Assessment of Geography Education with call to action [Press release]. Retrieved from http://press.nationalgeographic.com/2011/07/19/geo-literacy-coalition-responds-to-2010-national-assessment-of-geography-education-with-call-to-action/

National Research Council. (1997). *Rediscovering Geography: New Relevance for Science and Society.* Washington, DC: The National Academies Press.

National Research Council. (2001). *Knowing what students know: The science and design of educational assessment* (J. W. Pellegrino, N. Chudowsky, & R. Glaser, Eds.). Washington, DC: The National Academies Press.

National Research Council. (2010). *Understanding the changing planet: strategic directions for the geographical sciences.* Washington, DC: The National Academies Press.

National Research Council. (2011). *Assessing 21st century skills.* Washington, DC: The National Academies Press.

National Research Council. (2012). *A framework for K–12 science education: Practices, crosscutting concepts, and core ideas.* Committee on a Conceptual Framework for New K–12 Science Education Standards. Board on Science Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

No Child Left Behind Act of 2001, Pub. L. No. 107–110, 115 Stat. 1425 (2002).

Pattison, W. D. (1964). The four traditions of geography. *Journal of Geography, 63,* no. 5, 211–216.

Quellmalz, E. S., Griffin, M., Hurst, K., Kreikemeier, P., Rosenquist, A., & Zalles, D. (2004, April). Integrated performance assessments with technology (IPAT): Design model and prototypes. Presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Ruiz-Primo, M. A., Shavelson, R. J., Hamilton, L., & Klein, S. (2002). On the evaluation of systemic science education reform: Searching for instructional sensitivity. *Journal of Research in Science Teaching, 39*(5): 369–393.

Schell, E. M., Roth, K. J., & Mohan, A. (Eds.). (2013). *A road map for 21st century geography education: Instructional materials and professional development* (A report from the Instructional Materials and Professional Development Committee of the Road Map for 21st Century Geography Education Project). Washington, DC: National Council for Geographic Education. Retrieved from www.natgeoed.org/roadmap

Shavelson, R. J., Young, D. B., Ayala, C. C., Brandon, P. R., Furtak, E. M., Ruiz-Primo, M. A., Tomita, M. K., & Yin, Y. (2008). On the impact of curriculum-embedded formative assessment on learning: A collaboration between curriculum and assessment developers. *Applied Measurement in Education, 21*(4), 295–314.

Taaffe, E. J. (1974). The spatial view in context. *Annals of the Association of American Geographers, 64,* (1): 1–16.

U.S. House of Representatives. (2009). Departments of Transportation and Housing and Urban Development and related agencies appropriations Act, 2010. Conference report to accompany H.R. 3288. Retrieved from www.gpo.gov/fdsys/pkg/CRPT-111hrpt366/pdf/CRPT-111hrpt366.pdf

Virginia Department of Education. (2008). History and social studies standards of learning for Virginia Public Schools. Retrieved from http://www.doe.virginia.gov/testing/sol/standards_docs/history_socialscience/index.shtml

Vygotsky, L. S. (1978). *Mind and society: The development of higher mental processes.* Cambridge, MA: Harvard University Press.

Wiggins, G. P., & McTighe, J. (2005). *Understanding by design.* Association for Supervision and Curriculum Development.