Justifying the Use of Language Assessments: 
Linking Interpretations with Consequences

Lyle F. Bachman

Department of Applied Linguistics, University of California, Los Angeles
lfb@humnet.ucla.edu


Topics in this Presentation

- Genesis of our approach
- Limitations of other approaches
- Basic premises of our approach
- Uses of language assessments
- Accountability
- Assessment use argument (AUA)
- Qualities of claims in an AUA
- Summary

Genesis of our approach

- Increasing concern, among language testers, with issues of impact, consequences, ethics, and fairness in language assessment (e.g., Alderson & Wall 1996; Cheng 1999; Davies 1997; Hamp-Lyons 1997; Shohamy 2001)
- Recognition of the need to link validity issues with the consequences of using language tests (e.g., Hamp-Lyons; papers in Kunnan 2000b);
- Clearer conceptualizations of the roles of validity and fairness in language (e.g., Kunnan 2000a, 2004); and
- Argument-based approaches to validation in education.

Limitations of previous approaches to validity

1. Messick’s (1989) “unitary” view of validity provides a coherent theoretical framework for validity, but this provides no guidance on how to conduct validation research for a particular assessment.
2. However, in assessment practice, all validation is local. That is, it is aimed at supporting claims about the intended uses of a particular assessment, for a particular group of test takers, and in a particular setting.
3. Many current approaches to validity are essentially lists of qualities or standards that are not clearly related to each other (e.g., American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999; Bachman & Palmer, 1996; Kunnan, 2004; Messick, 1989)

These approaches are based on premise the that the particular set of qualities included will add up to some global notion of assessment value:

- Messick, Kane, Mislevy: (unitary) validity
- Bachman & Palmer: usefulness
- Kunnan: fairness

4. “Argument-based” approaches (e.g., Kane, 2006; Mislevy, Steinberg, & Almond, 2002) deal explicitly with the links between test takers’ performance and score-based interpretations, but are unclear about how these relate to assessment use.

That is, these fail to show how traditional concerns with the validity of interpretations are related to the consequences of assessment use.
5. Current approaches are focused either on test development (e.g., "evidence centered design" Mislevy, 2003) or on interpretations of assessments ("interpretive argument" Kane, 2006), but none appear to explicitly relate test development to test interpretation and use.

**Basic premises of our approach**

- Fairness, or fair test use, is an overriding concern in both the development and the use of language assessments.
- A concern for fairness implies a concern for consequences of the assessment use for stakeholders.
- A concern for consequences, in turn, implies the need for accountability to stakeholders.
- Therefore, any approach to developing and using a language assessment must begin with a consideration of accountability for the consequences of assessment use.
- Test developers and test users (decision makers) need the tools to enable them to be accountable to stakeholders.
- These tools are not provided by:
  - Abstract frameworks of validity, fairness, or usefulness
  - Increasingly sophisticated measurement models and analytic methods
  - Academic discussions of ethical issues and societal contexts.

**Uses of language assessments**

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

---

Links from test taker’s performance to intended uses
(after Bachman & Palmer, 2010, Figure 2.2, p. 23)
Types of decisions for which language assessments are used:

- Entrance, readiness
- Placement
- Changes in instruction
- Changes in approaches to or strategies of learning
- Achievement/progress (pass/fail)
- Certification
- Selection (e.g., employment, immigration)
- Allocation of resources

Many of these decisions are “high stakes”.
- Have major, life changes consequences for stakeholders
- Decision errors (false positives/negatives) are difficult to reverse

Therefore, need to ask:
- What information do we need to help us make the most equitable decisions?
- How can we gather this information?
  - Teacher judgments?
  - Classroom assessments?
  - Formal tests?
- How can we assure that the information we get will be meaningful and relevant to the decisions to be made?

**Accountability**

- We must be able to *justify* the use we make of an assessment.
- That is, we need to be ready if we are held *accountable* for the use we make of an assessment.
- In other words, we need to be prepared to *convince stakeholders* that the intended uses of our assessment are justified.

**Whom do we need to convince?**

All Stake holders:
- Ourselves
- Test takers (our students)
- School administrators (at various levels)
- Parents, guardians
- Government officials
- Other stake-holders (e.g., potential employers, funding agencies)

**How do we do this?**

- Through the process of *assessment justification*.
- *Assessment justification* is the process that test developers will follow to *investigate* the extent to which the intended uses of an assessment are justified.
- We need a conceptual framework to guide assessment justification, of the process of justifying the intended uses of our assessments to stakeholders.
- An “Assessment Use Argument” (AUA) (Bachman, 2005; Bachman & Palmer, 2010) provides such a framework.
- Assessment justification comprises two interrelated activities:
  1. Developing and Assessment Use Argument (AUA) that the intended uses of our assessment are justified, and
  2. Collecting backing (evidence), or being prepared to collect backing in support of the AUA

**Assessment Use Argument**

- Provides the conceptual framework for linking our intended consequences and decisions to the test taker’s performance.
TEST INTERPRETATION AND USE

Consequences

Decisions

Interpretations about test taker's ability

Assessment Reports/Scores

Assessment Performance
Also provides the rationale and basic for justifying the decisions we make in designing and developing an assessment. And thus provides a guide for designing and developing language assessments.
- **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

- **Warrants:** statements justifying the claims

- **Data:** information on which the claim is based.

- **Backing:** the evidence that we need to collect to support the claims and warrants in the AUA.

In this presentation I will only discuss claims and warrants.
Claims and warrants in an Assessment Use Argument
(after Bachman & Palmer, 2010, Figure 5-5, p. 104)
Qualities of Claims in an AUA

Claim 1
➤ Outcome: Consequences
➤ Quality: Beneficence

Claim 2
➤ Outcome: Decisions
➤ Qualities:
  • Values-sensitivity
  • Equitability

Claim 3
➤ Outcome: Interpretation
➤ Qualities:
  • Meaningfulness
  • Impartiality
  • Generalizability
  • Relevance
  • Sufficiency

Claim 4
➤ Outcome: Assessment record (score, description)
➤ Quality: Consistency

Summary
➤ Test developers and test users (decision makers) must be prepared to be held accountable to stakeholders for the uses (decisions and consequences) of their tests.

➤ An Assessment Use Argument provides an explicit rationale and conceptual framework for justifying the inferential links between assessment performance and assessment use.

➤ An assessment use argument can:
  • Guide the design and development of assessments
  • Guide the collection of backing (evidence) in support of the warrants and claims of the assessment use argument.

Conclusion
➤ An Assessment Use Argument, along with the backing for this, provides the justification we need in order to be held accountable for assessment use—the decisions that are based on assessment performance and the consequences of these for stakeholders.

References


