



Justifying the Use of Language Assessments

Lyle F. Bachman

Department of Applied Linguistics & TESL
University of California, Los Angeles

lfb@humnet.ucla.edu

Presentation based on Bachman, L. F. & Palmer, A. S.
(Forthcoming). *Language Assessment in Practice: Developing
Language Tests and Justifying their Use in the Real World.*
Oxford: Oxford University Press.

Genesis of this work

- Increasing concern, among language testers, with issues of impact, consequences, ethics, and fairness in language assessment;
- Recognition of the need to link validity issues with the consequences of using language tests;
- Clearer conceptualizations of the roles of validity and fairness in language testing; and
- Argument-based approaches to validation in education.

Limitations of previous work

1. Messick's "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. Many current approaches to validity are essentially lists of qualities or standards that are not clearly related to each other.
3. These approaches are based on premise the that the particular set of qualities included will add up to some global notion of assessment value.

Limitations of previous work

4. “Argument-based” approaches deal explicitly with the links between test takers’ performance and score-based interpretations, but are unclear about how these relate to assessment use—decisions and consequences.
5. That is, these fail to show how traditional concerns with the validity of interpretations are related to the consequences of assessment use.

Limitations of previous work

6. Current views of validity do not explicitly distinguish between decisions and consequences.
7. Current approaches are focused either on test development or on interpretations of assessments, but none explicitly relate test development to test interpretation and use.

Limitations of previous work

8. Most current views have grown out of a quantitative measurement tradition, and thus fail to take into consideration
 - a. assessments that yield descriptions (verbal, visual) that are *not* scores, and
 - b. evidence to support the assessment that may be collected through qualitative, naturalistic, or documentary methodologies.



Topics in this Presentation

- **Uses of language assessments**
- **Accountability**
- **Assessment use argument (AUA)**
- **Qualities of claims in an AUA**
- **Conclusion**



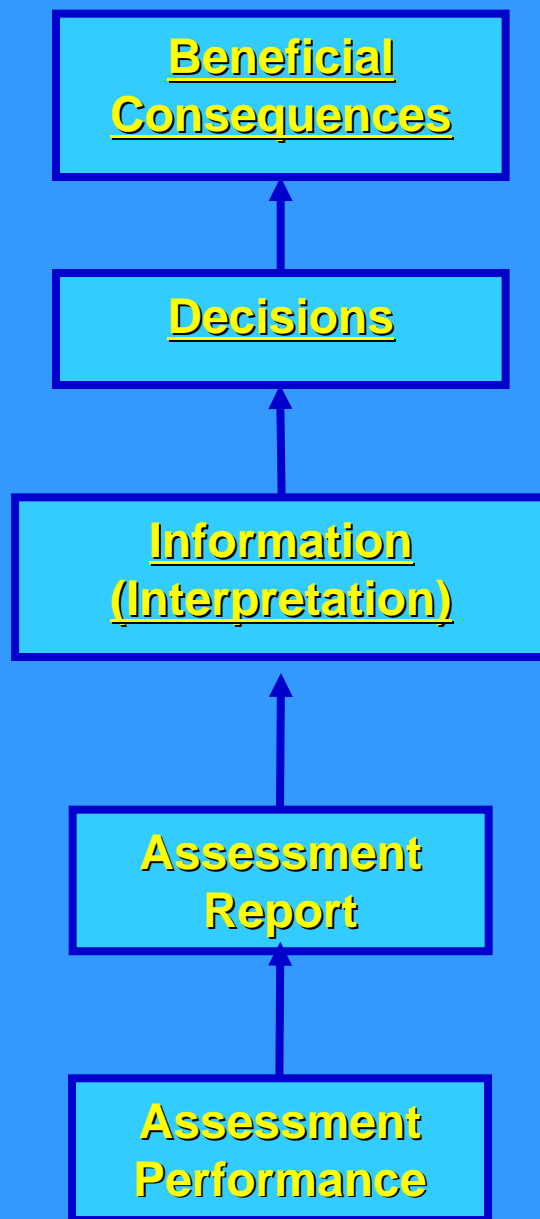
Uses of language assessments

The primary use of all language assessments is:

to gather information

to help us make decisions

that will lead to beneficial consequences for stake holders.





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
- Certification
- Selection (e.g., employment, immigration)



Uses of language assessments

- **Many of these decisions are “high stakes”.**
 - Have major, life changing consequences for stake holders
 - Decision errors (false positives/negatives) are difficult to correct or reverse.

- **Need to ask:**
 - What information do we need to help us make the most equitable decisions?
 - How can we gather this information?
 - ☞ Teacher judgment?
 - ☞ Classroom assessments?
 - ☞ Self assessments?
 - ☞ Formal tests?
 - How can we assure that the information we get will be meaningful and relevant to the decisions?



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 uses 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?

All Stake Holders:

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



How do we do this?

- **We need a conceptual framework to guide the process of justifying the intended uses of our assessments.**
- **An “Assessment Use Argument” (AUA) provides such a framework.**



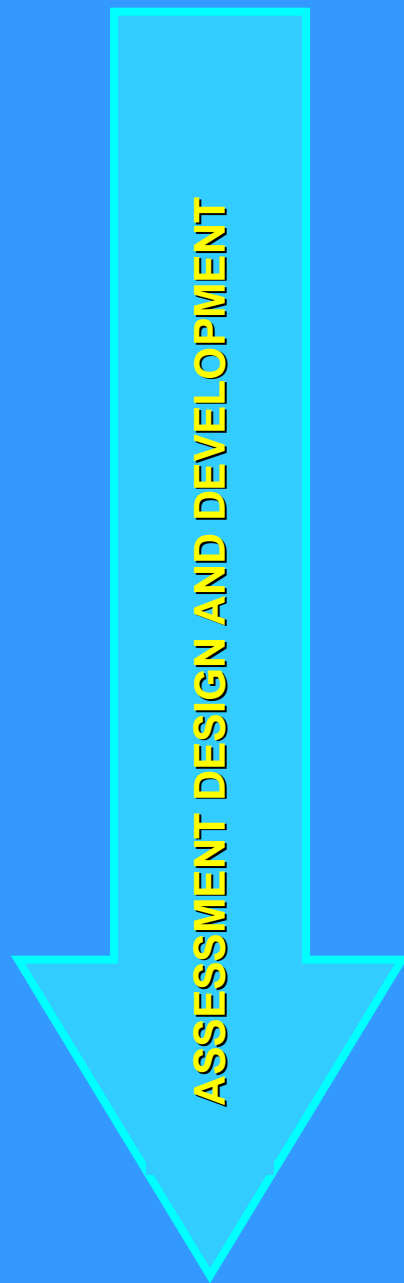
How do we do this?

1. **Develop an Assessment Use Argument (AUA) that the intended uses of our assessment are justified, and**
2. **Collect backing (evidence), or be prepared to collect backing in support of the AUA.**



Assessment Use Argument

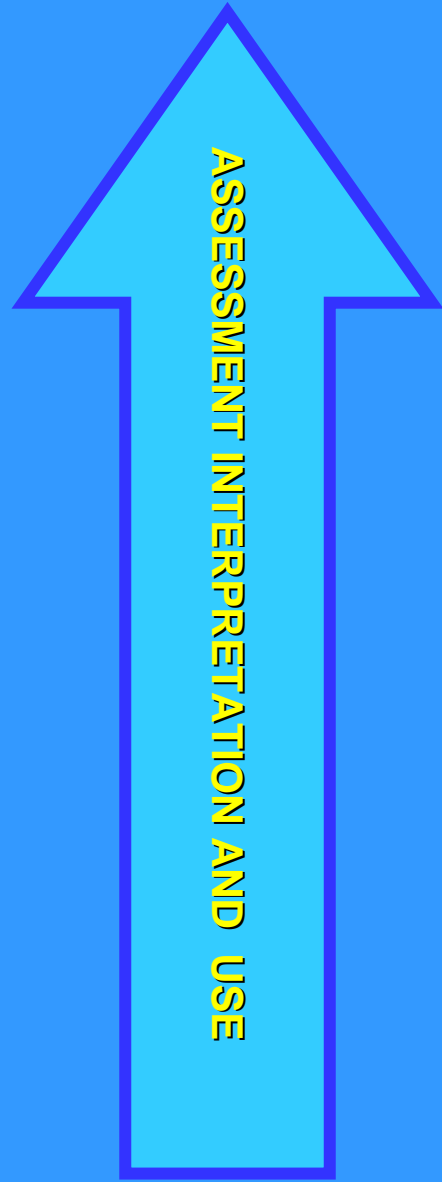
Provides the rationale and justification for the decisions we make in designing and developing an assessment.





Assessment Use Argument

Provides the conceptual framework for linking our intended consequences and decisions to the test taker's performance.





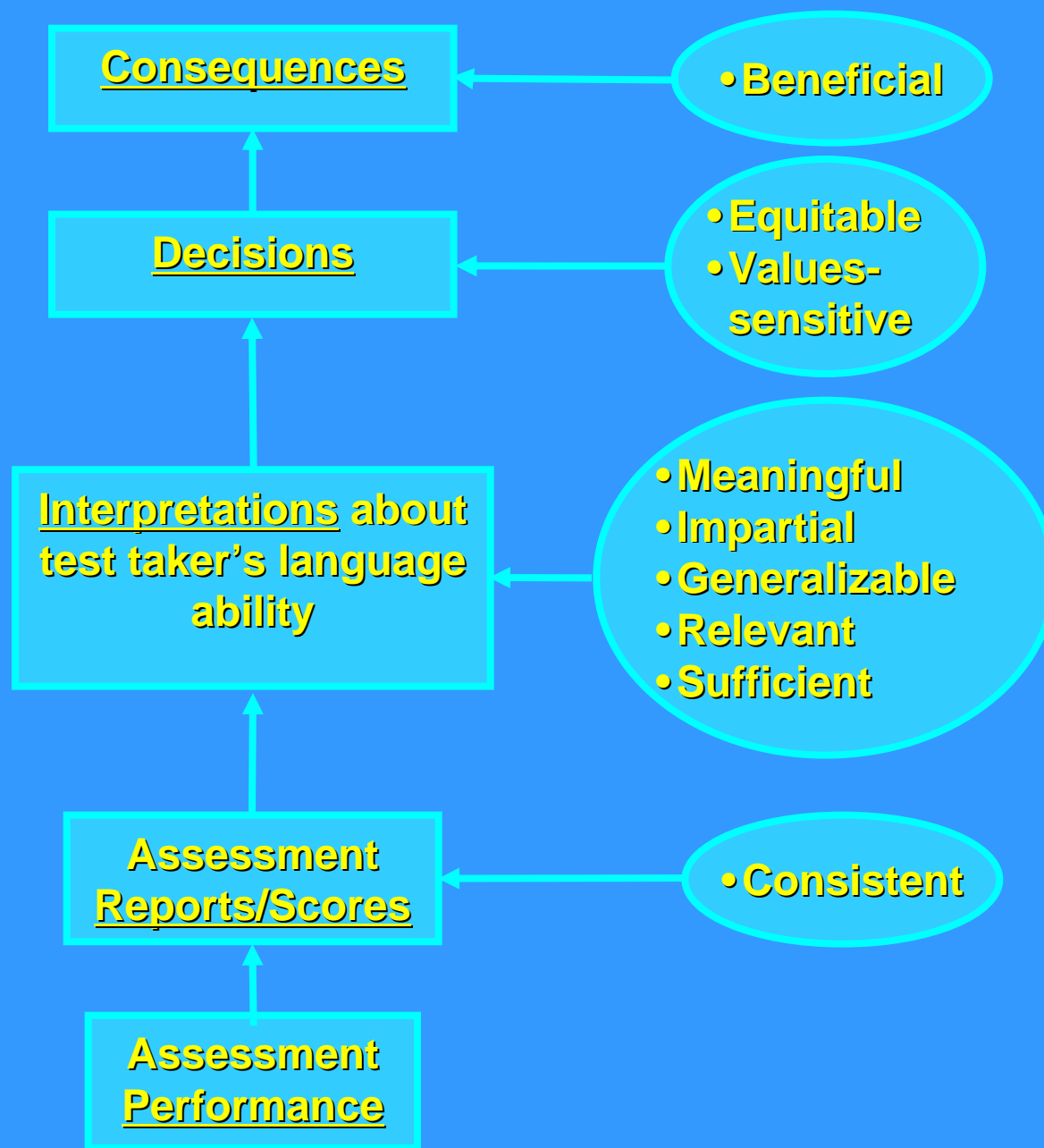
Parts of an Assessment Use Argument

- **Claims:** statements about our intended interpretations and uses of test performance; claims have two parts:
 - An outcome (consequence, decision, interpretation, record)
 - One or more qualities claimed for the outcome
- **Data:** information on which the claim is based.



Parts of an assessment Use Argument

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





Qualities of Claims in an AUA

Claim 1

- Outcome: Consequences
- Quality: Beneficence

Claim 2

- Outcome: Decisions
- Qualities:
 - Values-sensitivity
 - Equitability



Qualities of Claims in an AUA

Claim 3

- Outcome: Interpretation
- Qualities:
 - Meaningfulness
 - Impartiality (absence of bias)
 - Generalizability
 - Relevance
 - Sufficiency

Claim 4

- Outcome: Assessment record (score, description)
- Quality: Consistency



Conclusion

- **Test developers and test users (decision makers) must be prepared to justify the uses of their assessments to stake holders.**
- **An *Assessment Use Argument* provides an explicit rationale and justification for the links between assessment performance and assessment use.**



Conclusion

➤ **An Assessment Use Argument can:**

- **Guide the design and development of assessments, and**
- **Guide the collection of backing (evidence) in support of the warrants and claims of the assessment use argument.**



Conclusion

- **The Assessment Use Argument, along with the backing for this argument, provides the justification we need in order to be held accountable for uses—consequences and decisions—that are based on assessment performance.**



Thank you!



Example Toulmin argument

According to U.S. labor law, all employees who work overtime must be paid time and a half. Since Melissa worked overtime, she should have been paid time and a half, unless she is in an exempt category. Melissa's personnel file indicates that she is not in an exempt category, so we can conclude that she was paid time and a half.



Example Toulmin argument

- *Data:* Melissa worked overtime.
- *Claim:* Melissa was paid time and a half.
- *Warrant:* All employees who work overtime must be paid time and a half,
- *Backing:* According to U. S. labor law, all employees who work overtime must be paid time and a half.
- *Rebuttal:* Melissa is in an exempt category.
- *Rebuttal data:* Melissa's personnel file indicates that she is not in an exempt category.

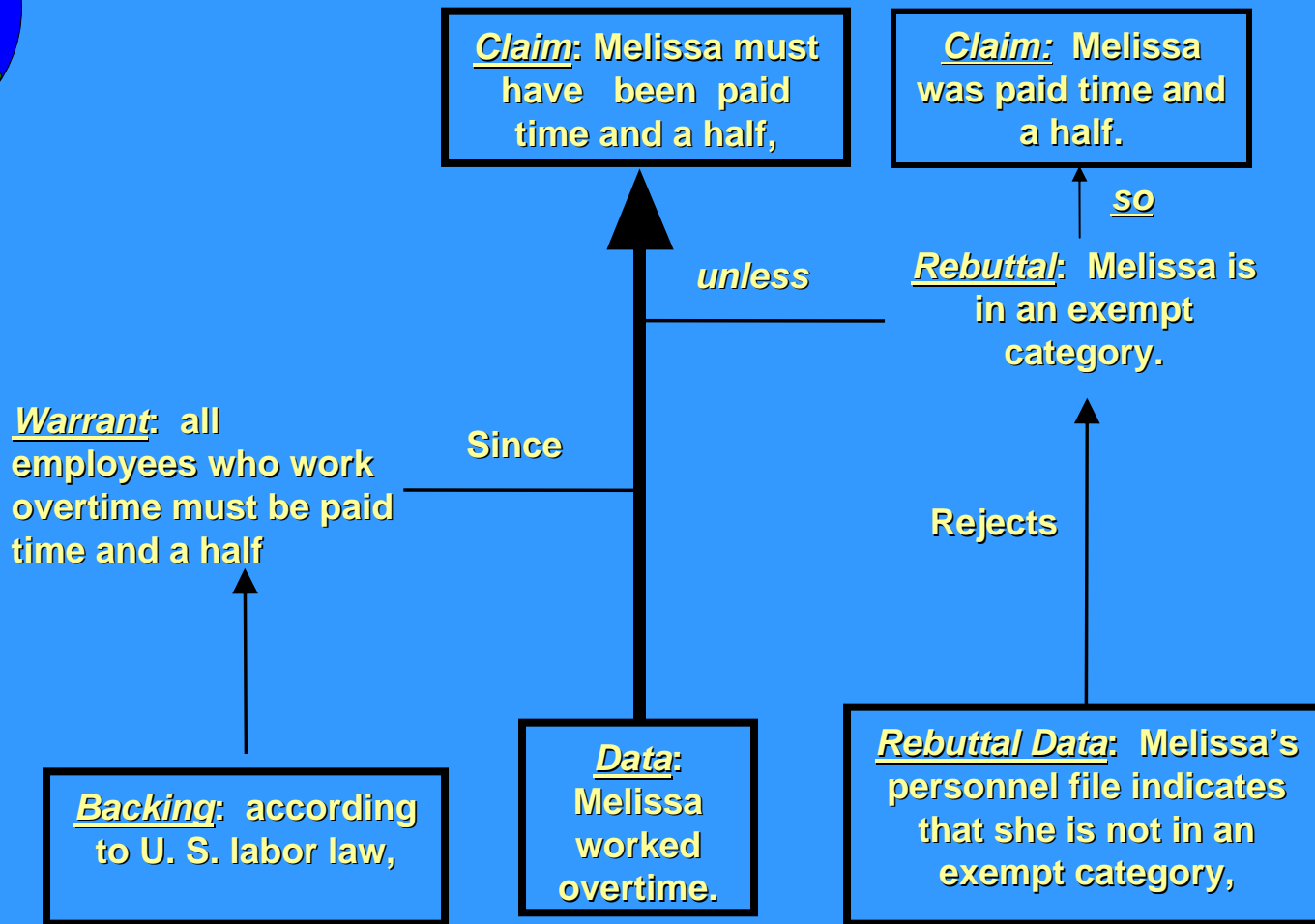


Figure 4: Toulmin Diagram for example argument