Steps in Developing an Assessment of an Educational Outcome

Facilitator: Larry Gruppen

There are several essential steps to developing an assessment of an educational outcome. Although these don't follow a strict sequence, there are some that must come before others. Also, some steps will entail more ambiguity and innovation than others, but which steps those are depends on the particular situation.

Step 1: What is the outcome/construct/"thing" you want to assess?

- This is the most important step.
- Clarity on what you want to assess is essential, otherwise you'll never know if the data are relevant or not.
- This is where theory becomes valuable as a guide to defining these outcomes.
- "Knowledge" in and of itself is seldom enough.

Step 2: What is the purpose of the assessment? What decisions do you want to make from the assessment?

- Decisions require data, which in turn, require assessment.
- Don't assess unless there is a decision to be made.
- Is the purpose formative (low-stakes) or summative (high-stakes)?
- Who cares (stakeholders)?

Step 3: Who are the assessees/learners?

- In whom do you want to assess the outcome? This influences the level of difficulty, logistics, authenticity, etc., of the assessment process.
- Different learners are more or less accessible, cooperative, etc.

Step 4: What observable behaviors reflect the outcome/construct/"thing"?

- We can only assess what we can observe.
- Even assessing knowledge requires the behavior of answering a question.
- Think broadly and creatively about these behaviors – there is seldom only one behavior that matters.

Step 5: What tasks or activities are likely to elicit these behaviors?

- How can you set up tasks or identify situations to observe these behaviors. Again, think creatively about this.
- Naturally occurring tasks and activities may be ideal but may also be very difficult to observe.
- "Artificial" tasks and activities may be more manageable, though possibly less valid.

Step 6: How will you capture or record these behaviors?

- Some behaviors are enduring (answers to MCQ), others are passing (interactions with patients).
- Passing behaviors require concurrent observations or scoring.

Step 7: How will you quantify or score the behaviors?

- Simplest score is a checklist – present vs absent.
- More complex and sophisticated scoring may include weights for different behaviors and complex combinations of scores.

Step 8: How will you use the scores to make your decision (see Step 2)?

- Do you need score cut-points and standard setting to help you make your decisions (pass/fail; honors, high pass, pass; admit/reject)?

Step 9: How will you communicate the results to the stakeholders?

- Do you need reference points for interpreting the scores (e.g., class average)?
- Are the scores linked to other assessments?
- Is there a portfolio or record? Score cards?

Step 10: What evidence can you assemble to convince others that the decisions you make on the basis of your assessment can be trusted?

- The five sources of evidence for construct validity can be used as a framework for evaluating an assessment or evaluation method.
- The following questions are intended to serve as a guide towards identifying and organizing evidence that you might gather to substantiate your claims about the quality and accuracy of your assessment model.
- Central to this analysis is a clear definition of WHAT you are trying to measure (the construct) – if you can't define it, you can't measure it.

Source of evidence	Guiding questions	Specific evidence *(construct = what are you trying to measure)*
Content *Does the measure map the content aspects of the construct?*	- What are the boundaries of the construct you are assessing (i.e., what tasks and content are relevant to this construct)? - Who has judged the relevance of the assessment content? - Did they judge the mapping of the assessment content to the underlying construct (i.e., identify gaps or duplication)? The relevance of the content? - What expertise or credibility do these judges have?	
Response Process *Does the measure map the cognitive or performance processes of the construct?*	- How are you verifying that your assessment is measuring the underlying processes relevant to the construct? - Does the assessment require examinees to do the same mental or physical procedures that the construct entails?	

Internal Structure *Does the structure of the assessment map onto the structure of the construct?*	● Are the scoring models or procedures consistent with the construct? ● Are there subscores relevant to sub-components of the construct? ● Is the combination of scoring elements appropriate to the construct? ● Do you have an "adequate" sample of tasks or items?	
Relations to Other Variables *Mapping of predicted relationships of scores with other measures*	● Do the scores from your assessment relate to other assessments or examinee characteristics in theoretically predicted ways? ● What other assessments should be similar and what should be different? ● Will scores on your assessment generalize to other situations, times problems, etc?	
Consequences *Are the judgments made on the basis of the scores appropriate?*	● What are the intended and unintended outcomes of using the assessment scores? ● What impact does passing or failing have on the examinee? ● To what other uses might these data be put, and are they appropriate? ● Are pass/fail decisions appropriate (not too easy or too difficult)?	