UNIVERSITÄT LEIPZIG

Task-Based Language Assessment and Testing for Proficiency: Where Do the Twain Meet?

Erwin Tschirner
L.E.A.R.N. Workshop
September 20-21, 2016
Universities at Shady Grove

Overview

- Origins of Task-Based Language Assessment
- Task-Based: Shades of Meaning
- Evidence-Centered Design (ECD)
- Relationship Between ECD and Proficiency Assessment
- Construct of the OPI
- Assessment Use Argument
- Future of Proficiency Assessment

Task-Based Language Assessment (TBLA)

The process of evaluating, in relation to a set of explicitly stated criteria, the **quality of the communicative performances** elicited from learners as part of goal-directed, meaning-focused language use requiring the **integration of skills and knowledge**" (Brindley, 1994: 74).

"The elicitation and evaluation of language use (across all modalities) for expressing and interpreting meaning **within a well-defined communicative context (and audience)**, for a clear purpose, towards a valued goal or outcome" (Norris, 2014: 25).

Early Considerations (Mislevy et al, 2002)

- Disenchantment with discrete-skills assessments focusing on grammar and vocabulary with largely decontextualized test items
  - Realization that linguistic competence (grammar and vocabulary) is not sufficient to use a language to achieve ends in social situations
- Alignment of assessment with (task-based, proficiency-oriented) communicative instruction
- Positive washback of assessment practices on instruction

Advantages

- Assesses the ability to use language to achieve ends in social situations (Mislevy et al., 2002)

- Assesses the integration of topical, social, and pragmatic knowledge with knowledge of the formal elements of language

- Aligns with Communicative Language Teaching and Task-based Language Instruction in Foreign Language Education

- Enables meaningful integration of teaching, learning, and assessment (Wiggins 1990)

Early Proponents

- **Foreign Service Institute Oral Interview (1950s)**
- Grant Wiggins (1990): Authentic Assessment
- Geoff Brindley (1994): Task-centered Assessment
  - Outcomes-based assessment (1998 and later)
- Lyle Bachman & Adrian Palmer (1996): Language Testing in Practice
- Tim McNamara (1996): Performance Testing
- Michael Long & John Norris (2000): Task-based Assessment
- Robert Mislevy (2002): Evidence-based Design

Shades of Meaning

- The influence of task features on test takers' cognitive processes and resulting performance (Skehan, 1998)

- Shared characteristics such as learner-centeredness, contextualization, and authenticity between assessment and instruction (Chalhoub-Deville, 2001)

- Ability to accomplish specific target tasks in particular communication settings (Long & Norris, 2000)

- Ability to engage in the kinds of activities characteristically encountered in communicative language teaching (Wiggins, 1990)

Evidence-Centered Design

'Data' become 'evidence' only when their relevance to some hypothesis, some inference, **some claim is established**. In task-based language assessment, this means that what we really need to understand first and foremost is the **inferential argument** associated with the assessment. **What is its purpose? What do we want to know, about what students know or can do, in what kinds of situations?** (Mislevy et al., 2002:492-3)

.

The Meaning of Test Results

- Data become evidence only when:

= Test Results attain meaning only when:

- Their relevance to a hypothesis, an inference, a CLAIM has been established.

➢ Toulmin's Argument Model

➢ Bachman/Palmer's (2010) Assessment Use Argument

➢ Mislevy et al's Evidence-Centered Design

Evidence-Centered Design

|	
What is the purpose of the assessment?	Task Analysis
What do we want to know about what students know or can do?	Ability
	Evidence
In what kinds of situations?	Task

Mislevy, Steinberg, & Almond (2002)
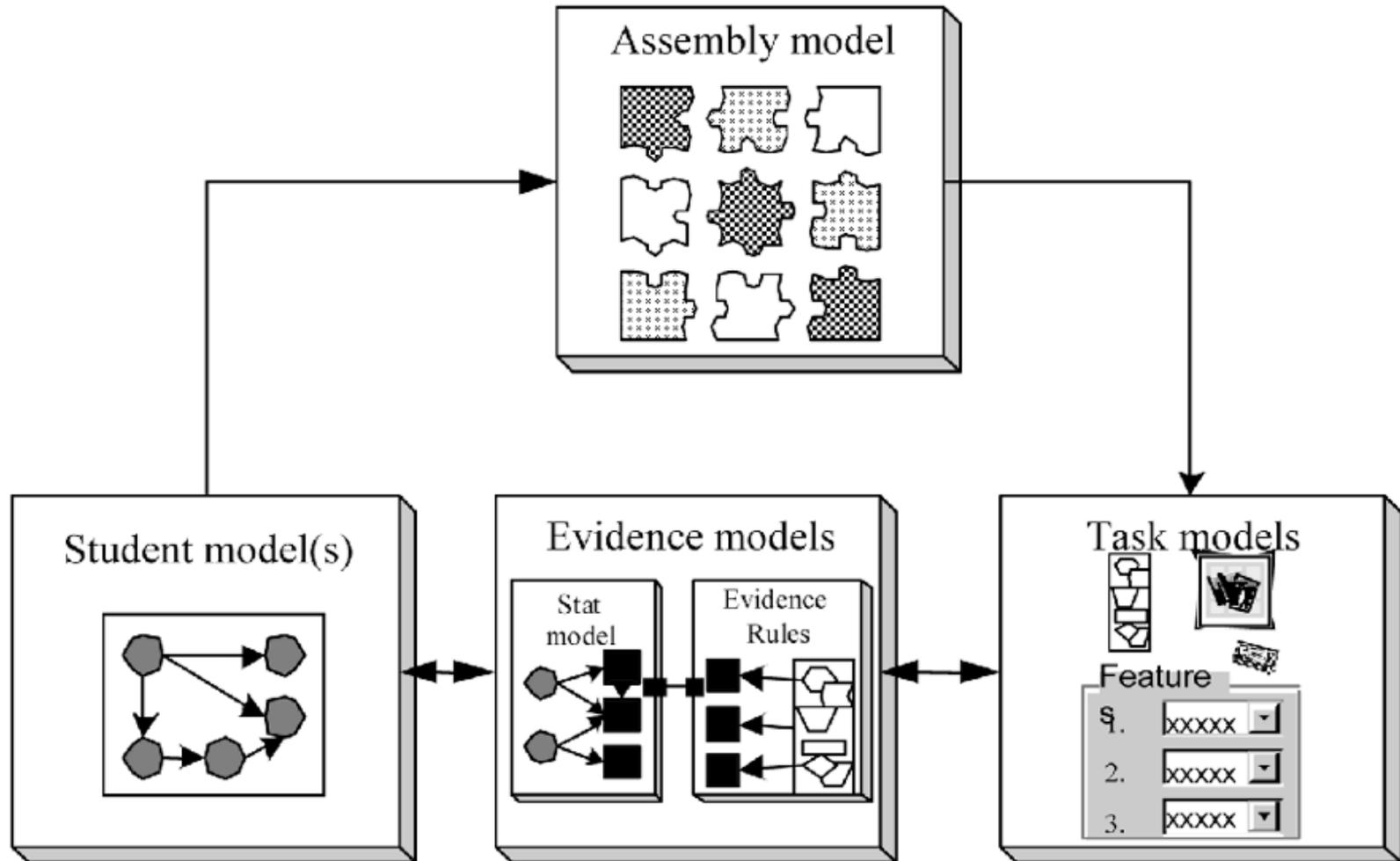
Assembly Model
Combining tasks
to produce an
assessment

Examinee Model
Ability
Knowledge
Skills

Evidence Model
Scoring
Measuring

Task Model
Schemas to elicit
evidence

Examinee Model

- Complex of knowledge and abilities to be assessed

- Unobservable

- Construct-driven (based on SLA theory)
  - Communicative Competence (e.g. Bachman/Palmer 1996)
  - Proficiency Level Descriptions (ILR, ACTFL)
  - Can-do statements (ILR, ACTFL, CEFR)

- Developmental stages/Implicational scaling

Evidence Model

- Observable behaviors that provide sufficient evidence about the knowledge or abilities we wish to measure

- Evaluation Component
  - Extracts the **salient characteristics** of what students say or do (Key aspects of the performance)
  - Based on evaluation rules (rubrics for rating scales)
  - What is valued, and how is it evaluated?

- Measurement Component
  - **Accumulation** of information to update beliefs about examinees: What do the observable variables tell us about the examinee's abilities?
  - How is evidence synthesized across multiple tasks and different performances? (When do you know, e.g., when you have enough?)

UNIVERSITÄT LEIPZIG

Task Model

- Tasks or situations that elicit the behaviors we need as evidence
- Schema for constructing and describing the situations in which examinees act
- Task-driven (based on task/domain/needs analysis)
  - Salient features of tasks
  - How do tasks influence (and constrain) performance?
  - What are key features of **target language uses** (TLUs) in tasks (authenticity)?
- Developmental stages/Implicational scaling
  - **Hierarchy of tasks**

Assembly Model

- Mix and number of tasks that are sufficient for an assessment
  - Determines the range of circumstances that need to be covered (to be able to generalize)
  - Controls the difficulty of tasks
  - Manages what information **accumulates** or does not accumulate

Assembly Model
Checks / Probes
Range of Topics
Range of Tasks
Kinds of Tasks

Examinee Model
Pragmatic Comp.
Socio-ling. Comp.
Text Competence
Gram./Lex. Comp.

Evidence Model
Success/Failure
Quality: How well?
Quantity: How much?

Task Model
Global Tasks
Context/Content
Role of Interlocutor

UNIVERSITÄT LEIPZIG

Speech Sample

- Provides **cumulative evidence** of the examinee's language ability

- Representative of the examinee's language ability
  - Functions, Textual organization, Register, Cultural competence
  - Variety of topics and contexts
  - Concern for affective schemata (warm-up etc.)

- Indicative of the examinee's control over the above
  - Random, emerging, developing, sustained

Task-centered and Construct-centered

Task-centered perspective

- Features of language-use situations that reveal:

- The language-use competences that are of interest; and

- The kinds of performances that contain evidence about language-use competences

Construct-centered perspective

- What performances in what situations tell us what about student abilities?

The Construct of the OPI

Communicative Competence

Language Ability (Bachman/Palmer 1996)

- Organizational ability
  - Grammatical ability (sentence level)
  - Textual ability (text level)
- Pragmatic ability
  - Functional ability
    - Ideational, instrumental, heuristic, imaginative
  - Sociolinguistic ability
    - Conventions, register, cultural references

Language Ability in OPIs

- Functional ability
  - Instrumental (negotiating daily interactions)
  - Ideational (describing and reporting)
  - Imaginative (narrating and hypothesizing)
  - Heuristic (explaining and arguing a point)
- Textual ability
  - Word- and sentence-length utterances
  - Connected texts of various complexity
- Sociolinguistic ability
  - Register, cultural competence

Developmental Hierarchies (ILR/ACTFL)

- Functional ability
  - Asking for and providing simple personal information
  - Describing and reporting
  - Narrating and explaining
  - Arguing and hypothesizing

- Sociolinguistic ability
  - None
  - One register
  - Two or more registers
  - High level of acculturation (control of conventions)

Developmental Hierarchies (ILR/ACTFL)

- Organizational ability
- Use of organizational types (text types)
  - Strings of words
  - Sentence-length utterances
  - Strings of sentences
  - Connected speech of various lengths (cohesion)

Development of organizational ability I

- **Gillian Brown** (1994): Modes of understanding
  - Hierarchies of cognitive load
- Identifying: **words**
- Procedural understanding: one **sentence** at a time
  - Fully supported by the external world; Partial understanding is sufficient; Immediate feedback
- Narrative understanding: **texts**
- Understanding of argument: **complex texts**

Development of organizational ability II

- **Manfred Pienemann** (1998): Processability theory
  - Grammatical memory: feature unification
- Lemma access: single words, no features   **Words**
  - Listing single words or expressions; no sense for syntax or morphology
- Category procedure
- Phrasal procedure
- Sentence procedure   **Sentences**
- Subordinate clause procedure   **Texts**

The Bigger Picture

The Assessment Use Argument

Test Development

Consequences

Decisions

Interpretations

Assessment Records
(Scores, Descriptors)

Performance

Task

Interpretation and Use
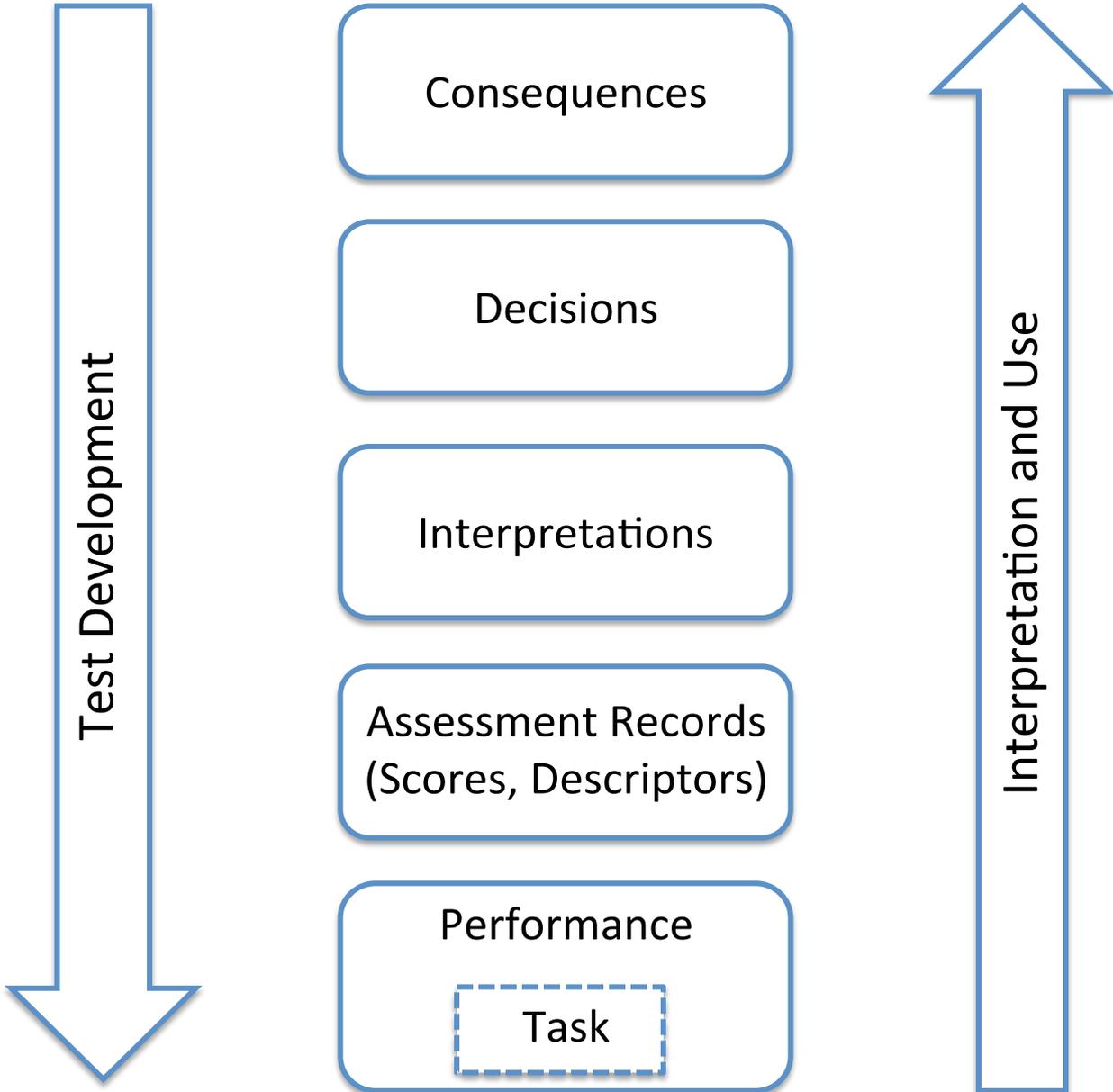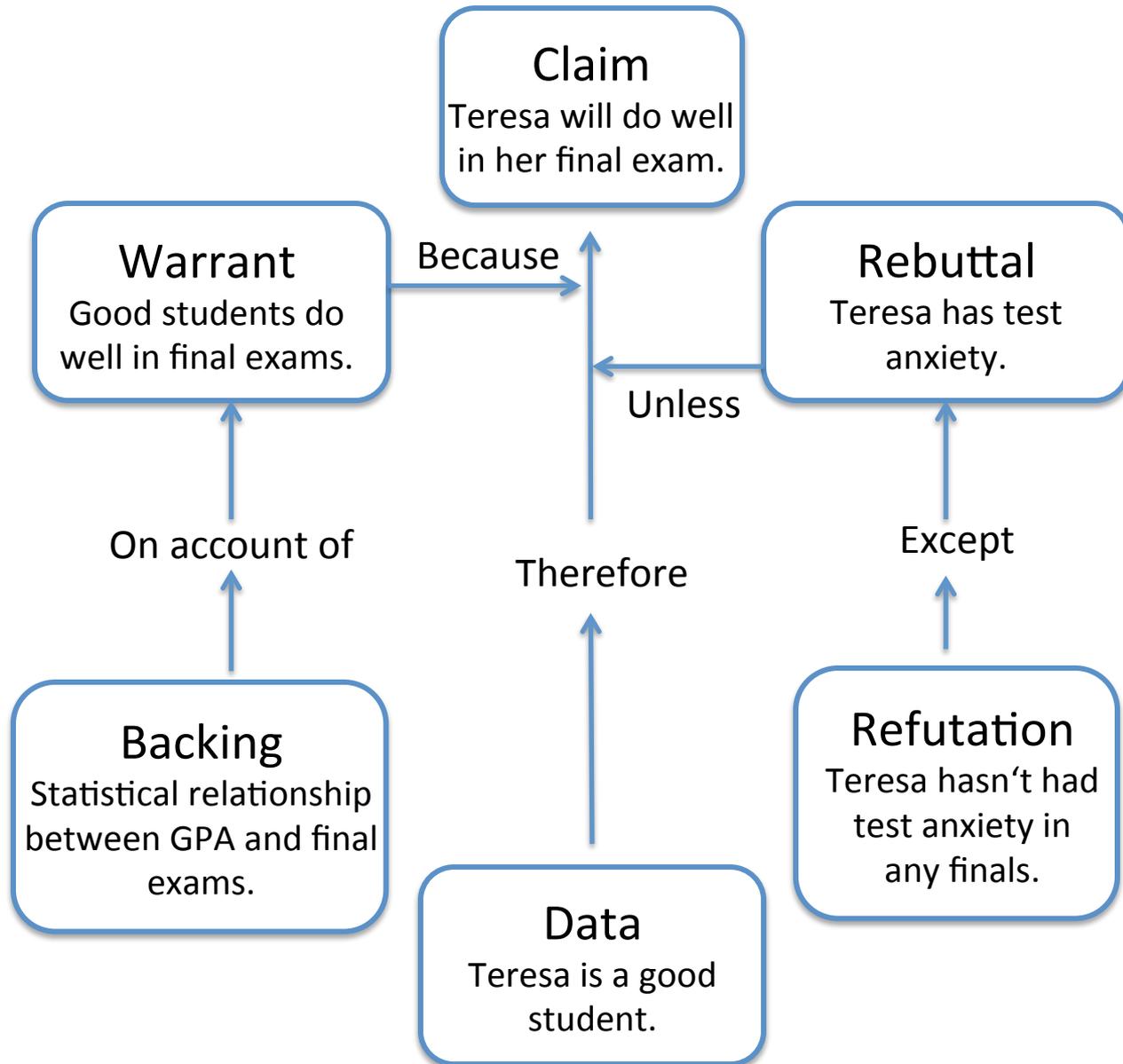
Argument Approach

Warrants

- Explicit
- Generalizable
- Provide initial support that the connection between the data and the claim is appropriate and legitimate

Backing

- Provides the evidence for the warrant
- Laws, rules, principles, facts
- Widely accepted assumptions or shared experience

Claim 4

↑ Warrants
Rebuttals

Data

- -

Claim 3

↑ Warrants
Rebuttals

Data

- -

Claim 2

↑ Warrants
Rebuttals

Data

- -

Claim 1

↑ Warrants
Rebuttals

Data
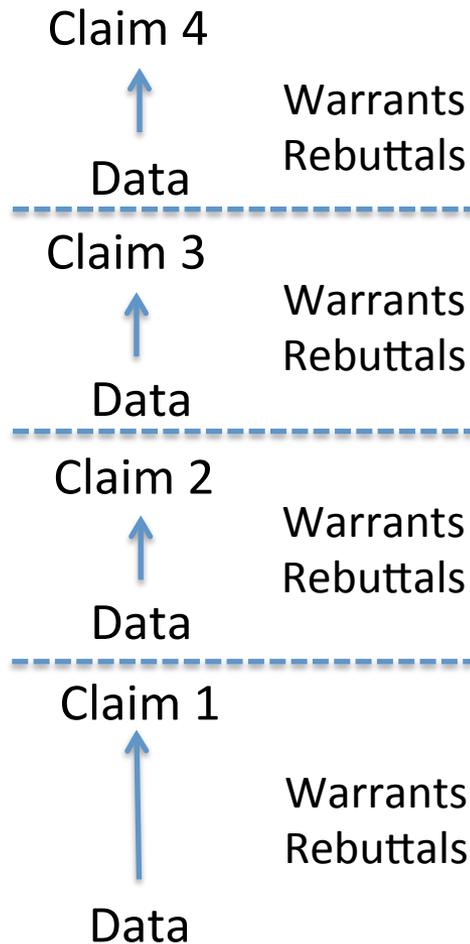
Consequences:
Beneficial

Decisions:
Values sensitive, Equitable

Interpretations:
Meaningful, Generalizable,
Impartial, Relevant, Sufficient

Assessment Records:
Consistent

Performance

Task

Example: Interpretations

- General warrant: The assessment record can be interpreted meaningfully.

- Specific warrant: The rating reflects what the examinee is able and not able to do linguistically.

- Backing: The construct of the assessment, i.e. proficiency, is described in comprehensive can-do statements.

- Rebuttal: The construct was developed experientially and is not theory-based.

- Repudiation: The construct aligns well with current SLA theories.

The Evolution of the (ACTFL) OPI

High Reliability / Quality Assurance

- Substantive and detailed Testing Protocol

- Substantive and detailed Rating Protocol

- Extensive tester training and certification procedures

- Extensive ongoing tester and rater norming

- Blind double ratings and arbitration

- Quality-controlled operational testing and rating (LTI)

What's Missing?

- … the evaluation of language use for expressing meaning within a well-defined communicative context and audience, for a clear purpose … (Norris, 2014)

- … meaningful integration of teaching, learning, and assessment … (Wiggins, 1998); … positive washback of assessment practices on instruction … (Mislevy et al., 2002)

Task Analysis
Domain Analysis
Interpretation

Curricular Tie-in
Consequences

Some Best Practice Examples

- Revised FSI Speaking Test (2014)
  - Task Analysis
- OPI+ (NSW) (2016)   Marla Federe (2016)
  - Task Analysis
  - Curricular Tie-in

Key Take-Aways

- Assessments need to be both, task-based and construct-based to yield evidence about an examinee's abilities
- Current validity/validation theories put in doubt the notion of a general proficiency test
- Current validity/validation theories call for:
  - Domain-specific proficiency tests
    - Even for general language programs in schools/universities
  - Substantial curricular tie-in
- The future of proficiency assessment may be the assessment of Proficiency+
  - Proficiency for a mission/job plus more general domains

Bibliography

- Bachman, L. F., & Palmer, A. S. (1996). *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford: Oxford University Press.
- Bachman, L. F. & Palmer, A. S. (2010). *Language Assessment in Practice: Developing Language Assessments and Justifying their Use in the Real World*. Oxford: Oxford University Press.
- Brindley, G. (1994). Task-centred assessment in language learning: The promise and the challenge. In Bird, N., Falvey, P., Tsui, A., Allison, D. & McNeill, A., eds., *Language and Learning: Papers Presented at the Annual International Language in Education Conference*. Hong Kong: Hong Kong Education Department, 73–94.
- Brown, G. (1994). Modes of understanding. In G. Brown, K. Malmkjær, A. Pollitt & J. Williams, eds., *Language and Understanding*. Oxford: Oxford University Press, 9-20.
- Chalhoub-Deville, M. (2001). Task-based assessments: Characteristics and validity evidence. In Bygate, M., Skehan, P. and Swain, M., eds., Researching Pedagogic Tasks: Second Language Learning, Teaching and Testing. New York: Pearson Education, 210–228.
- Federe, M. (2016, September). Authenticity: From Theory to Reality. Invited Paper. 2016 LEARN Task Based Assessment. Universities at Shady Grove, Rockville, MD.
- Long, M.H. and Norris, J.M. (2000). Task-based language teaching and assessment. In Byram, M., ed., *Encyclopedia of Language Teaching*. London: Routledge, 597–603.

Bibliography

- McNamara, T. (1996). *Measuring Second Language Performance*. New York: Longman.

- Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2002). Design and analysis in task-based language assessment. *Language testing*, *19*, 477-496.

- Norris, J. M. (2014, April). *How do we assess task-based performance?* Invited LARC/ CALPER testing and assessment webinar. Online: https://larc.sdsu.edu/testassesswebinar/jnorris/Norris_Task-BasedAssessment_PPT.pdf

- Pienemann, Manfred (1998). *Language Processing and Second Language Development: Processability Theory*. Amsterdam: John Benjamins.

- Skehan, P. (1998). *A Cognitive Approach to Language Learning*. Oxford: Oxford University Press.

- Toulmin, S. (1958). *The Uses of Argument*. Cambridge: Cambridge University Press.

- Wiggins, G. (1990). *The Case for Authentic Assessment*. ERIC Digest. Online: http:// files.eric.ed.gov/fulltext/ED328611.pdf.