SESSION: Principles of Assessment

Date: 4 December 2016

Time: (1 hour)

Learning Outcomes: By the end of this session participants will be able to:
- identify issues to consider when designing, implementing or evaluating an assessment
- explain the principles of validity, reliability, feasibility, and backwash

Materials: whiteboard and marker, dice, concept handouts

Procedures:
Step 1: Warm-Up-Substitution Table Riddle Race---------------------------------------**10 min**
Warm up?

Step 2: Reflecting on Assessment--**15 min**
Divide participants into five groups and have each group to answer the following questions, then share their answers with the whole group and review together:
- Who assesses students' learning? (teacher, but student self-assessment makes them more independent and more willing to accept teachers assessments)
- What is assessed? (individual progress, but also if class is achieving course goals)
- Why assess students' learning? (grades, identify needs, primarily to know if teaching is effective)
- How can you assess student's learning? (depends how they were taught- assess students in a way similar to how you taught them)
- What is done with the results of the assessment? (primarily to inform teaching, but also grades and reporting to others)

Step 3: Different kinds of tests for different purposes------------------------------**10 min**
Have participants at a group guess the purpose of the following kinds of tests one at a time, and talk through how those different kinds of tests are used, eliciting examples from their experience: proficiency, entrance, placement, diagnostic, achievement, and exit tests.

Step 4: Assessing assessments --- **20 min**
Divide participants into four groups and assign each one a principle (with definition handout): Validity, Reliability, Feasibility, and Backwash. Have participants present their principle with examples, then review as a group. Give example of a multiple choice national exam focused on academic vocabulary and grammar, with a later emphasis on communicative skills but no changes to the exam. Have groups assess Validity, Reliability, Feasibility, and Backwash

Step 5: Conclusion---**5min**
Summarize the session in general and provide time for final questions.

1.   **VALIDITY:**   A test is valid if it tests what it claims to test.

•   How do we know? Does the test assess only what it claims to assess and nothing else, or does it require the student to use OTHER knowledge or skills to complete the assessment?
•   In an end-of-unit vocabulary test, would it be good [valid] if several of the vocabulary items had not been taught in class?
•   Would it be good [or valid] to assess essay-writing ability with a multiple choice test on grammar and stylistic knowledge?
•   Would it be good [or valid] to assess listening ability by asking students to respond to the listening in short essay answers (that is, writing out their answers in paragraph?

2.   **RELIABILITY:**   Are the results consistent?

•   If the same person takes the same test again one week later (without any further study or preparation) and the results are very different, the test's reliability is in question.
•   If two people who have very similar abilities take the same test, but have very different results, the test's reliability is in question.
•   If two or more trained raters using the same criteria come up with very different evaluations of the same speaking or writing sample, the test's reliability is in question.

3.   **FEASIBILITY**: Are we happy with the "cost/benefit ratio" of the test?

•   That is, how much time/money/test taker goodwill is required in relation to the amount and type of information gathered?
o   For example, if you want to have a very accurate test of international students' ability to perform in academic classes at a U.S. college, you could have them spend a week participating and doing assignments in a college classroom. With observations, assignments, and instructor impressions you could predict very well how each student would do in her/his "real" college courses.
However, it would cost an incredible amount of time and energy and money to do this. Maybe we find ways to assess international students academic ability with a test (SAT? ACT? GRE?)-- that might take less time, be more cost-effective, and be good enough in terms of information.

4.   **BACKWASH** is the effect that testing has on learning.
Backwash can be positive or negative. A test that reinforces the way that the content was taught in class is positive backwash. If the test makes the teaching seem unhelpful—that it did not prepare the students for the test, then the backwash is negative.
•   For example, if I have been learning functions (permission and requests, giving directions, apologizing and thanking), and role plays have been a regular routine in class and I know that we will be assessed using similar role plays, the assessment will encourage my learning and have positive backwash.
•   However, if I am teaching a very communicative syllabus, but the final exam I am required to give is a multiple choice test on grammar rules and individual vocabulary items, the

backwash will be negative. Students may even stop putting effort into class activities and focus on rote memorization of items typically found on the multiple choice final exam.