When you need to assess your students, where do you begin?
1. The type of assessment task I will use
2. How I will maintain test security?
3. How I can help my students do well so that they will succeed after they finish my class?
4. How I can help my students do well so that they will make me look good?
5. All of the above?

Why do we teach?
The primary purposes of language teaching are to (e.g.):
- promote or facilitate learning;
- enhance learners’ linguistic, cognitive, emotional, and social development

Why do we assess?
Primary use of an assessment is to gather information to help us make decisions that will lead to beneficial consequences for stakeholders.

In the classroom, we use language assessments to inform two kinds of decisions: formative and summative:

Formative decisions relate to making changes in teaching and learning activities in support of, or to promote or enhance learning.

Formative decisions are made during the processes of teaching and learning.

Summative decisions relate to passing or failing students on the basis of their progress or achievement, or certifying them based on their level of ability.

Summative decisions are made after the processes of teaching and learning.

Modes of classroom assessment
Implicit mode: (“dynamic assessment”, “on-line assessment”, “continuous assessment”)
- Instantaneous and cyclical:
  - assessment – decision – instruction;
  - assessment – decision – instruction
- Learners are largely unaware that assessment is taking place.
- Used primarily for formative decisions.

---

Explicit mode: Assessment as “assessment”

- Separate activity from teaching
- Both teacher and learners know this activity is an assessment.
- Used for both formative and summative decisions.

**Explicit mode of classroom assessment**

*When do we assess?*

Whenever we need to make an instructional decision, or a decision about learners, we need to assess.

**Occasions for classroom assessment**

- Warm-up, revision (self-assessment, implicit assessment)
- Presentation (implicit assessment)
- Guided practice (implicit assessment)
- Independent practice (self-assessment)
- “Assessment” (explicit assessment)

**Types of decisions for which language assessments are used**

- Guiding teaching and learning
- Entrance, readiness
- Placement
- Achievement/progress
- Certification
- Selection (e.g., employment, immigration)

Many of these decisions are “high stakes”.

Therefore, need to ask some questions:

1. What beneficial consequences do we want to bring about?
2. What decisions do we need to make to help promote the intended consequences?
3. What information do we need collect in order to make the most appropriate decisions?
4. How can we gather this information?

**{SMALL GROUP WORK: FOR GROUP’S ASSESSMENT PROJECT, ANSWER THE FOUR QUESTIONS ABOVE}**

**Accountability**

- We must be able to justify the use we make of a language test.
- That is, we need to be ready if we are held accountable for the use we make of a language test.
- In other words, we need to be prepared to convince stakeholders that the intended uses of our test are justified.
Whom do we need to convince? (Who are the stakeholders?)

- Ourselves
- Our fellow teachers
- Test takers (our students)
- School administrators
- Parents, guardians
- Other stake-holders (e.g., potential employers, funding agencies)

{SMALL GROUP WORK: FOR GROUP’S ASSESSMENT PROJECT, DESCRIBE THE STAKEHOLDERS}

How do we do this?

We need a conceptual framework that will enable us to justify the intended uses of our assessments.

An “Assessment Use Argument” (AUA) (Bachman, 2005; Bachman & Palmer, 2010) provides such a framework.

Two activities in justifying the uses of our assessments:

- Develop an assessment use argument (AUA) that the intended uses of our assessment are justified, and
- Collect backing (evidence), or be prepared to collect backing in support of the AUA.

Assessment Use Argument

Provides:

- the rationale and justification for the decisions we make in designing and developing the test.
- the conceptual framework for linking the test taker’s performance to our intended consequences and decisions.

Parts of an Assessment Use Argument

Claims: statements about our intended interpretations and uses of test performance; claims have two parts:

- An outcome
- One or more qualities claimed for the outcome

Data: information on which the claim is based.

Warrants: statements justifying the claims

Rebuttals: statements about possible alternatives to the outcomes or to the qualities that are stated in the claims.

Backing: the evidence that we need to collect to support the claims and warrants, or to reject the rebuttals in the AUA.
Qualities of Outcomes of Claims in an AUA

1. Claim: *consequences* are
   - beneficial

2. Claim: *decisions* are
   - values sensitive
   - equitable

3. Claim: *interpretations* are
   - meaningful
   - impartial
   - generalizable
   - relevant
   - sufficient

4. Claim: *assessment records* are
   - consistent

Figure 1: Structure of an Assessment Use Argument
(Bachman & Palmer, 2010, p. 104)
Articulating Claims for Intended Uses

<table>
<thead>
<tr>
<th>AUA Articulation and Assessment Development Guide Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Consequences</strong></td>
</tr>
<tr>
<td>1. What beneficial consequences do we intend to promote by using the assessment and making decisions based on it, and whom do we intend to benefit from these consequences?</td>
</tr>
<tr>
<td>2. How can we assure that the consequences of using the assessment will be beneficial to each stakeholder group who are affected?</td>
</tr>
<tr>
<td>3. How can we assure that the consequences of the decisions that are made will be beneficial to each stakeholder group who are affected?</td>
</tr>
<tr>
<td>4. What are the potential detrimental consequences of false positive and negative classification decisions, and how might we mitigate these?</td>
</tr>
<tr>
<td><strong>Decisions</strong></td>
</tr>
<tr>
<td>5. What specific decisions need to be made to promote the intended consequences, and who will be responsible for making these decisions?</td>
</tr>
<tr>
<td>6. How can we assure that these decisions take into consideration existing educational and societal values and relevant legal requirements?</td>
</tr>
<tr>
<td>7. How can we assure that these decisions will be equitable?</td>
</tr>
<tr>
<td><strong>Interpretations</strong></td>
</tr>
<tr>
<td>8. What do we need to know about the test takers’ language ability in order to make these decisions?</td>
</tr>
<tr>
<td>9. How can we assure that the interpretations of ability are meaningful?</td>
</tr>
<tr>
<td>Answer:</td>
</tr>
<tr>
<td>10. How can we assure that the interpretations about test takers’ language ability are impartial for all groups of stakeholders?</td>
</tr>
<tr>
<td>11. How can we assure that the interpretation of language ability generalizes to the TLU domain of the decision?</td>
</tr>
<tr>
<td>12. How can we assure that our interpretations about test takers’ language ability are relevant to the decision?</td>
</tr>
<tr>
<td>13. How can we assure that our interpretation about test takers’ language ability provides sufficient information to make the decisions?</td>
</tr>
<tr>
<td><strong>Assessment records</strong></td>
</tr>
<tr>
<td>14. How will we assure that these assessment records are consistent?</td>
</tr>
<tr>
<td>15. How will we assure that assessment records are of comparable consistency across different groups of test takers?</td>
</tr>
</tbody>
</table>

Table 1: Assessment development guide questions

(Bachman & Palmer, 2010, pp. 164-166)

Claim 1
- Outcome: Consequences
- Quality: Beneficence

Articulate Claim 1 and list and describe:
- the intended consequences
- the stakeholders.
Generic version of Claim 1: The consequences of using an assessment and of the decisions that are made are beneficial to stakeholders.

{EXAMPLES OF CLAIM 1, pp. 2, 24}

{SMALL GROUP WORK: ADAPTING CLAIM 1 FOR PARTICIPANTS’ PROJECTS}

Claim 2
- Outcome: Decisions
- Qualities: Values-sensitivity, Equitability

Generic version of Claim 2: The decisions that are made on the basis of the interpretation take into consideration existing educational and societal values and relevant legal requirements and are equitable for those stakeholders who are affected by the decisions.

{EXAMPLES OF CLAIM 2, pp. 3, 27}

{SMALL GROUP WORK: ADAPTING CLAIM 2 FOR PARTICIPANTS’ PROJECTS}

BREAK: 10:30 – 11:00

====================================================================

PART II: 11:00 – 12:30

Claim 3
- Outcome: Interpretation
- Qualities: Meaningfulness, Impartiality, Generalizability, Relevance, Sufficiency

Generic version of Claim 3: The interpretations about the ability to be assessed are:

- meaningful with respect to a particular learning syllabus, a needs analysis of the abilities needed to perform tasks in the TLU domain, or a general theory of language ability or any combination of these.
- impartial to all groups of test takers,
- generalizable to the TLU domain,
- relevant to the decision to be made, and
- sufficient for the decision to be made.

{EXAMPLES OF CLAIM 3, pp. 5, 28}

Meaningfulness warrants define the ability we want to assess, with respect to one of more frames of reference, and specify the conditions under which test takers’ performance will be elicited.
Meaningfulness Warrant 1 provides the *conceptual definition* of construct.

*Generic version of meaningfulness Warrant 1:* The *conceptual definition* of construct is based on a frame of references such as teaching syllabus, a needs analysis, or current research and/or theory of language use, and clearly distinguishes the construct from other, related constructs

{EXAMPLES OF MEANINGFULNESS WARRANT 1, pp. 5, 28}

Meaningfulness Warrant 2 provides the *operational definition* of the construct.

*Generic version of meaningfulness Warrant 2:* The *operational definition* clearly specifies the conditions under which we elicit performance from which we can make inferences about the construct we intend to assess, along with the procedures we will use for scoring or otherwise reporting the results of the students’ performance.

{EXAMPLES OF MEANINGFULNESS WARRANT 2, pp. 5, 28}

{SMALL GROUP WORK: ADAPTING CLAIM 3 AND MEANINGFULNESS WARRANTS 1 and 2 AND CREATE EXAMPLE ASSESSMENT TASK FOR PARTICIPANTS’ PROJECTS}

Generalizability warrants describe:

1. the TLU domain
2. the tasks in the TLU domain, and
3. the correspondence between characteristics of TLU task and assessment task

{EXAMPLES OF GENERALIZABILITY WARRANTS, pp. 6, 29 }

{SMALL GROUP WORK: ADAPTING GENERALIZABILITY WARRANT FOR PARTICIPANTS’ PROJECTS}

{SMALL GROUP WORK:}

1. ADAPT THE GENERALIZABILITY WARRANT FOR YOUR PROJECT.
2. SPECIFY A TLU DOMAIN FOR YOUR PROJECT.
3. DESCRIBE THE CHARACTERISTICS OF A TLU TASK USING A TASK CHARACTERISTICS TEMPLATE.
4. CREATE AN ASSESSMENT TASK AND DESCRIBE ITS CHARACTERISTICS, A TASK CHARACTERISTICS TEMPLATE.
5. COMPARE THE CHARACTERISTICS OF TLU AND ASSESSMENT TASKS}
### Describe a TLU task

<table>
<thead>
<tr>
<th>Setting:</th>
<th>Description of Task Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical characteristics</td>
<td></td>
</tr>
<tr>
<td>Participants</td>
<td></td>
</tr>
<tr>
<td>Time of task</td>
<td></td>
</tr>
</tbody>
</table>

**Input:**

| Format (channel, form, language, length, vehicle, degree of speededness, type) |  |
| Language characteristics |  |
| Organizational (grammatical, textual) |  |
| Pragmatic (Functional, sociolinguistic) |  |
| Topical content |  |

**Expected Response:**

| Format |  |
| Language characteristics |  |
| Organizational |  |
| Pragmatic |  |
| Topical content |  |

**Relationship between Input and Response**

<p>| Type of external interactiveness |  |
| Scope |  |
| Directness |  |</p>
<table>
<thead>
<tr>
<th></th>
<th>Characteristics of TLU Task</th>
<th>Characteristics of Assessment Task</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Setting:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical characteristics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participants</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time of task</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Input:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Format (channel, form,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>language, length, vehicle, degree of speededness, type)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Language characteristics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organizational</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(grammatical, textual)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pragmatic (Functional,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sociolinguistic)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Topical content</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Expected Response:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Format</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Language characteristics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organizational</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pragmatic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Topical content</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Relationship between Input and Response</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type of external</td>
<td></td>
<td></td>
</tr>
<tr>
<td>interactiveness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scope</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Directness</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Summary:

Five things to think about before using an assessment

When developing a language assessment:

1. Begin with consequences.
   - What beneficial consequences do I want to bring about?
     A. How will using an assessment help my students improve their learning?
     B. How will using an assessment help me improve my teaching?
     C. How might using an assessment be detrimental to my students?

2. Consider decisions.
   - What decisions do I need to make?
     A. What decisions do I need to make to help my students improve their learning?
     B. What decisions do I need to make to improve my teaching?
     C. How can I make sure that my decisions are equitable and values sensitive?

3. Identify the information you need.
   - What information about test takers do I need in order to make these decisions?
     A. Do I need to know if students have mastered the learning objectives of the lesson or course?
     B. Do I need to know if students are ready for the next grade or level in the program?
     C. Do I need to know if students will be able to perform language use tasks at the university, or in a job?

4. Consider the quality of the information you need.
   - How will I make sure that the information I collect about my students is:
     A. Meaningful (e.g., reflects the content of the lesson or course, or the language needs in a TLU domain)
     B. Impartial (i.e., not biased for or against any particular student or group of students)
     C. Generalizable, relevant (i.e., tells me something about my students’ ability to use language in settings outside the test itself?)
     D. Sufficient (i.e., provides enough information for me to make a decision)

How can I get the information I need?

A. Can I obtain this from observing students in my class?
B. Do I need to make a conscious effort to informally assess my students more regularly and consistently?
C. Do I need to give my students a formal assessment or test?
D. How will I report the results of my observations or assessment? (e.g., scores, profile of strengths and areas for improvement, verbal descriptions as feedback on their work)
E. How will I make sure that my assessment reports are consistent?
Selected References


