

## **Assessing Projects: Types of Assessment Validity and Reliability of Formative Assessment**

### **Collecting Good Assessment Data**

Teachers have been conducting informal formative assessment forever. It is human nature, to form judgments about people and situations. Most of these kinds of judgments, however, are unconscious, and many result in false beliefs and understandings. For the data collected from formative assessments to be valid, it must assess what it claims to assess, and to be reliable, it must provide information that can be replicated.

Valid assessments accurately target specific skills, strategies, and knowledge. Answering multiple-choice questions about problem solving in mathematics, for example, does not really give a teacher information about how well students solve problems. Answering these questions correctly may show that students have memorized how to use a problem-solving strategy or show that they have highly-developed guessing skills, but it will not show how students perform under authentic problem-solving situations. Rarely are these easy-to-score assessments valid for 21st century skills.

Stiggins (2004) warns, "We have not invested in ensuring the accuracy of classroom assessments. Thus the chances of inaccurate assessment and therefore ineffective decision making at all over levels clearly increase" (p. 25). When teachers make decisions about students' knowledge and abilities too quickly with too little information, their conclusions can inhibit growth rather than encourage it.

Araison (2001) describes some threats to validity:

1. Stereotyping, drawing conclusions based on personal impressions or previous biases
2. Logical errors, evaluating students' abilities based on irrelevant characteristics, such as how they are dressed or the achievements of their siblings. (These judgments are usually unconscious and teachers are unaware of making them.)
3. Inadequate sampling, making judgments based on just one observation or piece of information
4. Generalizing, assuming that when students behave in a certain way in one situation, they will behave the same way in other situations

Data collected about student performance must also be reliable. Reliable information is consistent and typical. Any assessments of students' thinking collected, for example, the day before a long holiday, are likely to be unreliable since student's behavior is bound to be atypical.

For assessment data to help teachers draw useful conclusions it must be both valid, showing something that is important, and reliable, showing something that is usual. Researchers use the term "triangulation" to describe the process that is used to draw conclusions from data. Like a journalist who seeks corroboration before printing evidence of a crime, a teacher needs more than one piece of information before drawing a conclusion about a student's ability. Even then, a conclusion must be tentative and open to contradictory data. This means, for example, that a teacher may see that a child has difficulty generalizing in a group project and in a learning log entry. Later, however, the child may show that she can generalize in a different subject area. The teacher can make a tentative conclusion that the child's inability to generalize is connected to her insufficient subject-area knowledge, not necessarily her thinking expertise.

Most teachers are alert and maintain continuous awareness of their students. They cannot help but notice how students are behaving and what they are saying. Unfortunately, they rarely consider this kind of informal observation as formative assessment and do not record what they see in a systematic manner. These kinds of observations, when used without careful analysis, can result in skewed perspectives and faulty decisions because they do not consider enough data. Instruction based on data collected in haphazard or unsystematic ways can impede student progress. The careful collection and consideration of information about students derived from formative assessments takes time and planning, but the effect this kind of assessment has on student learning and motivation makes it well worth the effort.