# The Design of Place-Based, Culturally Informed Geoscience Assessment

Emily M. Geraghty Ward,[1,2,a] Steven Semken,[3] and Julie C. Libarkin[2]

## ABSTRACT

We present a mixed-methods approach to community-based assessment design that engages tribal college and university faculty, students, and science educators, as well as experts in cultural knowledge from the Blackfeet and Diné (Navajo) nations. Information from cultural experts, gathered through a combination of sequential surveys and focus group sessions, was analyzed to identify important themes with regard to assessment and geoscience content within the context of these communities. While experts use a variety of assessment approaches in their classrooms, only pre- and posttesting and portfolios were found to be most valuable. Experts indicated that the primary role of assessment was to monitor student progress, steer instruction, and prepare students for success; thus, assessment should be tied to the course goals. Experts differed on their views regarding sources of bias in testing, but overall they agreed that test language and content were both strong sources of bias. They indicated that input on assessment would help to incorporate local context and provide a mechanism for combating bias. Surveys completed by tribal college faculty and Native American students from Blackfeet Community College (BCC) and Arizona State University (ASU) provided information on the themes of geoscience, native science, place, and culture. Participants provided a variety of examples of important geoscience concepts that focused on (1) traditional geoscience concepts (e.g., the composition of Earth materials), (2) Earth system concepts (e.g., the environment and ecosystems), and (3) interactions between native culture and geoscience (e.g., incorporation of native language in science curriculum). Combined, these data offer the basis for developing place-based and culturally informed geoscience assessments by revealing geoscience content that is important to the local community. To aid in assessment design, one-on-one interviews with tribal college faculty and science educators, as well as students from BCC and ASU, provided specific feedback on the question validity of select items from an existing instrument: the Geoscience Concept Inventory (GCI). Emergent themes from the interview transcripts address assessment content, language, and format and reference school science, cultural knowledge, physical places, and connections to the local landscape (e.g., sense of place). Together, these data (1) address the validity of the GCI as a standardized assessment measure in these student populations and (2) provide the basis for developing open-ended assessment questions and concept inventory–like questions that incorporate this feedback. © *2014 National Association of Geoscience Teachers*. [DOI: 10.5408/12-414.1]

*Key words:* Native American, assessment, place-based, culture, validity

## INTRODUCTION

To attain educational equity and help students cross from the culture of everyday life to the culture of the classroom, culturally relevant curriculum, pedagogy, and assessment are necessary (Lee, 1997; Aikenhead and Jegede, 1999; Bang and Medin, 2010). Culturally relevant assessment, in particular, remains an important and mostly unresolved issue in science education research (Lee and Luykx, 2007). While cultural relevance has been incorporated into some course-based assessment (Nelson-Barber and Trumbull, 2007), few standardized and research-based assessments have addressed this issue (Solano-Flores et al., 2002). Standardized assessment works against diversity in science, because it assumes all students have similar access to knowledge and resources (Lee, 2001). Solano-Flores and Nelson-Barber (2001) express concern about the cultural validity of science assessments specifically, arguing that assessment development must consider the sociocultural context of the population to be tested, because this influences interpretation and solution of questions. For Native Americans and other indigenous communities, the fundamental sociocultural context for learning in the geosciences and related natural sciences is place (Cajete, 2000; Deloria and Wildcat, 2001; Aikenhead and Michell, 2011). Given the lack of valid and reliable place-based assessment in the sciences, the purpose of this study is to develop place-based geoscience assessments for the Blackfeet and Diné (Navajo) communities that focus on conceptual understanding.

A place is any locality imbued with meaning by direct or vicarious human experience (Relph, 1976; Tuan, 1977; Gould and White, 1986). Places populate a sociocultural landscape made by humans (Sauer, 1925), which is built upon and interconnected with the physical landscape formed by geologic, hydrologic, atmospheric, and biologic processes. Rich and diverse meanings, both humanistic and scientific, become associated with places, encoding the interplay of environment and culture in them from prehistoric times to the present day. As they make meanings in places, people also form emotional attachments (or aversions) to places.

*Received 28 December 2012; revised 31 May 2013; accepted 1 June 2013; published online 26 February 2014.*

[1]Geology Program, Rocky Mountain College, 1511 Poly Drive, Billings, Montana 59102, USA
[2]Geocognition Research Laboratory, Department of Geological Sciences, Michigan State University, 206 Natural Science, East Lansing, Michigan 48824, USA
[3]School of Earth and Space Exploration, Arizona State University, P.O. Box 871404, Tempe, Arizona 85287-1404, USA
[a]Author to whom correspondence should be addressed. Electronic mail: emily.ward@rocky.edu. Tel.: 406-238-7264. Fax: 406-259-9751

 

The collective set of meanings and attachments held by an individual or a community is called the sense of place (Brandenburg and Carroll, 1995).

In turn, sense of place informs and infuses place-based education, which is characterized by attentiveness to local landscapes; synthesis of place-relevant disciplines, including the natural sciences; teaching and learning out of doors or in similarly authentic settings; community outreach; and promotion of the cultural and environmental sustainability of the places studied (Elder, 1998; Woodhouse and Knapp, 2000; Sobel, 2004; Semken, 2005; Ault, 2008). The sense of place has also been shown to function as an authentic learning outcome and assessment measure for place-based education (Semken and Butler Freeman, 2008; Williams and Semken, 2011; Kudryavtsev et al., 2012).

Native American and other indigenous philosophies and practices of teaching and learning have always been place based: invested with culturally defined biophilic and geophilic place attachment and informed by time-integrated observation of and reflection on Earth system processes, phenology, animal behavior, and human history. Place-based indigenous teachings empower successive generations to thrive communally and self-sufficiently amid the climatic, hydrologic, ecologic, and geomorphic patterns and cycles specific to their homelands (Basso, 1996; Kawagley and Barnhardt, 1999; Cajete, 2000; Aikenhead and Michell, 2011). In contrast, this philosophy was largely absent from mainstream formal education until the emergence of public environmental consciousness and environmentally focused education (Orr, 1992; Sobel, 2004; Ault, 2008; Semken and Brandt, 2010). To encourage greater diversity in the geoscience community, place-based pedagogy is advocated and practiced (e.g., Semken and Morgan, 1997; Riggs and Semken, 2001; Riggs, 2005; Semken, 2005, 2011; Gibson and Puniwai, 2006; Riggs et al., 2007; Palmer et al., 2009; McCoy, 2011; Hammersley et al., 2012; Unsworth et al., 2012) as a means to better recruit, engage, and retain Native American, Native Hawaiian, and Hispanic students, who possess rich, culturally rooted senses of the places studied in these curricula. Advocates of place-based teaching for Native American students typically speak from expertise in indigenous cultures, experience in teaching Native American students, and expressed intentions to help Native American and other underrepresented-minority students gain better access to scientific careers (Semken, 2005). The limited extent to which place plays a role in assessment, particularly in place-intensive domains such as geoscience, suggests significant gaps between needed and available assessments.

Assessment can serve many purposes. Within the context of a specific course or informal experience, collection of data prior to instruction can inform educators about student needs and abilities, whereas assessment after instruction provides instructors with evidence of student learning. Assessment can also be conducted for research purposes, generating data that argue for or against the use of specific instructional practices. Course-based assessment in the geosciences engages students in an array of assessment activities, from traditional, close-ended exams to field mapping and independent research. Such course-based assessment offers rich insight into student thinking.

Research-based assessment in the geosciences, typically collection of data for the purposes of characterizing a population or evaluating the impacts of a program, tends to focus primarily on conceptual understanding (e.g., Libarkin et al., 2011, and references therein), although some effort has been made to evaluate the impact of instruction on affective variables (e.g., Libarkin, 2001; Jolley et al., 2012). Most instruments used for geoscience assessment are founded on the ideal of standardized assessments that can be used across populations. Standardization at its most ideal is as free of bias as possible and aims to produce assessments that are acultural and globally applicable. This standardization can be invaluable for comparing learning across normative settings, such as has occurred with the Geoscience Concept Inventory (GCI) (e.g., Petcovic and Ruhf, 2008; Teed and Slattery, 2011). At the same time, the value of assessment is limited when individuals have experiences and cultural settings that are significantly different from the normative settings in which assessments are developed (Solano-Flores and Nelson-Barber, 2001; Demmert, 2005; Lee and Luykx, 2007; Luykx et al., 2007; Nelson-Barber and Trumbull, 2007; Coles-Ritchie and Charles, 2011). In the context of geoscience, assessments are often limited by their focus on overarching concepts rather than place-specific constructs. This inhibits the ability to assess learning, particularly for those communities where place plays a significant role in world view, such as has been suggested for the Blackfeet and Diné.

The purpose of this study is to identify appropriate place-based content and language that can be incorporated into the design of new geoscience assessments. This paper presents the results from a mixed-methods approach to assessment design that engages tribal college and university faculty, students, and science educators, as well as experts in cultural knowledge from the Blackfeet and Diné nations (Fig. 1). The study employed a two-stage research process: (1) experts, faculty, and students identify effective assessments and relevant geoscience content for their communities, and (2) data collected from experts, faculty, and students are synthesized to collaboratively construct valid and reliable place-based assessments. This paper focuses on the methods used to collect the baseline data for assessment design, which we have termed Stage 1.

## Research Participants

Two groups of participants were included in this study (Table I). The first group consisted of nine cultural experts from the Blackfeet ($n = 7$; four female, three male) and Diné ($n = 2$; male) communities. These individuals were selected because they have an understanding of the history, language, and traditional knowledge of Blackfeet and Diné cultures, are involved with formal, informal or both types of science education in these communities, or both. Cultural experts included Blackfeet and Diné studies and language scholars, science educators and education researchers from informal science and outreach settings, an ethnobotanist, and a K–12 educator. The expertise among the cultural experts was deliberately diverse and allowed feedback on geoscience assessment from multiple perspectives, including the value of assessment, the important geoscience concepts, and the influence of place in the geosciences.

The second group of participants consisted of faculty ($n = 4$) and students ($n = 4$) from Blackfeet Community College (BCC) and Native American students from Arizona State University (ASU) ($n = 15$) and provided information on

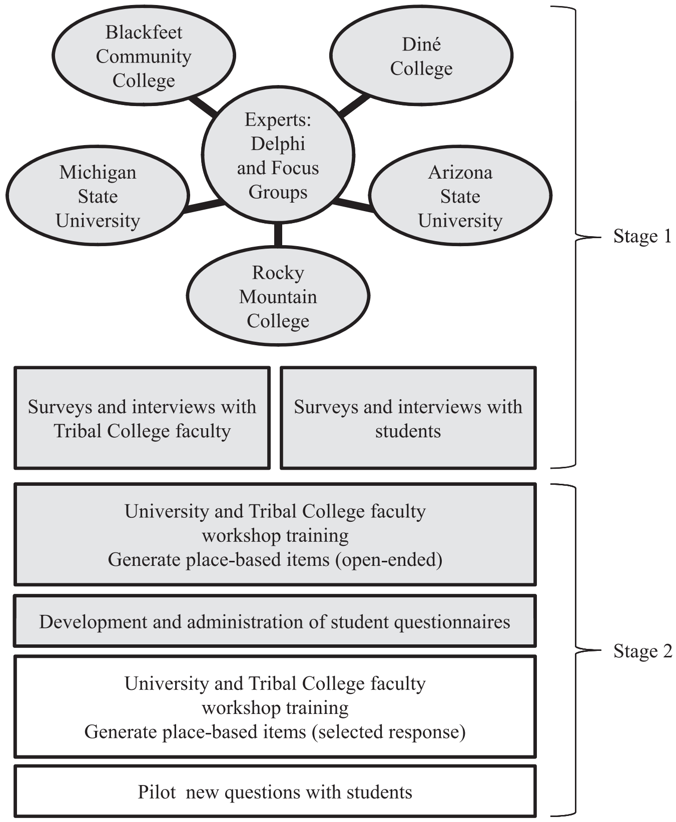


**FIGURE 1: Conceptual diagram of research design. Stage 1 of assessment development included administration of Delphi surveys and facilitation of focus group sessions with cultural experts. Faculty and student surveys and one-on-one interviews were also part of Stage 1. These data guided the development of place-based open-ended items for student questionnaires (early Stage 2). Research activities that are complete are in gray.**

ideas of culture, place, native science, and geoscience. BCC faculty are defined here as both professors (one of Blackfeet Studies and one of Natural Resources) and science educators affiliated with the college (one American Indian Science and Engineering Society chapter adviser–learning center coordinator and one science instructor at the alternative high school). BCC and ASU students included both science and nonscience majors, as well as nontraditional students, as indicated by the standard deviation for average age (Table I). These same faculty and students provided feedback on select items from the GCI that addressed the content, language, and format of the items.

Equal numbers of total female and male participants make up experts, faculty, and students. A large percentage of participants had completed biology coursework, and at least half had previously taken geology or Earth Science classes (Table I). All but one of the BCC participants were affiliated with the Blackfeet tribe of northwest Montana and southern Canada, while multiple tribes of the southern Rocky Mountains and the Southwest were represented by ASU students (Fig. 2).

**TABLE I: Age, gender, academic training in science, and tribal affiliation for the faculty and students involved in the assessment validation.**

| | Expert Group ($n = 9$) | BCC Faculty ($n = 4$) | BCC Students ($n = 4$)[1] | ASU Students ($n = 15$)[2] |
|---|---|---|---|---|
| Average age (y) | 45 (±10) | 56 (±5) | 29 (±11) | 28 (±8) |
| Gender | 4 female, 5 male | 1 female, 3 male | 2 female, 2 male | 9 female, 6 male |
| Students with science coursework completed in high school or college | | Biology ($n = 4$), physics ($n = 1$), geology ($n = 3$), chemistry ($n = 2$) | Biology ($n = 4$), physics ($n = 3$), geology ($n = 2$), chemistry ($n = 2$) | Biology ($n = 14$), physics ($n = 9$), geology ($n = 11$), chemistry ($n = 9$) |
| Tribal affiliation | Blackfeet ($n = 7$), Diné ($n = 2$) | Blackfeet ($n = 4$) | Blackfeet ($n = 3$) | Assiniboine, Diné, ($n = 8$), Pascua Yaqui, White Mountain Apache ($n = 2$), Salt River Pima-Maricopa, San Carlos Apache, Chemehuevi |

[1]One student is not affiliated with a tribe.
[2]Tribal affiliation for ASU students is $n = 1$ unless otherwise indicated.


**FIGURE 2: Location of participating institutions and hometown zip codes of participants (map: Hannah Clark; projection: Lambert Conformal Conic).**

## METHODS

A mixed-methods approach was chosen to elicit expert, faculty, and student views on assessment and geoscience to use as baseline data in the design of new place-based geoscience assessments. The study required that cultural experts complete sequential surveys and participate in a focus group to identify important themes with regard to assessment and geoscience content within the context of the Blackfeet and Diné communities. Faculty, science educators, and students from BCC and ASU completed a different survey that contained both open-ended and Likert-scale items. Among the survey questions, participants were asked to define the terms *geoscience, native science, place,* and *culture* and to articulate those geoscience concepts that were important for students to understand. Survey data analysis identified points of convergence and divergence regarding participant views on assessment and geoscience content within the context of these two communities. Transcripts from one-on-one interviews, conducted with faculty and students to review select GCI items, were coded thematically, with codes emerging from the interview data. The feedback collected from the interviews addressed the construct and communication validity of this assessment instrument.

## 2-Tier Delphi Instrument: Blackfeet Expert Group

In this study, open-ended and Likert-scale items related to assessment, native science, geoscience, and meanings of place and culture were collected from the Blackfeet expert group. The Delphi method was used to identify convergence of opinions within an expert group and seek opinions from experts through sequential surveys that approach opinion consensus (Ziglio, 1996; Hasson et al., 2000; Rowe and White, 2001; Osborne et al., 2003; Okoli and Pawlowski, 2004; Hsu and Sandford, 2007). Participants responded to an initial 20-item survey (Round 1) containing both open-ended items ($n = 12$) and closed-response items ($n = 8$). These were then summarized and served as the basis for the response options in a close-ended second survey (Round 2). As a consequence, open-ended items used in Round 1 were presented as Likert-scale items in Round 2. For example, an open-ended item in Round 1 of the Delphi was implemented as "List the things that you think might make assessments unreliable or invalid." Based on participant responses, this item was changed to a Likert-scale item in Round 2: "Please select the level to which you agree with the following as a source of bias in testing of native students: (a) Lack of input from instructors, students, other people involved in instruction; (b) Lack of instructional resources; (c) Lack of knowledge of instructor; (d) Inconsistency (e.g., given a post assessment, with no preassessment given to compare); (e) Stress of student/test anxiety; (f) Lack of context (e.g., not based on the particular students/content/environment being assessed); (g) Gender."

In Round 2 of the Delphi, participants were presented with eight Likert-scale items, along with a summary of responses from Round 1. Participants were asked to respond to the items in Round 2, providing rationale and additional comments for their responses as needed. Convergence on common themes was determined by identifying the highest-ranking mean response value for each item in Round 2.

## Focus Group

Content and questions from the Delphi survey were used as the initial foundation for the focus group sessions with both Diné ($n = 2$) and Blackfeet ($n = 7$) expert groups. Focus groups served different purposes for the two populations. While Blackfeet participants completed surveys as described earlier, items from Round 1 of the Delphi were administered orally to the Diné participants during the focus group in alignment with a cultural practice of oral questioning. Diné participant responses to the Delphi items were audio recorded and transcribed, with note-taking during the focus group helping to focus subsequent analysis. The focus group discussion conducted with Blackfeet experts was guided by the themes of the Delphi survey: assessment, locally relevant geoscience content, native science, and current geoscience issues affecting the Blackfeet community. Responses from the focus group were summarized in handwritten notes, transcribed, and then reviewed by the expert group to verify the information collected (Bertrand et al., 1992; Kitzinger, 1995).

## Survey Instrument: Faculty and Students

Faculty and student participants from BCC and ASU were asked to describe what the terms *culture, place, geoscience,* and *native science* meant to them. Surveys also asked participants to articulate those geoscience concepts that were important for students to understand. Surveys were administered prior to the one-on-one interviews. Thematic content analysis (the process of identifying patterns or themes in qualitative data) of these open-ended responses revealed common themes that (1) provide evidence for consensus among our participants to building culturally relevant assessments and (2) will shape the design of new geoscience assessments that incorporate this cultural perspective.

## Interviews: Faculty and Students

A subset of items from the GCI (Libarkin and Anderson, 2005) was selected for validation during interviews with tribal college faculty and students ($n = 23$). Eight GCI items[1] referencing topics of glaciers, wind, mountains, clouds, fossils, volcanoes, rocks, and minerals were chosen for review (Table II). Items were selected based on how well the item content related to the geologic features present in the northern and southern Rocky Mountains and the Southwest (here used to encompass the Colorado Plateau, Arizona Transition Zone, and Basin and Range Province). For example, glaciers, mountains, wind, and clouds are familiar to residents of the Blackfeet Reservation, which is located in the northern Rocky Mountains of Montana and adjacent to Glacier National Park. Similarly, mountains, volcanoes, wind, and clouds are familiar to the traditional homelands for tribes in the Southwestern deserts. During the interviews, participants were asked to comment on their comprehension of the item; their frames of reference, reasoning, and judgment about the concept; the importance of the concept (construct validity); any personal meaning or experience surrounding the concept; any meaning the concept might have to their peers and community (cultural validity); and the question format. Participants read each item aloud and were asked a series of interview questions to elicit discussion about concepts and to assess item validity for each participant (see the Appendix).

Themes in the interviews were identified through content analysis of the transcripts. The common themes identified then became the categories for content analysis of the transcripts. The initial coding of transcripts focused on the following categories: school science (content and language common in a science classroom), physical landscape, place (defined here as a cultural connection with the landscape), and culture. A coding rubric was constructed to draw out these themes with respect to assessment content and language. An additional "format" code was included in the rubric to identify instances in which participants addressed the construction of the item and response options. Two researchers (the lead and second author of this paper) coded three interviews to establish interrater reliability. These authors have extensive backgrounds working with these Native American communities. The second author has a long history working with tribes of the southwestern U.S., having taught geologic sciences at a tribal college on the

[1]Not all questions were used in every interview, depending on the length of the interview (~1 h).

**TABLE II: Selected GCI items for validation by faculty and students. Numbers reported indicate the number of participants that responded to each item.**

| GCI Item | BCC Participants ($n = 8$) | ASU Participants ($n = 15$) |
|---|---|---|
| Where do you think glaciers can be found today? | 6 | 1 |
| Which of the following can be caused by wind? | 6 | 14 |
| During a recent trip to Canada, a traveler visited two mountains made up of the same type of rock. The sketches below represent the outlines of these two mountains. Which of the following reasons best explains the differences in the two drawings? | 5 | 1 |
| What is the connection between clouds and rain? | 3 | 13 |
| Are rocks and minerals alive? | 3 | 15 |
| Which of the following best describes mountains? | 1 | 10 |
| Fossils are studied by scientists interested in learning about the past. Which of the following can become fossils? | 1 | 1 |
| On continents, where does most volcanic material come from? | 0 | 1 |

Diné Nation for 15 y prior to moving to ASU. The first author has collaborated with BCC staff and faculty for 10 y on various professional development and informal science projects. With this experience, the raters coded the interview transcripts for indications of culture and place within the context of these communities.

Initial disagreement between the two coders regarding the content, format, and language were reconciled by narrowing the definitions of these codes so that *format* pertained to the structure of the item (e.g., the number of response options) and *language* referred to specific terms. Disagreements between culture and place coding prompted clarification that the content code of *place* be applied to participant descriptions of any activity (including a cultural practice) that made specific reference to one or more named places or physical localities, whereas the content code of *culture* be applied to references to all other cultural practices. Interrater reliability of the rubric between the two raters yielded initial agreement of 86% on hand-recorded transcripts and 70% agreement on transcripts generated from audio recordings. After addressing areas of disagreement, a fourth transcript from the audio-recorded interviews was coded and interrater reliability reached 79%. In general, reliability values from 75% to 90% are considered an acceptable level of agreement (Graham et al., 2012). With interrater reliability values within acceptable range, the remaining interviews were coded by both raters. One rater coded the remaining 13 hand-recorded transcripts, the other coded the remaining 6 audio-recorded transcripts.

### Validity in Instrument Design

The survey instruments and interview protocol were externally evaluated for communication and cultural validity (cf. Solano-Flores and Nelson-Barber, 2001) and revised where appropriate before administration to study participants. The external evaluators were chosen based on their experience with science education projects in Native American communities. The three-person external evaluation team was composed of the diversity director for the National Center of Earth System Dynamics at the University of Minnesota and two Fond du Lac tribal members involved in science education. The tasks of the external evaluators were to (1) provide feedback on the overarching project plan and (2) review validation and professional development methods for this study (e.g., surveys, interview protocols, etc.). Content for Likert-scale items in the Round 1 Delphi were derived from the published literature to speak to content and construct validity of the questionnaire items (Williams and Vaske, 2003; Wiggins and McTighe, 2005; Nelson-Barber and Trumbull, 2007; ESLI, 2009). Notes from the focus group with experts were summarized in a typewritten report and shared with focus group participants. Participants were encouraged to provide feedback to ensure the validity of information recorded during note-taking. Interview probes incorporate elements of published protocols (Libarkin, 2008; Solano-Flores and Li, 2009), targeting communication and construct validity in a cultural context. These probes, together with those eliciting personal meaning and experience with the question content, allowed interviewees opportunities to validate the assessment instrument within the cultural context. Participants were also given access to a public blog that documented the activities of the project (http://geoedplaces.blogspot.com) and were encouraged to review and comment on the blog.

## DATA

### Experts: Delphi Survey and Focus Groups

#### *Assessment: Use and Value*

The majority of experts (eight of nine) indicated that assessment should be used before, during, and after instruction. More than half of the Blackfeet and Diné experts indicated that they use pre- and posttesting (seven of nine) and multiple-choice questions (six of nine) to assess their students. About half (five of nine) used creative works, essay questions, portfolios, and oral exams for assessment, whereas less than half used written reports (nine of nine) and short-answer and standardized tests (both two of nine). Experts indicated that pre- and posttesting and portfolios are the most valuable types of assessment, in sharp contrast to standardized tests (Fig. 3).

The list of assessment types on the Delphi survey was viewed as incomplete by Diné experts. In particular, oral exams, programmatic rubrics, and community-based assessment were considered important assessment practices in Diné courses. For example, one interviewee indicated that Diné instructors use a broad range of rubric-based assess-

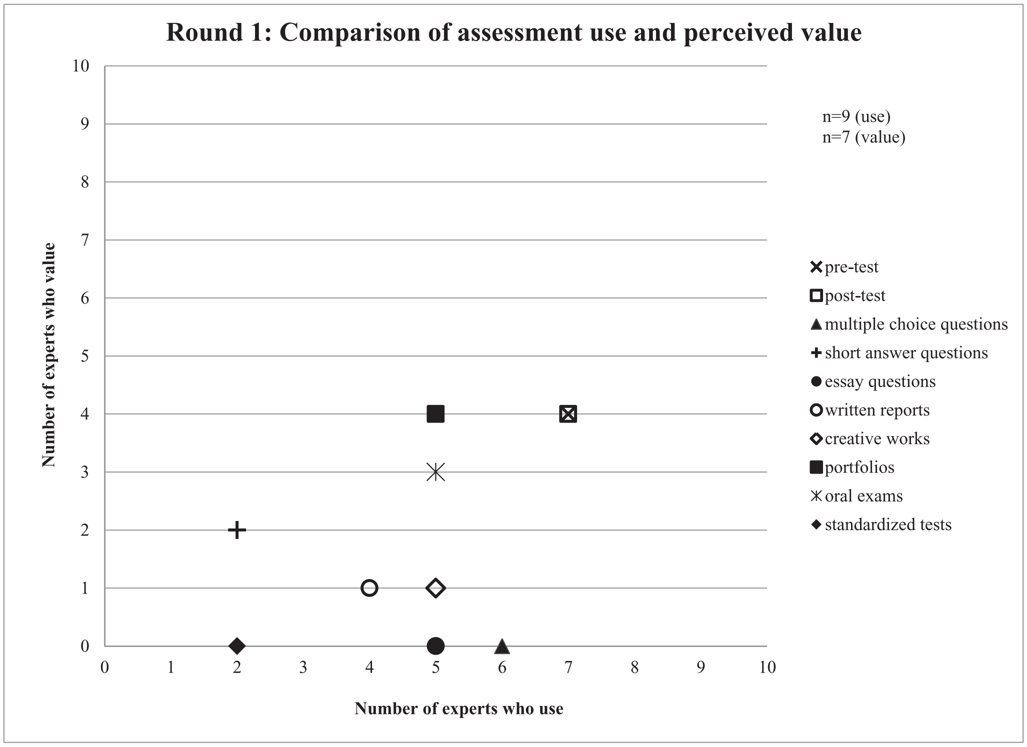


**FIGURE 3: In Round 1 of the Delphi, experts indicated that they use a variety of assessment approaches but only found a few to be "most valuable." The majority of experts used and valued pre- and posttest assessment. Though multiple-choice tests were used by many, they were perceived to have little value by the experts. (Use $n = 9$, value $n = 7$; Diné experts only commented on use of assessment.)**

ments in which rubrics are typically aligned to degree programs rather than course outcomes. Furthermore, many assessments are administered orally owing to the Diné oral tradition. Unlike traditional assessment, Diné instructors often used community-based assessment, in which the students survey family and friends using specified procedures. This expert feedback provided context on use and style of assessment in these communities, incorporating cultural traditions into classroom assessment.

Blackfeet participants also identified valued assessments not initially considered by the research team. In Round 2 of the Delphi, Blackfeet experts continued to rank pre- and posttesting as highly valuable (4.7 and 4.5, respectively, on a 5-point scale) but added reflection (4.8) and participant observation (4.3) as valuable types of assessment (Fig. 4). Both reflection and participant observation were listed by Blackfeet experts in Round 1 of the Delphi as additional types of assessment that they used in their classrooms. Blackfeet experts noted different contextual information that spoke to the use and value of assessment:

*Pre and post are valuable, but sometimes I feel you need someone solely responsible for administration and analyzing.*

*I tend to feel like oral exams are not accurate for some kids, as they may not be able to express their understanding orally. Written reports and portfolios are more likely to assess a student's understanding of a topic, although these types of assessments are more time consuming and not always possible.*

### *Assessment: Role*

Expert ideas about the role of assessment in curriculum development focused on four primary themes: student knowledge, instructional goals, student progress and growth, and informing instruction (Table III). With regard to student knowledge, progress, and growth, Blackfeet experts responded that they used assessment to determine what content students might need help with, when students have learned the material, and what content might need to be reviewed again to facilitate student progress. In terms of instruction, Blackfeet experts indicated that the assessment should be tied to instructional goals and embedded in curriculum development. Diné experts spoke similarly about the links among goals, assessment, and curriculum design, speaking about the importance of clear objectives and goals. They mentioned that "lesson planning is based on these and may expand on the initial set of objectives" and assessment is then used "to see if the course has met the goals."

In reference to the role of assessment in instruction, experts focused on the following themes: student progress and success, informing instruction, and student reflection (Table III). With regard to student progress and success, Blackfeet experts indicated that assessment can measure student success and should be used to determine whether and how students understand the information, which is useful to both instructor and student. Assessment may also be used to steer instruction and to help instructors know when their students are ready to move on to the next topic.

Expert discourse about the role of assessment in student learning focused on three themes: student preparation and

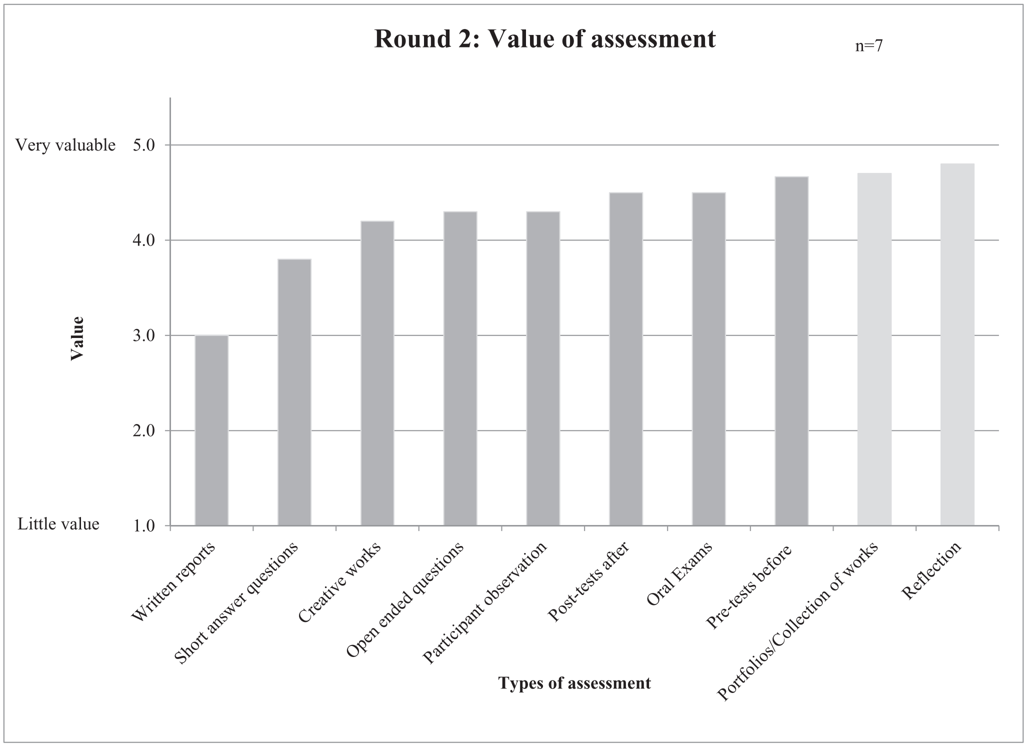


**FIGURE 4: When asked to rank the value of assessment (1 = of little value, 5 = very valuable), Blackfeet experts continue to rank pre- and posttesting high but added reflection and participant observation as valuable types of assessment (light gray columns) ($n$ = 7).**

success, student reflection, and adapting instruction and content to facilitate student learning (Table III). One expert mentioned that assessment "sets the stage" for student learning and helps instructors tailor their teaching for their students. Others mentioned that assessment is useful for students in that it provides opportunities for students to reflect on their learning and progress.

These data were used to generate a ranking instrument implemented in Round 2 of the Delphi. Blackfeet experts indicated the most important roles for assessment in terms of curriculum development, instruction, and student learning. The expert group consistently ranked three themes highly on a 5-point scale (Fig. 5). The most important roles of assessment, according to the Blackfeet experts, were to direct instruction (4.8), monitor student learning (4.8), and help students reflect on their learning (5). Note the shift in support from inform instruction (two of nine in Round 1) to direct instruction (average rank 4.8 of 5 in Round 2), providing an example of how expert feedback helps guide consensus using the Delphi technique.

Specific discourse about the role of assessment provides additional nuanced understanding of expert conceptualization of the role and value of assessment. For example, one Blackfeet expert offered a view that illustrated the poor fit of many standardized tests to curriculum:

> *My experience with standardized tests is that the information/questions being given doesn't always pertain to all students; it is sometimes, too often, too broad and doesn't connect to what really needs to be assessed. All students are different.*

On student learning, the same expert noted,

**TABLE III: Expert views on the role of assessment in curriculum development, instruction, and student learning ($n$ = 9). Numbers reflect the number of respondents to mention these themes regarding the role of assessment.**

| Curriculum Development[1] | | Instruction | | Student Learning | |
|---|---|---|---|---|---|
| Theme | $n$ | Theme | $n$ | Theme | $n$ |
| Tie to goals | 4 | Monitor student progress and success | 3 | Prepare students for success | 3 |
| Student knowledge | 3 | Steer instruction | 3 | Metacognitive | 2 |
| Monitor student progress and growth | 3 | Formative and metacognitive | 1 | Tailor to meet the needs of students | 2 |
| Inform instruction | 2 | | | | |

[1]Diné experts ($n$ = 2) only commented on the use of assessment with regard to curriculum development. Some experts listed more than one theme in their answer.

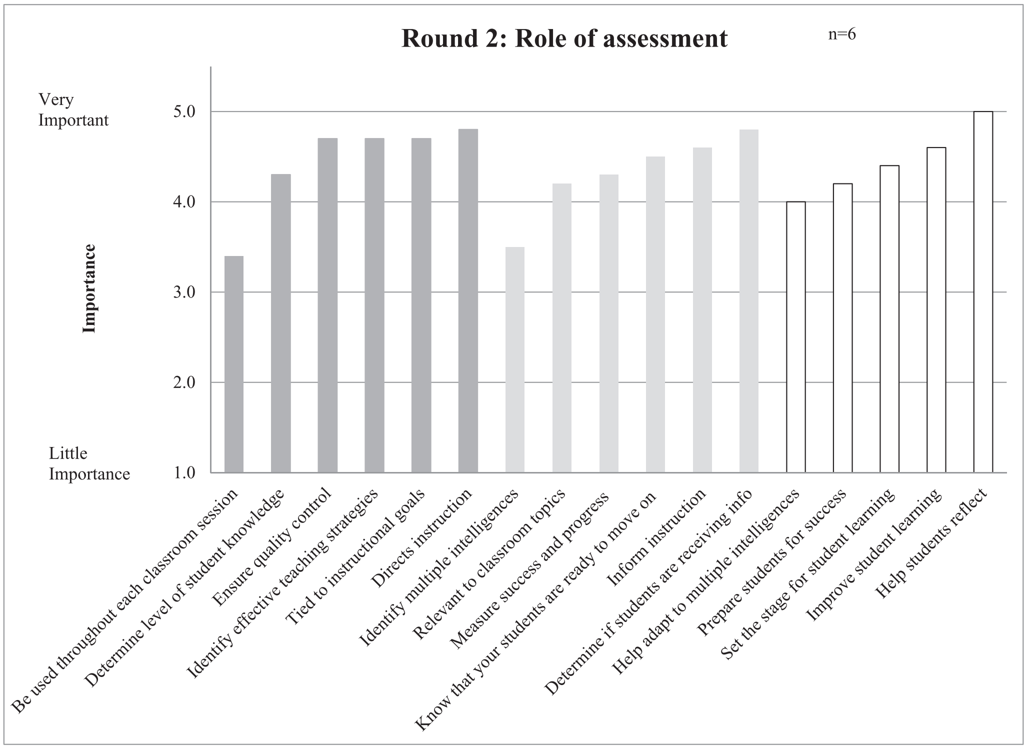


**FIGURE 5: When asked to rank the importance of the role of assessment in the areas of curriculum development (dark gray columns), instruction (light gray columns), and student learning (white columns), Blackfeet experts agreed that the role of assessment is to help inform instruction, gauge student understanding, and help students reflect on their learning. Rankings were based on a 5-point scale (1 = not important, 5 = very important) ($n$ = 6). One expert did not answer the question.**

> *It is great to set the stage for what a child/participant needs to learn based on his or her own needs. I think a culture—not necessarily Native American culture but just the culture they are growing up in—needs to be considered.*

### *Assessment: Validity and Reliability*

Blackfeet and Diné experts listed a variety of factors that might make assessment invalid or unreliable in Round 1 of the Delphi. These included the language used and style of assessment, insufficient knowledge or resources, disregard of educational context, inconsistency in conducting assessment, and stress. Gender was also noted as a possible source of bias, although the exact nature of this bias was unclear. In the second round, Blackfeet experts ranked the level to which they agreed with these sources of bias on a 5-point scale (1 = strongly disagree, 5 = strongly agree) (Fig. 6). Sources of bias identified by participants were then categorized according to published sources of assessment bias (Nelson-Barber and Trumbull, 2007). Lack of input, referring to student input, as well as input from faculty, peers, and the community, ranked highest on the list of sources of bias (4.8), followed closely by inconsistency (4.7) and lack of context (4.7). One Blackfeet expert noted that assessment should look for "buy-in" on the part of the students to encourage more effort on their behalf. The buy-in may be closely tied to context in this case, with another Blackfeet expert noting that assessment "has to serve a purpose, mean something to the learner." Inconsistency referred to test administration, with a Blackfeet expert noting that inconsistency in test administration can cause students stress.

Diné experts mention similar assessment bias with regard to test language and content in Round 1 of the Delphi. One expert noted:

> *Many existing references used in teaching were created by non-Navajos, but unless assessment is designed by a Navajo, it is not relevant to a Navajo. . . . A major issue is the use of third person by non-Navajos versus first person by Navajos. Teaching about "my" culture is more effective than teaching about "Navajo" culture.*

Another Diné expert emphasized the importance of context. He described a textbook for Diné language that was designed by a nonnative speaker and contained acultural questions and scenarios in the third person, rather than in the more Diné-appropriate first person. The expert indicated that this would make it difficult for a student to understand, interpret, and answer questions. Rather, he emphasized that the use of the present tense and the inclusion of practical and relevant phrases that students could use immediately were important for engagement of Diné students.

In rating the sources of test bias identified by Nelson-Barber and Trumbull (2007) (5-point scale; 1 = strongly disagree, 5 = strongly agree) experts agreed that test use and format were sources of bias in Round 1 of the Delphi survey (3.8 and 3.7, respectively) (Fig. 7). By Round 2 experts shifted their rankings to agree that test language (3.7) and test

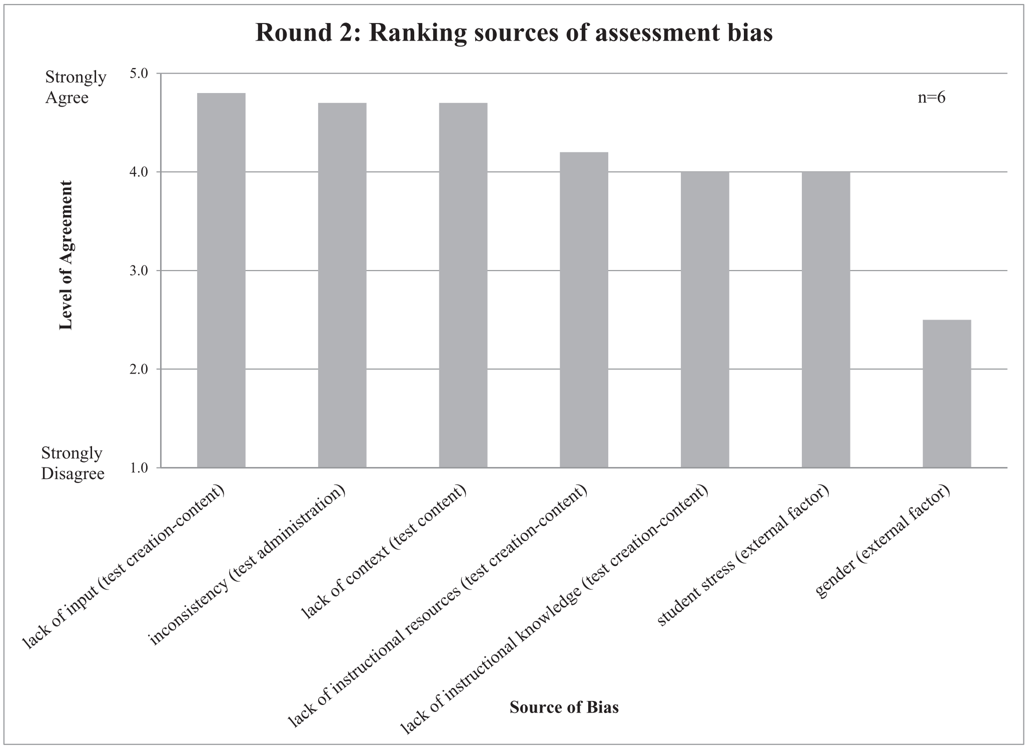


FIGURE 6: Rank-order list of sources of bias in assessment according to Blackfeet experts. In Round 2, experts ranked the list by the level to which they agreed with the source of bias (1 = strongly disagree, 5 = strongly agree). Sources were then categorized according to published sources of bias (Nelson-Barber and Trumbull, 2007). Experts suggest that bias occurs during the stages of test creation, as well as in administration. External factors refer to factors that are not inherent to the assessment but rather are factors of the test taker ($n = 6$). One expert did not answer the question.

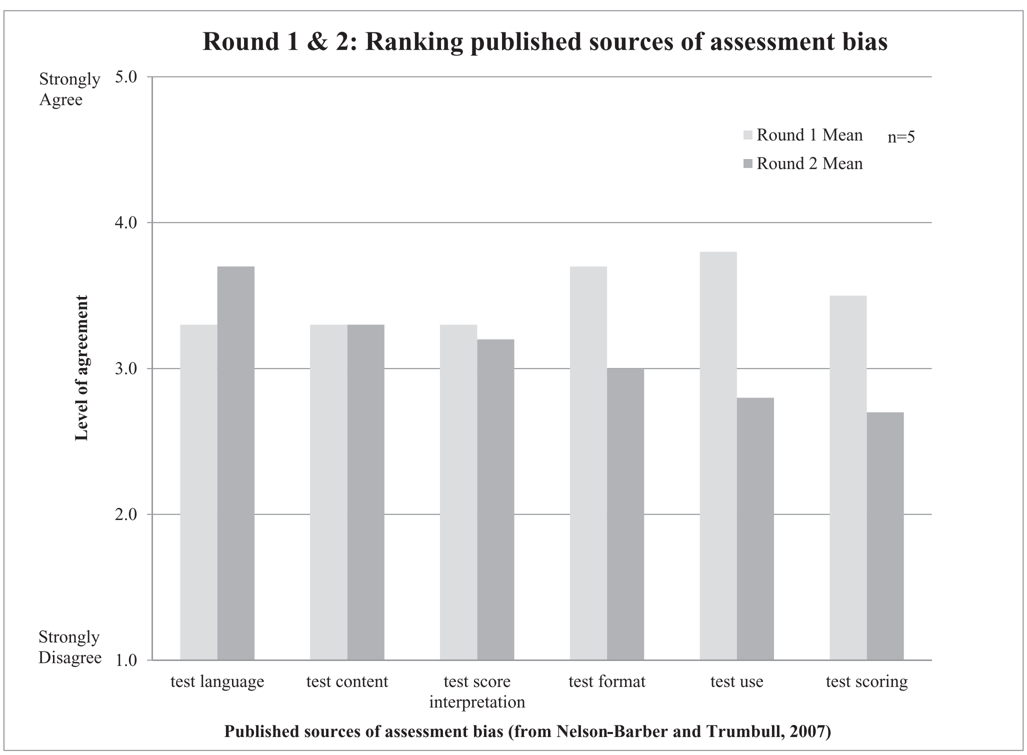


FIGURE 7: Expert rankings of published sources of bias in assessment (Nelson-Barber and Trumbull, 2007) in the 2-tier Delphi (1 = strongly disagree, 5 = strongly agree). Values reported are mean expert rankings ($n = 5$). Some experts did not respond to the question.

content (3.3) were primary sources of test bias (Fig. 7). Test language encompassed both academic language (e.g., technical language) that is different from the "everyday English" students understand and language proficiency of English-language learners compared with native English speakers (Nelson-Barber and Trumbull, 2007). These rankings are in agreement with those bias sources identified by the participants in open-ended responses in Round 1 of the Delphi (Fig. 6). Experts mentioned test design (lack of input and context or language) and inconsistent test administration as sources of bias (Fig. 6).

## Experts, Faculty, and Students: Surveys

### *Culture, Place, and Native Science*

Experts, faculty, and students were asked to describe what the terms *culture, place,* and *native science* meant to them. Overall, participants agreed that *culture* is a set of values, beliefs, and practices that are shared by a community of people, with such practices including language and traditions. *Place* meanings varied from descriptions of the physical landscape to personal connections such as "home" and "identity." One Diné expert described place as a "strong connection to where you are born, where your umbilical cord is buried, where your grandparents live, where your childhood experiences were, laughter, joy, where your wealth [e.g., livestock] is based. If you leave and then return home, you relive those connections over again."

Culture and place both played prominent roles in participant views of native science, which both Blackfeet and Diné experts explain as emphasizing the interconnectedness of natural features and phenomena, and cyclic processes, largely concordant with the Western Earth-system science perspective (e.g., Semken and Morgan, 1997; Cajete, 2000). One Blackfeet expert commented that native science incorporates "native knowledge to make science more clear and relevant," while a Diné expert explained that the Diné scientific method emphasizes time-integrated observation of natural phenomena over sampling and analysis. Native scientists need to travel to the subject of study (i.e., "go to the mountain") and reside there to conduct research. In addition, this participant articulated a nuanced view of science that values epistemology: Diné scientists do not just consider the physical but also the metaphysical role of thought. Language is such an important part of native science such that "the terms that Diné use to describe the universe emerge from our language. Language is perception." (Diné expert citing Benjamin Whorf, cited in Carroll, 1964).

Other participants articulated commonalities between native science and Western science, commenting that "Native science is a natural study of observing and understanding the relationships within the natural universe and participating in the process" (Blackfeet expert) and "The studies of the Earth's surroundings pertinent or relevant to a particular culture." (BCC student). Another participant thought the distinction between native and Western science was artificial, noting "I think it's a foreign concept because there is no separation. Science is the study of life, living entities, and that is interwoven into tribal practices. As you look at native science, you are looking at traditional knowledge of plants, land—again, it ties into resources." (ASU student). These and similar comments provided an understanding of the lenses through which Diné, Blackfeet, and other Native American students and faculty situate geoscience within the larger context of culture and place. By finding commonalities among comments and understanding the way in which overarching themes emerge from native discourse, we begin to identify meaningful concepts that can be incorporated into new assessment.

### *Geoscience Content and Instruction*

Experts, faculty, and students were asked to provide geoscience concepts that were considered important for understanding the Earth from their perspective. Overall, we recognized concepts from three categories that focused on (1) traditional geoscience concepts, (2) Earth system concepts (connections between geoscience and other science disciplines), and (3) interactions between native culture and geoscience. Several themes that aligned with traditional geoscience instruction include respect for Earth, rocks and minerals, maps and the physical landscape, natural resources (e.g., water and petroleum), geologic features (e.g., rivers) and processes (e.g., erosion and deposition), and technology used to study Earth (e.g., global positioning systems and geographic information systems). Earth system concepts included the seasons, environment, weather (e.g., wind and rain) and climate, and ecosystems (e.g., relationship between water in drainage basins and fish populations). Other discourse focused on concepts closely aligned with a cultural perspective, such as traditional elements (life, land, water, air, and fire), sustainability and utility of the land, artifacts, and cultural relevance of concepts to the place where one lives. The cross-disciplinary nature of these concepts, both in terms of multiple sciences and incorporation of cultural references, suggest the importance of an Earth-system science framework for any instruction or assessment within these communities (Semken and Morgan, 1997; Barnhardt and Kawagley, 2005; Aikenhead and Michell, 2011). The important geoscience topics identified by experts, faculty, and students address the construct validity of the geoscience content selected for use in assessment. Construct validity focuses on whether strong support for the content of the items exists (Libarkin, 2008).

## Faculty and Students: Interviews

One-on-one interviews ($n = 23$) with BCC faculty and students and with ASU students were undertaken to consider the extent to which GCI items were applicable in these Native American contexts. Participants were asked to comment on question content, language, and format. Thematic content analysis of their responses by two raters identified themes related to school science, landscape, place, and culture. Concepts that aligned with school science included those typically addressed in the classroom, such as plate tectonics, earthquakes, and climate change. Landscape themes emerged when participants mentioned specific physical locations that were often locally or regionally situated, such as a nearby mountain range or a local outdoor setting. Place and culture themes were identified and distinguished according to the criteria described earlier. Raters coded interviews to highlight instances in which participants mentioned any of the four aforementioned themes.

Data are reported as a percentage of the total number of instances that each of these themes appeared in participant interviews (Fig. 8). Overall, participants tended to focus their

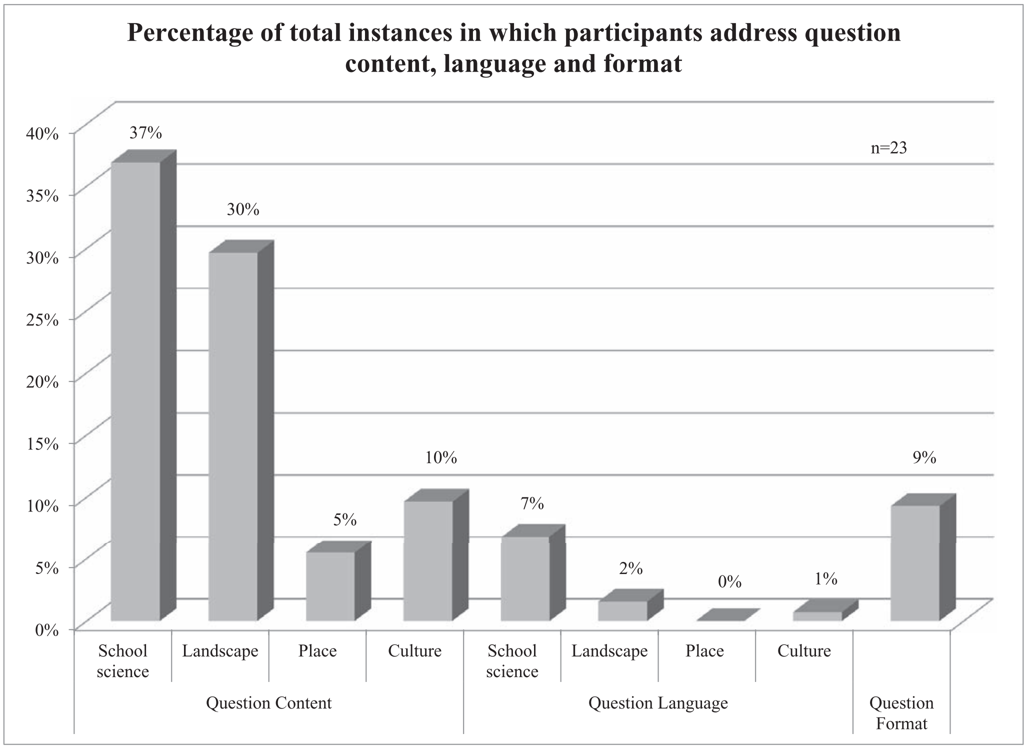


**FIGURE 8: BCC and ASU participants in interviews ($n = 23$) discussed question content more than question language and format overall. Participants focused primarily on the school science content of the GCI questions and often related the question content to the local landscape. The roles of place and culture were secondary focuses in participant discourse with regard to question content.**

feedback on question content rather than question language. Participants had suggestions on changes to question format that typically addressed the number of response options and the use of figures with question text. The data reveal that, by far, participants focused on question content with regard to school science (37%) and local landscape (30%). Content referencing culture (10%) and place (5%) was addressed less frequently by interview participants.

In general, participants referenced what they saw and experienced every day when they engaged with each item. Some participants indicated that the content was important for them because it related to a particular physical feature or landscape of cultural significance. Others spoke of the connections they saw between content of the question and significant issues facing their communities, such as the availability and quality of water or soil resources. Participants spoke of knowledge they gleaned from elders and parents and indicated that people might respond differently to the item depending on their age. Some noted that members of the older generation may invoke cultural knowledge when interpreting or responding to an item.

Participants identified technical language (e.g., *plate tectonics, erosion,* and *mountain building*) as terms that might keep students from fully understanding the question and response options. This may be especially true for nonscience majors who have had minimal exposure to geologic terminology and for nontraditional students who may be returning to school after many years away from formal education. Participants commented on the format of the questions, indicating that some response options "all sounded the same" and made it seem like the test was trying to "trick" the test taker into getting it wrong. Others mentioned that the response options did not contain responses that they would use to answer the question. Overall, participants were split on how they felt about multiple-choice questions versus open-response formats: some indicated that they liked being able to work through the possible response options to determine the correct answer, whereas others preferred to express what they know through essays.

### *Statistical Analyses: Question Content and Language*

Correlation analyses were performed to determine the relationship among specific interview codes identified earlier and demographic information for the interviewees. The correlations were conducted for all content, language, and format codes across all questions for each interview participant ($n = 23$). Statistically significant correlations between how often participants addressed question content related to school science and physical landscape ($r = 0.569^{***}$)[2] and physical landscape and place ($r = 0.413^{**}$) were identified, whereas there was a negative correlation between how often participants addressed school science and cultural content ($r = -0.351^{*}$). The physical landscape was often mentioned by participants when addressing both question content and language ($r = 0.629^{***}$), and mention of the physical landscape in question language occurred with

[2] $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

discussion of the school science content of the assessment questions ($r = 0.521^{**}$). While these data indicate that school science and physical landscape were often addressed with regard to question content and language, they were both negatively correlated with cultural content (school science language and culture content $r = -0.479^{**}$ and physical landscape language and culture content $r = -0.396^{*}$). Negative correlations were also identified with content related to school science topics and the physical landscape with respect to question format ($r = -0.466^{**}$ and $r = -0.359^{*}$, respectively).

Participant gender, expertise, and residence time (e.g., number of years living in one's hometown) correlated with different aspects of question content and language. For example, female participants mentioned the cultural content of questions more often than their male counterparts ($r = 0.424^{**}$). Participants who completed physics coursework mentioned cultural content more often when discussing question content ($r = 0.427^{**}$) but did not mention the physical landscape as often when discussing question language ($r = -0.429^{**}$). Participants who completed geology coursework showed the opposite effect, speaking more about the physical landscape with regard to question language ($r = 0.388^{*}$) than to cultural content ($r = -0.366^{*}$). No significant correlations were identified between either of the other major scientific disciplines (chemistry and biology) and the interview codes. Finally, those participants who had lived longer in their hometowns were more likely to mention place in their review of assessment questions than those who had become residents more recently ($r = 0.461^{**}$).

Regression analysis of significant correlations identified the most influential among the multiple interview codes and demographic data. Multiple regressions indicated that discussion of school science content in the assessment questions was influenced by discussion of the physical landscape in question language ($\beta = 0.421^{*}$) over the cultural and physical landscape content of the question. Similarly, discussion of the physical landscape with regard to question content and language was strongly related ($\beta = 0.653^{***}$), whereas question content regarding place did not have significant influence on the discussion of question content with regard to the physical landscape. These data show that the interview discourse focused primarily on the school science content of the question and the question language that incorporated the physical landscape. Furthermore, discussion of the physical landscape in question language occurred in tandem with discussion of the physical landscape content of the question.

### *Interview Exchanges: Question Content*

Blackfeet students often focused on school science topics and physical landscape in their discussions of the GCI questions, with minor mention of culture and place as part of question content. Often the student participants would relate the content of the question to the local landscape, particularly geologic features that they experienced every day. The following exchange highlights the emphasis placed on presenting geoscience content in a local context. The participant explains how incorporating a local context into the GCI question would help him answer the question given his prior knowledge of the landscape. The participant was addressing the similarity of response options in a GCI item pertaining to mountains:

Which of the following best describes mountains? **Choose all that apply.**

(A) Old mountains are taller than young mountains because old mountains have been growing longer.
(B) Old mountains have gentler slopes than young mountains because old mountains have been wearing down longer.
(C) Old mountains have more vegetation than young mountains because old mountains have had plants growing on them longer.
(D) Old mountains have rougher surfaces than young mountains because old mountains have been around longer.
(E) All mountains are roughly the same age regardless of shape, size, vegetation or roughness.

**Interviewer:** *You said that you kind of got tripped up on it a little bit [by the question]. What was confusing about it? You mentioned the similarity of answers ...*

**Participant:** *Yeah, yeah. I really didn't understand where it was going ... you know what I mean, or what are we trying to get at? So I guess you could probably change the question a little bit into more of a descriptive answer. You know which of the following best describes the Rocky Mountains or a certain region of mountains, then you'd have a more directive answer so we could [choose them] a little better.*

**Interviewer:** *So you think by putting a specific location here that would help interpret the answers?*

**Participant:** *So it would be easier for me to give an educated guess if that was to say Rocky Mountains you know, how did they get to be the way they were and whatnot, so.*

**Interviewer:** *Why would the Rocky Mountains be easier for you to make an educated guess?*

**Participant:** *You know what, probably because I've lived here. I've lived in the Rocky Mountains for somewhat a time, so I mean I know a little about them. You know, like I say, I don't know a lot, a whole bunch [about mountains] I would probably be able to relate to them to their physical properties a little better than just a general question on mountains.*

### *Interview Exchanges: Question Language*

Participants noted use of technical language in the GCI items, noting terms such as tectonic plates, erosion, and mountain building might be difficult to interpret for some, and suggested that these terms be replaced with more descriptive response options. While the use of these terms might be valid for assessing geoscience majors, they may be invalid in assessment measures of nonscience majors or the public. The following excerpt highlights these issues with communication validity. The participant had read aloud the following GCI item and options pertaining to wind:

Which of the following can be caused by wind? **Choose all that apply.**

(A) Tectonic plate motion
(B) Waves

(C) Earthquakes
(D) Mountain building
(E) Erosion

***Interviewer:*** *What were you thinking about while you were reading that question?*

***Participant:*** *Well, I was just picking out some things that I thought would be caused by the wind as I was reading it, you know, the waves and the erosion. The first—can I just tell you something really, really quick about me?*

The participant went on to describe how he had grown up speaking the native language and learned English when he was nearly middle-school age. He mentioned that even today, he comes upon English words that are unfamiliar to him and he uses a dictionary to look up the definitions. After his explanation, he ended with:

***Participant:*** *... and the first thing hit me, tectonic plate motion. I don't BS myself into thinking, well, I know what this is. If I don't know it, I don't know it and that's flat the way it is.*

### *Interview Exchanges: Culture*

Aspects of native culture were discussed by participants in response to individual prompts throughout the interviews; however, the content of one particular GCI item sparked discussion about native culture in the context of rocks and minerals with both Montana and Arizona participants. Participants read the following GCI item and options aloud:

Are rocks and minerals alive?

(A) Yes, rocks and minerals grow.
(B) Yes, rocks are made up of minerals.
(C) Yes, rocks and minerals are always changing.
(D) No, rocks and minerals do not reproduce.
(E) No, rocks and minerals are not made up of atoms.

In the following excerpts, participants described how the question could be interpreted in different ways depending upon the context in which the question was asked. Depending on the context, people might interpret the word *alive* in different ways; one might interpret *alive* as being "living and breathing" and another might interpret *alive* as something with "energy" or "power." These different meanings for the word *alive* have their roots in the school science (e.g., living and breathing) taught in the classroom, whereas others come from the traditional native knowledge passed down from previous generations (e.g., energy and power).

***Interviewer:*** *What were you thinking about while reading/answering the question?*

***Participant:*** *Um, when I think of this question, I think of the two different aspects of what I was taught in a book and what I was culturally aware of. By you asking me these questions I want to give you "right" answers. I would tell you from the book learning I got, the answer you wanted to hear, but also thinking of what my grandfather and grandmother taught me, that everything is connected and everything is alive.*

Participants were asked to discuss whether the GCI item had any personal meaning and whether they thought it would have meaning to their peers and local community. In their responses, participants discuss both native knowledge and scientific knowledge about rocks and minerals. The following participant discusses the difference between the two knowledge bases and the validity of each.

***Interviewer:*** *Would you find this question meaningful to your students?*

***Participant:*** *Yes, again in both aspects. It depends on where you teach and if you have native students. If you have native students, you want them to develop the difference between native and scientific values. Know that both can be possible. They are different but both valid. It's important to have both values and know neither is wrong.*

Another participant interpreted the GCI item through a scientific lens to answer the interviewer's question:

***Interviewer:*** *Can you relate the content of this question to any personal experiences you've had?*

***Participant:*** *Well, the question "rocks and minerals, are they alive," especially from a native perspective. We recognize that certain materials have power. It does not necessarily make them alive, but it makes them important. People tend to separate alive versus not alive as if one was important and one was not. I was thinking as a scientist when I read that question.*

In other exchanges, the participant spoke of generational differences among people in their community. Differences in discourse were identified in younger and older participants, with younger participants typically referencing scientific knowledge gleaned from the classroom, whereas older participants were able to move between the different worldviews more readily, discussing both the traditional and the scientific meaning of rocks and minerals.

***Interviewer:*** *Would you find this question meaningful to your local community?*

***Participant:*** *Yes. I feel like Elders would say yes, they are alive, everything on Earth is alive. Kids would say no, it's just a rock. It shows the difference between generations.*

Overall, the interviews provided important context in which to view the construct, communication, and cultural validity of the GCI items. These data indicate that by incorporating locally significant landscapes, removing technical language, and having an awareness of how certain everyday terms (e.g., *alive*) might be interpreted differently are crucial in validating assessment in these communities. This highlights the importance of seeking input on assessment from local faculty and students.

## DISCUSSION

Through the sequential Delphi surveys and focus groups, cultural experts from the Blackfeet and Diné communities provided insight into the use and value of assessment; the role of assessment in instruction, curriculum development, and student learning; and the sources of bias in assessment. The data gleaned from these surveys suggest that though a variety of assessments are used in the classroom, assessment that is most highly valued aligns with the traditions of the Blackfeet and Diné communities. This valuing includes acknowledgment of common language, such as use of tense, and appropriate assessment styles, such as oral or written, that align with community norms. Experts highly value portfolios as an assessment style as well, echoing the sentiment of many in the broader science education community (e.g., Slater, 1997; Mintzes et al., 2005). Experts use and value pre- and posttests but do not value standardized tests, an interesting finding. The experts understand the utility of pre- and posttesting as way to gauge student learning from a course or intervention; however, they choose not to use standardized questions as the means to do this. Rather, they prefer to use contextualized questions to measure student learning in their classrooms.

The items within the GCI are considered a standardized approach to pre- and posttesting. Assessments designed in this project can be implemented in pre- and posttesting that are valued by experts and will include open-ended and concept inventory-type items. Concept inventory items developed by participants will incorporate meaningful content (e.g., features of the landscape) and language (both scientific and native languages) familiar to these Native American communities involved in the project. Furthermore, the concept inventory items could address expert concerns about the time involved in both test administration and analysis as they can become part of the GCI database, which enables testing and analysis to be done online.

Pre- and posttesting would also align with the roles of assessment that experts found most important (indicated in italics). Pretests inform instructors about the knowledge their students bring into the classroom prior to instruction, allowing instructors to *make decisions on the direction of instruction*. Using pre- and posttesting, instructors can (1) identify whether *students are receiving the information* and (2) encourage students to *reflect* on how they answer the questions and whether their understanding has broadened or changed as a result of instruction. Expert feedback guides the assessment design such that the type of the assessment aligns with what is used and valued by the community, as well as what fills the roles of assessment important in curriculum development, instruction, and student learning.

Experts identified sources of test bias that aligned with what researchers in educational assessment have found to be areas of concern (Nelson-Barber and Trumbull, 2007), including how tests are created (lack of input from community) and how they are administered (used inconsistently). Experts indicate that context is important in assessment design and that question content should incorporate a local context. By having Native American instructors and students involved in the assessment design process to comment on question content, language, and format, local context can be identified and incorporated into the assessment.

Feedback provided by experts, tribal college faculty, and students from BCC and ASU characterized types of contextual elements that can be incorporated into assessment to make it more meaningful to diverse student groups. Themes of culture, place, native science, and geoscience are common to these elements. Participants described how ideas of culture and place can be embodied in the physical landscape (e.g., physical locations with cultural significance). With native science and geoscience in particular, participants describe important concepts and topics that are more ecologically oriented and compatible with an Earth-system science framework. Addressing important and meaningful geoscience topics with assessment, and situating such concepts within an Earth-system science framework, ultimately addresses the cultural validity of assessment and enhances both communication and construct validity.

Thematic content analysis of interview transcripts that included discussion of select GCI questions clearly shows the importance of the physical landscape in question content and language. Participants appeared to focus more on the content than the language of the GCI questions in the interviews. This is likely an artifact of the type of questions selected for review. The GCI is guided by the research-based "rules" of multiple-choice item construction (e.g., Haladyna and Downing, 1989; Frey et al., 2005); therefore, it contains questions that are already relatively free of jargon. That said, participants focused on the physical landscape nearly as often as they discuss the school science content of the assessment items. Furthermore, participants discussed question content related to place when discussing content addressing the physical landscape. These significant correlations emphasize the collective importance of physical landscape and cultural landscape (i.e., places) in assessment content. Discussion of the role of cultural attributes and practices not directly situated in places in the question content occurred less frequently, perhaps indicating that participants were able to make clear connections with the assessment content more readily by discussing the local landscape and their connection to it, rather than invoking specific cultural knowledge and traditions. Alternatively, given the importance of place in native cultures, perhaps incorporating content that addresses place provides a link between culture and physical landscape. Validation of the existing assessment items identified important content and appropriate language, as well as highlighting content and language that would be meaningful to the test taker and leverages familiar knowledge of the local landscape.

Participants indicated that geoscience content should incorporate features of the local landscape (ideally those that combine cultural and geologic significance) to make assessment more relevant to students. Older faculty and students were able to find meaning in many of the questions, noting applications of the content to their culture and traditions. Younger students seemed to focus more on the school science content of the question, and some did not readily relate question content to their personal experiences outside of the classroom. The relationship between age and place is supported by the strong correlation between residence time (how long the participant had lived in their hometown) and discussion of place content in the assessment questions. This trend could indicate a difference in life experiences among the different age groups or different levels of understanding of their culture as related to geoscience, or it could be an artifact of the exchanges between interview and interviewer. Aikenhead and Jegede's 1999 paper on "border crossings" between Western science and native science describe how some indigenous students compartmentalize these two frameworks and "cross borders" to participate in each, depending on the situation. In this scenario, students have to deal with dual identities, separating what they have learned in everyday life from life in the classroom. Bang and Medin (2010) argued that culturally based science education helps students navigate these borders more readily and is effective in motivating students to pursue science. Our collaborative work with cultural experts, faculty, and students works to ease the border crossings by finding commonalities between the two frameworks and to produce assessment that is perhaps a better measure of conceptual understanding than those that do not incorporate this local context. Our data suggest that the local landscape may be a key link in designing meaningful geoscience assessment, given that it can be both culturally and geologically significant in these communities.

## CONCLUSION

Data gathered through mixed methods offer the basis for developing place-based and culturally informed geoscience assessment and reveal important local context. This feedback addresses the cultural validity of new and existing assessment and acts to bridge local assessment needs and culturally valid practice with "standardized" instruments (e.g., GCI). Cultural validation is a community process, as indicated by the collaboration among cultural experts, tribal college faculty and students, and university faculty, staff, and researchers.

Using the data collected from focus groups, surveys, and one-on-one interviews, the Blackfeet expert group has since written several open-ended assessment items that incorporate the local landscape and address important geoscience content. These assessment items are linked to both educational standards developed by the Blackfeet educators (BCCRSI, 2005) and content standards outlined in the Earth Science Literacy Principles document collaboratively developed by the geoscience community (ESLI, 2009). The resultant assessment items incorporate the expert, faculty, and student feedback described in this paper. Future work will gather student answers through open-ended questionnaires and develop items in a variety of assessment styles that will include both multiple-choice and question clusters.

### Acknowledgments

The authors thank the numerous participants from Diné College, BCC, and ASU and members of the Blackfeet and Diné communities for their work on this project. Thanks also to Hannah Clark who created the participant map in ArcGIS. This project is supported by the National Science Foundation (NSF GEO-1034909 and GEO-1034926). 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 NSF. We thank the two reviewers and editorial board for their thoughtful comments that improved this manuscript.

## REFERENCES

Aikenhead, G., and Michell, H. 2011. Bridging cultures: Scientific and indigenous ways of knowing nature. Don Mills, ON, Canada: Pearson Canada.

Aikenhead, G.S., and Jegede, O.J. 1999. Cross-cultural science education: A cognitive explanation of a cultural phenomenon. *Journal of Research in Science Teaching,* 36(3):269–287.

Ault, C.R. 2008. Achieving *querencia:* Integrating a sense of place with disciplined thinking. *Curriculum Inquiry,* 38:605–637.

Bang, M., and Medin, D. 2010. Cultural processes in science education: Supporting the navigation of multiple epistemologies. *Science Education,* 94(6):1008–1026, doi:10.1002/sce.20392.

Barnhardt, R., and Kawagley, A.O. 2005. Indigenous knowledge systems and Alaska Native ways of knowing. *Anthropology and Education Quarterly,* 36(1):8–23.

Basso, K.H. 1996. Wisdom sits in places: Landscape and language among the western Apache. Albuquerque, NM: University of New Mexico Press.

BCCRSI (Blackfeet Community College Rural Systemic Initiative). 2005. Blackfeet education standards implementation guide. Browning, Montana: Blackfeet Community College.

Bertrand, J.T., Brown, J.E., and Ward, V.M. 1992. Techniques for analyzing focus group data. *Evaluation Review,* 16(2):198–209.

Brandenburg, A.M., and Carroll, M.S. 1995. Your place or mine? The effect of place creation on environmental values and landscape meanings. *Society and Natural Resources,* 8:381–398.

Cajete, G. 2000. Native science: Natural laws of interdependence. Santa Fé, NM: Clear Light Publishers.

Carroll, J.B., ed. 1964. Language, thought, and reality: Selected writings of Benjamin Lee Whorf. Cambridge, MA: MIT Press.

Coles-Ritchie, M., and Charles, W. 2011. Indigenizing assessment using community funds of knowledge: A critical action research study. *Journal of American Indian Education,* 50(3):26–41.

Deloria, V., Jr., and Wildcat, D. 2001. Power and place: Indian education in America. Golden, CO: Fulcrum Resources.

Demmert, W.G. 2005. The influences of culture on learning and assessment among Native American students. *Learning Disabilities Research and Practice,* 20(1):16–23.

Elder, J. 1998. Stories in the land: A place-based environmental education anthology. Great Barrington, MA: The Orion Society.

ESLI (Earth Science Literacy Initiative). 2009. Earth science literacy principles: The big ideas and supporting concepts of earth science. Available at http://earthscienceliteracy.org/es_literacy_22may09.pdf (accessed 28 December 2012).

Frey, B.B., Petersen, S.E., Edwards, L.M., Pedrotti, J.T., and Peyton, V. 2005. Item-writing rules: Collective wisdom. *Teaching and Teacher Education,* 21:357–364.

Gibson, B.A., and Puniwai, N. 2006. Developing an archetype for integrating Native Hawaiian traditional knowledge with Earth system science education. *Journal of Geoscience Education,* 54(3):287–294.

Gould, P., and White, R. 1986. Mental maps, 2nd ed. Boston: Allen & Unwin.

Graham, M., Milanowski, A., and Miller, J. 2012. Measuring and promoting inter-rater agreement of teacher and principal performance ratings. Rockville: MD: Center for Educator Compensation Reform, p. 9.

Haladyna, T.M., and Downing, S.M. 1989. Validity of a taxonomy of multiple choice item-writing rules. *Applied Measurement in Education,* 2:51–78.

Hammersley, L.C., Levine, R., Cornwell, K., Kusnick, J.E., and Hausback, B.P. 2012. The geology of Mexico: A quantitative evaluation of a course designed to increase the number of Hispanic students participating in the geosciences at California State University, Sacramento. *Journal of Geoscience Education,* 60(2):189–198.

Hasson, F., Keeney, S., and McKenna, H. 2000. Research guidelines for the Delphi survey technique. *Journal of Advanced Nursing,* 32(4):1008–1015.

Hsu, C.-C., and Sandford, B.A. 2007. The Delphi technique: Making sense of consensus. *Practical Assessment, Research & Evaluation,* 12(10):1–7.

Jolley, A., Lane, E., Kennedy, B., and Frappé, T.-P. 2012. Perceptions of Earth and ocean science courses, development of an instrument to measure (SPESS). *Journal of Geoscience Education,* 60:83–91.

Kawagley, A.O., and Barnhardt, R. 1999. Education indigenous to place: Western science meets native reality. *In* Smith, G.A., and Williams, D.R., eds., Ecological education in action: On weaving education, culture, and the environment. Albany, NY: State University of New York Press, p. 117–140.

Kitzinger, J. 1995. Qualitative research: Introducing focus groups. *BMJ,* 311:299–302.

Kudryavtsev, A., Krasny, M.E., and Stedman, R.C. 2012. The impact of environmental education on sense of place among urban youth. *Ecosphere,* 3(4):29, 1–15.

Lee, O. 1997. Scientific literacy for all: What is it, and how can we achieve it? *Journal of Research in Science Teaching,* 34(3):219–222.

Lee, O. 2001. Culture and language in science education: What do we know and what do we need to know? *Journal of Research in Science Teaching,* 38(5):499–501.

Lee, O., and Luykx, A. 2007. Science education and student diversity: Race/ethnicity, language, culture, and socioeconomic status. *In* Abell, S.K., and Lederman, N.G., eds., Handbook of research in science education, 2nd ed. Mahwah, NJ: Lawrence Erlbaum Associates, p. 171–197.

Libarkin, J.C. 2001. Development of an assessment of student conception of the nature of science. *Journal of Geoscience Education,* 49:435–442.

Libarkin, J.C. 2008. Concept inventories in higher education science. *In* National Research Council Promising Practices in Undergraduate STEM Education Workshop 2. Washington, DC: National Academies Press.

Libarkin, J.C., and Anderson, S.W. 2005. Assessment of learning in entry-level geoscience courses: Results from the Geoscience Concept Inventory. *Journal of Geoscience Education,* 53:394–401.

Libarkin, J.C., Ward, E.M.G., Anderson, S.W., Kortemeyer, G., and Raeburn, S.P. 2011. Revisiting the Geoscience Concept Inventory: A call to the community. *GSA Today,* 21:26–28.

Luykx, A., Lee, O., Hart, J., and Deaktor, R. 2007. Cultural and home language influences on children's responses to science assessments. *Teachers College Record,* 109(4):897–926.

McCoy, T.J. 2011. Asiihkiwi neehi kiisikwi: A multi-generational, culturally embedded Earth and sky curriculum for the Myaamiaki [Abstract]. *In* Proceedings of the 42nd Lunar and Planetary Science Conference. Houston, TX: Lunar and Planetary Institute. Available at http://www.lpi.usra.edu/meetings/lpsc2011/pdf/2227.pdf (accessed 10 April 2011).

Mintzes, J.J., Wandersee, J.H., and Novak, J.D., eds. 2005. Assessing science understanding: A human constructivist view. Burlington, MA: Elsevier Academic Press.

Nelson-Barber, S., and Trumbull, E. 2007. Making assessment practices valid for indigenous American students. *Journal of American Indian Education,* 46(3):132–147.

Okoli, C., and Pawlowski, S.D. 2004. The Delphi method as a research tool: An example, design considerations and applications. *Information & Management,* 42:15–29.

Orr, D.W. 1992. Ecological literacy: Education and the transition to a postmodern world. Albany, NY: State University of New York Press.

Osborne, J., Collins, S., Ratcliffe, M., Millar, R., and Duschl, R. 2003. What "ideas-about-science" should be taught in school science? A Delphi study of the expert community. *Journal of Research in Science Teaching,* 40(7):692–720.

Palmer, M.H., Elmore, R.D., Watson, M.J., Kloesel, K., and Palmer, K. 2009. Xoa:dau to Maunkaui: Integrating indigenous knowledge into an undergraduate Earth systems science course. *Journal of Geoscience Education,* 57:137–144.

Petcovic, H., and Ruhf, R. 2008. Geoscience conceptual knowledge of preservice elementary teachers: Results from the Geoscience Concept Inventory. *Journal of Geoscience Education,* 56:251–260.

Relph, E. 1976. Place and placelessness. London: Pion.

Riggs, E.M. 2005. Field-based education and indigenous knowledge: Essential components of geoscience education for Native American communities. *Science Education,* 89:296–313.

Riggs, E.M., Robbins, E., and Darner, R. 2007. Sharing the land: Attracting Native American students to the geosciences. *Journal of Geoscience Education,* 55(6):478–485.

Riggs, E.M., and Semken, S.C. 2001. Culture and science: Earth science for Native Americans. *Geotimes,* 46:14–17.

Rowe, G., and Wright, G. 2001. Expert opinions in forecasting: the role of the Delphi technique. *In* Armstrong, J.S., ed., Principles of forecasting: A handbook for researchers and practitioners. Boston: Kluwer Academic Publishers, p. 125–144.

Sauer, C.O. 1925. The morphology of landscape. *University of California Publications in Geography,* 2(2):19–54.

Semken, S. 2005. Sense of place and place-based introductory geoscience teaching for American Indian and Alaska Native undergraduates. *Journal of Geoscience Education,* 53:149–157.

Semken, S. 2011. A sense of the American Southwest: Place-based Earth system science for diverse students. *In the Trenches,* 1(3):1–4.

Semken, S., and Brandt, E. 2010. Implications of sense of place and place-based education for ecological integrity and cultural sustainability in contested places. *In* Tippins, D., Mueller, M., van Eijck, M., and Adams, J., eds., Cultural studies and environmentalism: The confluence of ecojustice, place-based (science) education, and indigenous knowledge systems. New York: Springer, p. 287–302.

Semken, S., and Butler Freeman, C. 2008. Sense of place in the practice and assessment of place-based science teaching. *Science Education,* 92:1042–1057.

Semken, S.C., and Morgan, F. 1997. Navajo pedagogy and Earth systems. *Journal of Geoscience Education,* 45:109–112.

Slater, T.F. 1997. The effectiveness of portfolio assessments in science. *Journal of College Science Teaching,* 26(5):315–318.

Sobel, D. 2004. Place-based education: Connecting classrooms and communities. Great Barrington, MA: The Orion Society.

Solano-Flores, G., and Li, M. 2009. Generalizability of cognitive interview-based measures across cultural groups. *Educational Measurement: Issues and Practice,* 28(2):9–18.

Solano-Flores, G., and Nelson-Barber, S. 2001. On the cultural validity of science assessments. *Journal of Research in Science Teaching,* 38(5):553–573.

Solano-Flores, G., Trumbull, E., and Nelson-Barber, S. 2002. Concurrent development of dual language assessments: An alternative to translating tests for linguistic minorities. *International Journal of Testing,* 2(2):107–129.

Teed, R., and Slattery, W. 2011. Changes in geologic time understanding in a class for preservice teachers. *Journal of Geoscience Education,* 59:151.

Tuan, Y.-F. 1977. Space and place: The perspective of experience. Minneapolis, MN: University of Minnesota Press.

Unsworth, S., Riggs, E.M., and Chavez, M. 2012. Creating pathways toward geoscience education for Native American youth: The importance of cultural relevance and self-concept. *Journal of Geoscience Education,* 60(4):384–392.

Wiggins, G.P., and McTighe, J. 2005. Understanding by design, 2nd edition. Alexandria, VA: Association for Supervision & Curriculum Development.

Williams, D., and Semken, S. 2011. Ethnographic methods in analysis of place-based geoscience curriculum and pedagogy. *In* Feig, A.P., and Stokes, A., eds., Qualitative research in geoscience education: Geological Society of America Special Paper 474. Boulder, CO: Geological Society of America, p. 49–62.

Williams, D., and Vaske, J.J. 2003. The measurement of place attachment: Validity and generalizability of a psychometric approach. *Forest Science,* 49(6):830–840.

Woodhouse, J.L., and Knapp, C.E. 2000. Place-based curriculum and instruction: Outdoor and environmental education approaches. Digest EDO-RC-00-6. ERIC Document Reproduction Service ED448012. Charleston, WV: ERIC Clearinghouse on Rural Education and Small Schools, Appalachia Educational Laboratory.

Ziglio, E. 1996. The Delphi method and its contribution to decision-making. *In* Adler, M., and Ziglio, E., eds., Gazing into the oracle: The Delphi method and its application to social policy and public health. Bristol, PA: Jessica Kingsley Publishers, p. 3–33.

## APPENDIX

### THINK ALOUD PROTOCOL

#### Date and Time

I am going to show you a series of questions. Please read the question and answer options aloud. I want to emphasize that the purpose of this interview is to get your feedback on the content and format of the question, and therefore you need not attempt to answer the question.

#### Probes[3]

(1) What do you think this question is about?[4] (comprehension)
(2) What were you thinking about while reading/answering the question? (frame of reference)
(3) Were you confused by the question? If yes, in what way? (reasoning and judgment)
(4) Do you think that what the question is asking is important for you/your students to understand? (Construct validity[5])
(5) **Do you find this question meaningful?**
(6) **Would you find this question meaningful to:**
  a. **Your students?**
  b. **To the local community?**
(7) **Can you relate the content of this question to any personal experiences you've had?**
(8) How would you change this question to be more meaningful or relevant to the local community?
(9) (Geosci faculty only) From your perspective, does the question actually measure some aspect of geoscience understanding? (Content validity[3])
(10) END: In general, how do you feel about multiple choice questions? Do you prefer other question formats? Please explain why.

[3]Probes in bold speak to cultural validity (Solano-Flores and Nelson-Barber, 2001).
[4]From Solano-Flores and Li (2009).
[5]From GCI (Libarkin, 2008).

## Erratum

In the article by Emily M. Geraghty Ward, Steven Semken, and Julie C. Libarkin, "The Design of Place-Based, Culturally Informed Geoscience Assessment," which appeared in Volume 62, Number 1 on pages 86–103 of the *Journal of Geoscience Education*, there was an error introduced in Table I during the production process. The corrected table is printed below in its entirety.

**TABLE I: Age, gender, academic training in science, and tribal affiliation for the faculty and students involved in the assessment validation.**

| | Expert Group ($n$ = 9)[1] | BCC Faculty ($n$ = 4) | BCC Students ($n$ = 4)[2] | ASU Students ($n$ = 15)[3] |
|---|---|---|---|---|
| Average age (y) | 45 (±10) | 56 (±5) | 29 (±11) | 28 (±8) |
| Gender | 4 female, 5 male | 1 female, 3 male | 2 female, 2 male | 9 female, 6 male |
| Science coursework completed in high school or college (% of group) | | Biology ($n$ = 4), physics ($n$ = 1), geology ($n$ = 3), chemistry ($n$ = 2) | Biology ($n$ = 4) physics ($n$ = 3), geology ($n$ = 2), chemistry ($n$ = 2) | Biology ($n$ = 14), physics ($n$ = 9), geology ($n$ = 11), chemistry ($n$ = 9) |
| Tribal affiliation | Blackfeet ($n$ = 6), Navajo ($n$ = 2) | Blackfeet ($n$ = 4) | Blackfeet ($n$ = 3) | Assiniboine, Navajo ($n$ = 8), Pascua Yaqui, White Mountain, Apache ($n$ = 2), Salt River Pima-Maricopa, San Carlos Apache, Chemehuevi |

[1]One expert was not affiliated with a tribe.
[2]Tribal affiliation for ASU students is $n$ = 1 unless otherwise indicated.
[3]One student is not affiliated with a tribe.

1089-9995/2014/62(4)/736/1