BACKGROUND INFORMATION FOR PRESENTERS

Assessing for Learning in Outdoor Schools

Overview

Assessment of learning can be a complex and sometimes controversial topic. Assessment is an essential part of education—to gain insight into both students' learning progress and whether teaching strategies are appropriate and effective. Grant Wiggins, a leading commentator on issues in educational assessment, notes that the root of the word *assessment* comes from *to sit with*. In its original sense, it's about connecting with the person being assessed in a way that helps gain a close understanding of what they know and understand. The purpose of assessment, then, is to gather evidence of learning—whether it is for the outdoor instructor, program, teacher, student, parent, district, state, or even nation. It's important that learning be recognized for the complex process it is. Currently, educators use many ways to assess what students know. Some of these methods look like traditional tests; others look very different. The kind of assessment you use depends on what information you want. We use the term *assessment* to refer to all those activities instructors use that gather information to be used as feedback to modify teaching and learning activities.

> **"When the cook tastes the soup, that's formative. When the guests taste the soup, that's summative."**
>
> *-Robert Stake, educational researcher*

Assessment is much more than tests! Many people think of assessment as simply tests students take before and after instruction. Those *are* assessments, but so is listening in on student discussions, watching students work, reading what students have written, asking students questions, and listening to their responses. Summative assessments are those given to students primarily when instruction is finished, to find out what students learned and maybe to assign a grade. Formative assessments take place *during* instruction and are used to figure out how to best adjust instruction to meet the needs of students. Embedded assessments are those that are part of learning experiences. Embedded assessments are often formative, but sometimes they can be summative. This session focuses on embedded formative assessments that field instructors can use to improve their instruction. The vocabulary and specific definitions of types of assessment can be confusing and are not the focus of this session. For the sake of not overwhelming instructors with vocabulary, the terms *formative* and *summative* are not part of the session; however, *embedded* (which we've found to be more intuitive to understand) is part of the session.

> **"Assessment specialists have found it useful to describe assessment as a process of reasoning from evidence—of using a representative performance or set of performances to make inferences about a wider set of skills or knowledge. The process of collecting evidence to support inferences about what students know and can do is fundamental to all assessments—from classroom quizzes, standardized achievement tests, or computerized tutoring programs, to the conversations students have with their teachers as they work through an experiment."**
>
> *-Developing assessments for the Next Generation Science Standards. National Research Council*

Teaching and learning must be interactive. One reason to emphasize and support assessment practices with your instructors is the tendency of many educators to overestimate what their students learn from their instruction. It can be a huge aha! moment when an instructor who has been teaching an activity or topic for a while, under the assumption that their students were really getting it, thoughtfully assesses student learning only to find out that many of those students are leaving with incomplete or inaccurate understandings about the topic. At the moment of this awareness, they may suddenly feel like less effective instructors, but this increased awareness from assessment is an important aspect of becoming more effective instructors. It helps instructors have a more realistic idea about

what their students are learning and gives them information they can use to improve instruction. One of the most important aspects of a quality educational program is an environment of reflective practice. That's when instructors aren't just leading a series of the same activities with each group, but are intentionally setting goals, trying out approaches and strategies to accomplish those goals, assessing to see how they went, and then making adjustments the next time they teach.

Field instructors need to know about their students' progress and struggles with learning so they can adapt their teaching to meet students' needs—needs that are often unpredictable and that vary greatly from one individual to another. Instructors can find out what they need to know in a variety of ways, including asking questions and listening when talking with students during *Walk & Talks* and *Turn & Talks*, observing students in the field, listening carefully during discussions, and reading students' written work. These are all considered embedded assessments because they are part of effective instruction. Such assessment becomes formative assessment when the evidence and information gained is actually used to adapt instruction to better meet students' needs. Opportunities for students to express their understanding should be designed into any and all pieces of teaching, because this enables instructors to receive and act on feedback directly from learners in order to improve and refine their teaching skills as well as enhance the curriculum. It kicks off an interactive connection between embedded assessment, teaching, and learning. Discussions, performance-based student activities, and journal prompts can all be used to give students those opportunities. Then it's vitally important for the instructor to look at or listen carefully to the talk, the writing, and the actions to get an idea of the state of student understanding.

Assessment is about gathering observations and making explanations from evidence. During quality teaching, the instructor is frequently gathering information about student understandings through listening in on discussions, asking questions, looking at student work, and so on. Making explanations from evidence is a key aspect of assessment because that's when an instructor attempts to make sense of their observations and tries to understand what students are thinking; that's the riskiest part of the process. It's risky because it's impossible to have certainty about anything that's going on in students' heads. It's also risky because it's easy to misinterpret and inaccurately explain your findings. Our unconscious biases may influence our explanations and cause us to jump to conclusions. We all have unconscious biases, and if we have unconscious biases about certain groups or individuals, we may unconsciously and unfairly judge their participation, work, and understandings, giving more benefit of the doubt to some while making negative assumptions about others. Without an awareness of these issues, an instructor can cause harm. Instructors need to intentionally work to avoid these biases and assumptions by spending time becoming more aware of their biases, by keeping all explanations *firmly grounded in evidence*, by maintaining uncertainty and open-mindedness about explanations, and by brainstorming multiple explanations rather than assuming that the first explanation that comes to mind is true. If this sounds familiar, it should, because it's very much like scientific thinking. Similarly with scientific thinking, as humans it's easy for us to slip into less scientific thinking and get carried away with our explanations, losing sight of the original observations. Instructors should hungrily gather observations during instruction and be careful, wary, and tentative when making explanations from their observations. Instructors can also maintain a focus on what the assessment says about their instruction, rather than only on what it says about students. Instruction is something instructors have control over, and keeping this perspective in mind helps improve instruction (and is far less risky).

> **"...assessment is always a process of reasoning from evidence. By its very nature, moreover, assessment is imprecise to some degree. Assessment results are only estimates of what a person knows and can do. Every assessment, regardless of its purpose, rests on three pillars: a model of how students represent knowledge and develop competence in the subject domain, tasks or situations that allow one to observe students' performance, and an interpretation method for drawing inferences from the performance evidence thus obtained."**
>
> *-Knowing What Students Know, National Academy of Sciences*

Emphasis on Formative Embedded Assessment

Best practices in teaching recommend that formative assessment should be woven into all stages of learning. We need to assess *before* new material is presented (to assess current knowledge), *during* the learning process (to see how well a student is grasping a concept), and *after* the lesson has been taught (to see how well the student understands the new material and to assess the effectiveness of the teaching). This lets an instructor evaluate and modify the learning experience in time to make a difference—rather than discover a problem just as it's time to move on to another subject.

Curriculum-embedded formative assessment practices create a continuous flow of information between instructor and student that guide the next steps in learning. It's important to actively involve students in the assessment process. Direct instructor feedback and expectations are shared with students to help them envision performance goals and understand what is needed for improvement. For example, when an instructor gives students feedback on how they are actively changing their minds about ideas and appropriately using language of uncertainty, the instructor is being explicit about expectations and about what students can do to improve and is also providing important feedback on how they're doing. When students clearly understand what's expected of them and receive specific feedback, then formative assessment information becomes an important guide for instructional improvement for both the instructor and the student. When an instructor allows students to be more in charge of their own learning, their motivation to learn is greatly increased. Research has shown that, when carried out effectively, formative classroom assessment that includes constructive feedback to students significantly raises the levels of student achievement. In this sense, quality formative assessment is indistinguishable from good teaching practices.

Informal assessments using narrow questions can be unproductive and can even inhibit learning. When students get to talk about their understandings in their own ways, it helps increase their knowledge and improves their understanding. Talking with a group of students also can give an instructor the chance to respond to and reorient their thinking. But there are many examples of discussions in which instructors have unconsciously responded in ways that inhibit student learning. Often this happens when the instructor is looking for a particular response and doesn't have the flexibility or confidence to deal with unexpected responses from students. Many instructors informally assess their groups by asking a series of narrow recall questions to check for understanding. Unfortunately, this natural and direct way of checking on learning is often unproductive. Another common problem is that, following a question, instructors don't wait long enough for students to think through their answers. When an instructor calls on the first raised hand or answers their own question after only two or three seconds (often because a minute of thoughtful silence seems intolerable), there isn't enough time for students to think deeply about the question. Sometimes, unintentionally, the instructor tries to direct the student with hints toward saying the answer they're looking for. In manipulating the dialogue toward a narrow goal, the instructor can sometimes block any unorthodox, though often thoughtful, attempts by students to work out their own answers. Over time, students get the message that they are not required to think out their own answers. The object of the exercise, they assume, is to work out—or guess—what answer the instructor expects to see or hear.

Assessment for learning is when students have:

- Understanding of instructional goals.
- Opportunity to develop skills and knowledge necessary to achieve goals.
- Ownership of their learning.
- Motivation to learn.

Instructors using formative assessment to support learning should:

- Develop clear learning goals for students.
- Provide opportunities to learn content in a deep way.

- Align assessment tasks with curriculum (what is taught) and instruction (how it is taught).
- Embed assessments at key points throughout instruction.
- Make changes to curriculum, instruction, and assessments based on assessment evidence from students.

Scoring guides and rubrics. These assessment tools are often used with assessments in formal classroom settings to help teachers evaluate different levels of student responses. They can be part of both formative and summative assessments and can be used to monitor progress in knowledge as well as abilities. Developing accurate and useful scoring guides and/or rubrics requires looking at many samples of student responses in order to capture a broad range of student understanding and to be able to represent the most common ways that learners build understanding of a particular concept.

Emphasis on Formative Embedded Assessment

Teachers in classrooms use summative assessments to measure student progress at the end of a unit or year of instruction. These kinds of assessments are more typically included in curriculum programs and textbooks and are designed to help find out what students have learned from the instructional materials. Since they take place *after* instruction, when students are preparing to move on to other topics, they have a very different purpose than formative embedded assessments. Summative assessments do provide feedback to the teacher about the success of the lessons, but this information can only guide instruction for the next time the materials are used, which is often not until the following school year. In this case, summative assessment can be referred to as assessment *of* learning, in direct contrast to formative assessments that are used throughout instruction as assessment *in support of* learning. In outdoor science programs, on the other hand, an instructor may often get a chance to apply the information to their instruction only a week or so later. Given the rapid cycle of instruction and repetition in many outdoor schools, the information from summative assessments may actually be used in a formative way.

Biases in assessment. We all have unconscious biases, and becoming more aware of these biases and then working to minimize them is important in assessment and instruction. Often, these biases can get "baked into" systems, such as assessments that are used regularly by programs but aren't regularly evaluated for their effectiveness. "Culture and background may lead to the inaccuracy of assessments, including traditional tests and alternative assessments. Standardized tests intend to measure intelligence and general knowledge, but they are normed based on the knowledge and values of the majority groups, which can create bias against minority groups, including gender, race, community status, and persons with different language backgrounds, socioeconomic status, and culture" (Kim & Zabelina, 2015). Teachers' different expectations for students of various racial or social classes are often correlated with less effective instructional practice (Leacock, 1969; Murray, 1996). Teachers often fail to identify potentially qualified students for selective programs in schools, especially when students are culturally different from them (Peterson, 2000).

To make any assessment more equitable, the prompt/task/question should be based on information and experiences that all your students have access to and should be both relevant and interesting to them. In outdoor science, assessments are ideally based on a common engaging experience in which the group has participated. Vocabulary can be a stumbling block to understanding how to respond, so it's important that the language be accessible to all students. Any difficult words should be used extensively and in context before appearing in assessments. Students who have more experience developing test-taking skills may be better at figuring out how to be successful on the assessment in ways that are not actually related to their understanding of the content. It's important to be clear about your expectations for responding, so all students know how to respond. Review your instruction and assessments (including questions you ask throughout your programs) for relevance and comprehensibility. Ideally, this should include guidance by a team that reflects the background of students you serve. Assessments represent one of many important reasons that people in leadership roles at your organization, including on the board, on the executive team, and as educators, should be representative of the communities of

students that you serve.

Stereotype threat. There is a large body of research on stereotype threat. Subtle and not-so-subtle perceptions of stereotypes have been shown to affect performance on assessments, even if the individual doesn't believe in the stereotype. If there is a negative societal stereotype about a particular group—for example, that black students underperform academically—the self-threat is thought to affect the intellectual performance of these students during assessments if they think the assessment measures intelligence. In this example, if students don't think the assessment measures intelligence, their performance is much better. "...[E]ven passing reminders that someone belongs to one group or another (e.g., asking the test taker to identify their race or gender on the test form), that has been stereotyped as inferior in academics, can wreak havoc with test performance.... By subtly altering the test situation to remove stereotype threat, Aronson and his colleagues have demonstrated dramatic improvement in standardized test scores among members of negatively stereotyped groups" (American Psychological Association, 2006). The book *Whistling Vivaldi and Other Clues to How Stereotypes Affect Us*, written by the social psychologist Claude Steele, is a very accessible read about stereotype threat and the underperformance of minority students in higher education.

There are things an instructor can do to reduce stereotype threat and bias:

- Reflect on unconscious biases they have and how their language might trigger certain responses from students, particularly during times when students feel they are being evaluated.
- Be transparent about how they are using information from journals and performance tasks: to help students formulate and express their ideas, to help themselves as instructors to understand what they got out of an experience, to help instructors improve their program and their teaching—and specifically *not* to evaluate the students.
- Have assessments reviewed by multiple stakeholders who represent the students served.
- Encourage students to have a growth mindset and to recognize that intelligence is not something you're just born with, but can be increased through effort.
- Encourage students to do self-affirmation, writing and/or talking about a value that is important to them, or a positive attribute they have. In 2006, researchers Geoffrey L. Cohen, Julio Garcia, Nancy Apfel, and Allison Master did a brief written self-affirmation with middle school students and found that it improved the grades of black students and reduced the racial achievement gap by 40%.
- Help students socially self-identify as part of the academic world. Help them see themselves as scholars/scientific thinkers.
- Communicate clearly that diversity is valued.
- Facilitate students' engaging with other students who are different from them.
- Reduce stress by helping all students feel seen and socially connected in the group.

Summative Assessment and Evaluation in Outdoor Science Programs

Many residential and multi-day outdoor science programs are interested in gathering summative information about student learning at the end of their programs. For this reason, student and teacher surveys are often used to get feedback about the effectiveness of field instruction and the success of specific program activities. These kinds of self-reporting surveys can indeed provide some information about important aspects of the program, particularly student attitudes, behavior, and social interactions. However, getting valid and reliable assessment information about what students have learned as a result of outdoor learning experiences conducted in an informal setting can be problematic. See the excerpts from the *Learning Science in Informal Environments* report highlighted below to read more about the complex conditions under which these kinds of assessment are developed and administered. This report provides both a broad description of science learning in informal environments and a detailed review of the

evidence of their impact on science learning. Published in 2009 by the National Research Council, it synthesizes literature across multiple disciplines to identify a common framework of educational goals and outcomes and provides insights into educational practices in informal settings such as outdoor science programs.

"The characteristics of informal learning environments make it very difficult to develop practical, evidence-centered ways to assess learning outcomes…[The assessment itself is] logistically complex, but also the data gathered are hard to interpret. It can be difficult to separate the effects of a single visit from other factors that could be contributing to positive learning outcomes. And arranging for tests before and after the experience or setting up other traditional measures…can be disruptive, or even inappropriate for the purpose that assessment may serve…[It] is important to consider the rationale for assessing learning in informal science learning settings.

[Informal learning] experiences cannot fully be prescribed or predetermined. Rather, the environments are learner-centered; so much of what happens emerges during the course of activities…The challenge thus becomes how to document the learning that occurs while not sacrificing the freedom and spontaneity that is integral to the experience.

The collaborative and social aspects inherent in many informal experiences also pose a challenge for assessing learning. Participants in summer camps, science centers, family activities, hobby groups, and such are generally encouraged to take full advantage of the social resources available in the setting to achieve their learning goals…Thus, assessments that focus on an individual's performance alone may "under-measure" learning…Assessing whether participants working in a group have grasped the science is important, but measuring the role that collaboration and problem solving have played in learning may be equally so.

…A first step in developing assessments is identifying the anticipated learning goals…It is equally important to understand what knowledge, skills, and beliefs the target audience brings to the learning situation… In fact, defining outcomes and target audiences for informal science learning experiences can be the most challenging tasks in the assessment process because it requires a deep understanding about purpose and the various ways in which informal experiences may be connected to past and future learning experiences… The development of assessments appropriate for science learning in informal environments should be guided by three criteria.

First, the assessments must address the range of capabilities that the designers have in mind, including not only cognitive outcomes, but also attitudinal, behavioral, and social outcomes…

Second, assessments should fit with the kind of participant experiences that make informal learning environments attractive and engaging. Any assessment activities undertaken in these settings should not undermine the very features that make for effective learning.

Third, the assessments must be valid; that is, they should measure what they purport to measure (construct validity) and align with opportunities for learning that are present in the environment (often referred to as ecological validity). In short, assessment measures should capture as much of the breadth of learning that a reasonable audience could experience, should align with the nature of the learning experience, and should represent in some faithful way the learning that actually occurs. Doing so is not easy."

-Surrounded by Science: Learning Science in Informal Environments , National Research Council