Mapping ILR Skill Level Descriptions to an Assessment Metric for the DLPT5

Heeyeon Y. Dennison, Ph.D.

Test Development Division

DISCLAIMER

- This presentation is authorized by the Defense Language Institute Foreign Language Center and the Department of Defense. Contents of this presentation are not necessarily the official views of, or endorsed by, the U.S.Government, or the Department of the Army.

- All material displayed within this presentation is for educational purposes only.

- All third party information featured in the presentation slides remain the intellectual property of their respective originators. All use of information is done under the fair use copyright principal, and the author(s) of this presentation do not assert any claim of copyright for any quotation, statistic, fact, figure, data or any other content that has been sourced from the public domain.

- The content of this presentation is the sole responsibility of the presenter.

Background on DLPT5 (1)

- Defense Language Proficiency Test Generation 5

- General foreign language proficiency test for reading and listening

- For U.S. DoD military and civilian personnel

- Interagency Language Roundtable (ILR) reading and listening skill level descriptions as core criteria for developing test items (http://www.govtilr.org)

Background on DLPT5 (2)

- Items passed through multi-stage reviews and field test analyses are assembled to parallel forms by test development experts

- Web-based test administration

- Test results used to classify examinee proficiency against the ILR standards

- ILR-based proficiency classification generates real world implications on DoD agencies' decision making process

Building a Validity Argument Framework (Kane, 2013)

© 2010 Alpine Testing Solutions, Inc.

# *Accumulating Validity Evidence in DLPT5 Development*

Premises

1) Test items are written to measure specific abilities tied to particular ILR levels

2) Parallel test forms consist of representative items properly evaluated and distributed across the targeted ILR levels

3) Scoring and analysis procedure meets psychometric industry standards

Current verification methods

➤ Holistic judgments from various ILR experts at multiple stages of item development

➤ Field tests/ TAP, Item analysis and selection using parameters from Classical Test Theory and Item Response Theory

➤ Standard-setting studies, Psychometric analyses, Consultation with various technical and policy groups

Example: Items' Intended ILR levels by Empirical Item Difficulty (b)
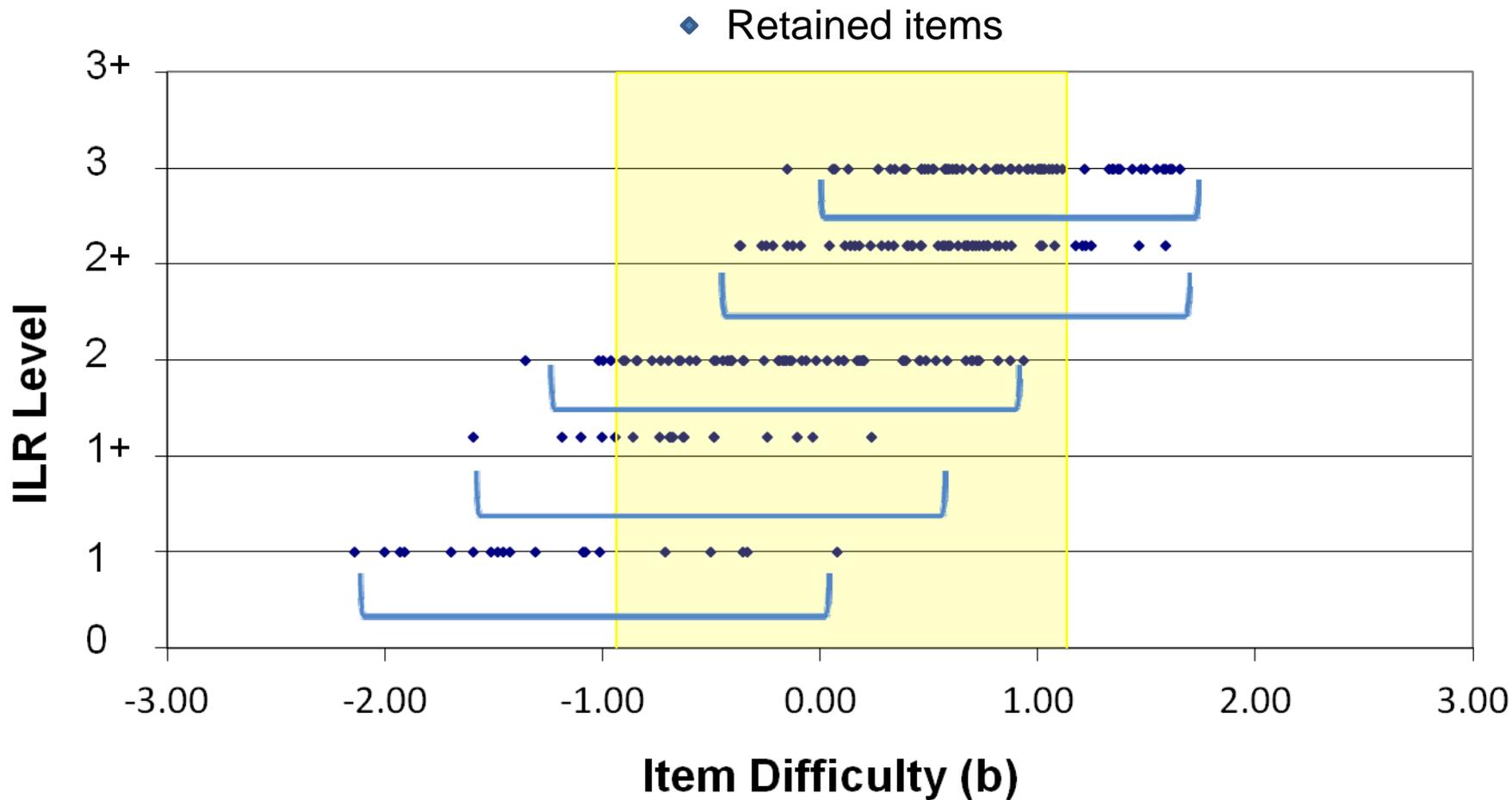
Example: Expected Proportion Correct by Proficiency (Theta)
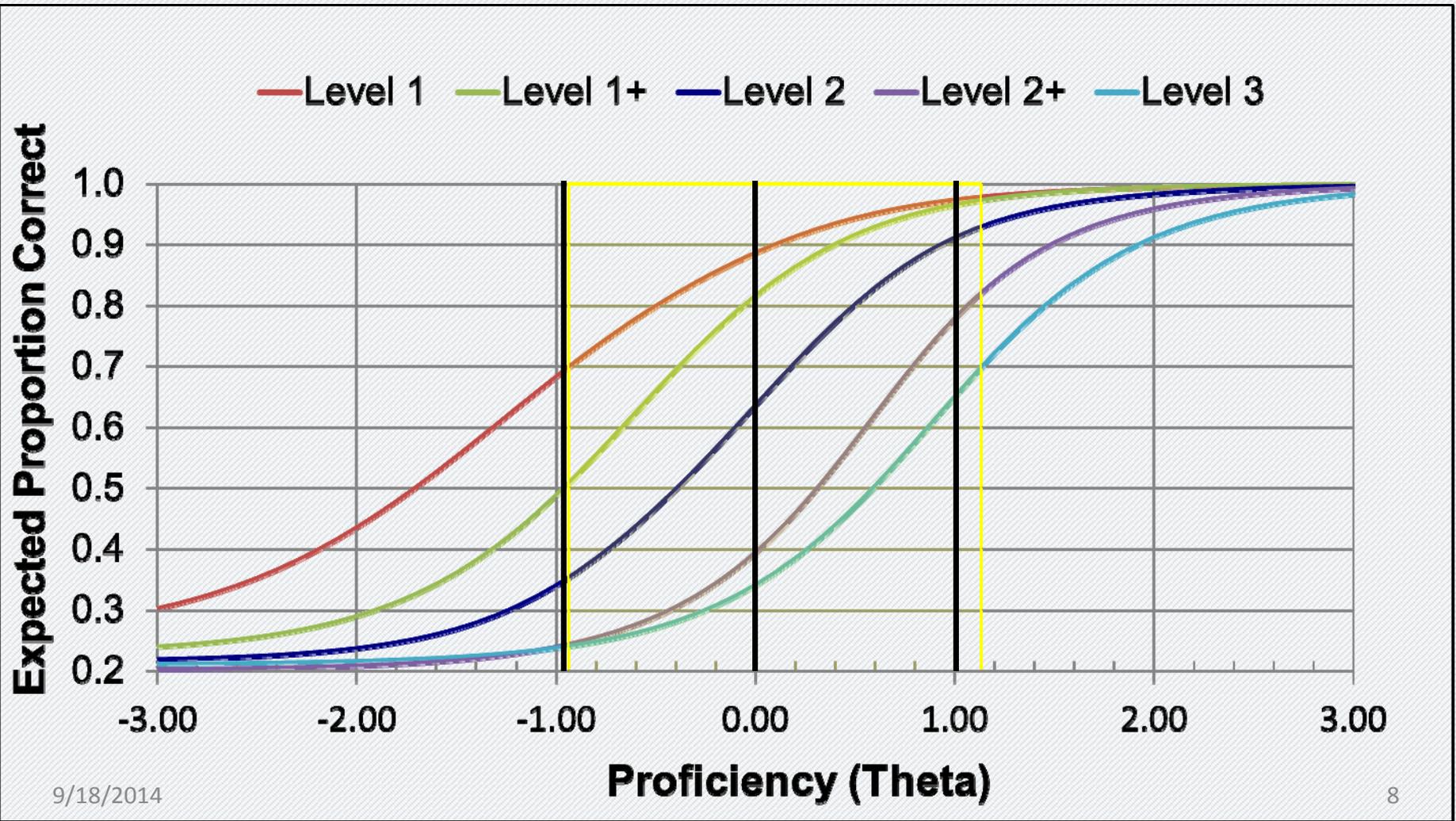
Needs Identified for DPLT5 (1)

- Detailed item bank specifications
  - To better understand the relationship between items' theoretical correlates and item difficulty parameters (e.g., Downing & Haladyna, 2006)
    - Building a validity argument framework in support of test score interpretations and uses (e.g., Bachman & Palmer, 2010; Chapelle, Enright, & Jamieson, 2008; Kane, 2013)

Needs Identified for DPLT5 (2)

- Detailed item bank specifications
  - To move toward automatically generated test forms with precise content balancing (Computerized Adaptive Testing, CAT; e.g., Wainer, 1990)

Example of CAT Logic (Wainer, 1990; p. 108)

1. Begin with Provisional Proficiency Estimate

2. Select & Display Optimal Test Item

3. Observe & Evaluate Response

4. Revise Proficiency Estimate

5. Is Stopping Rule Satisfied?

NO

YES

6. End of Test

7. End of Battery?

NO

YES

8. Administer Next Test

9. Stop
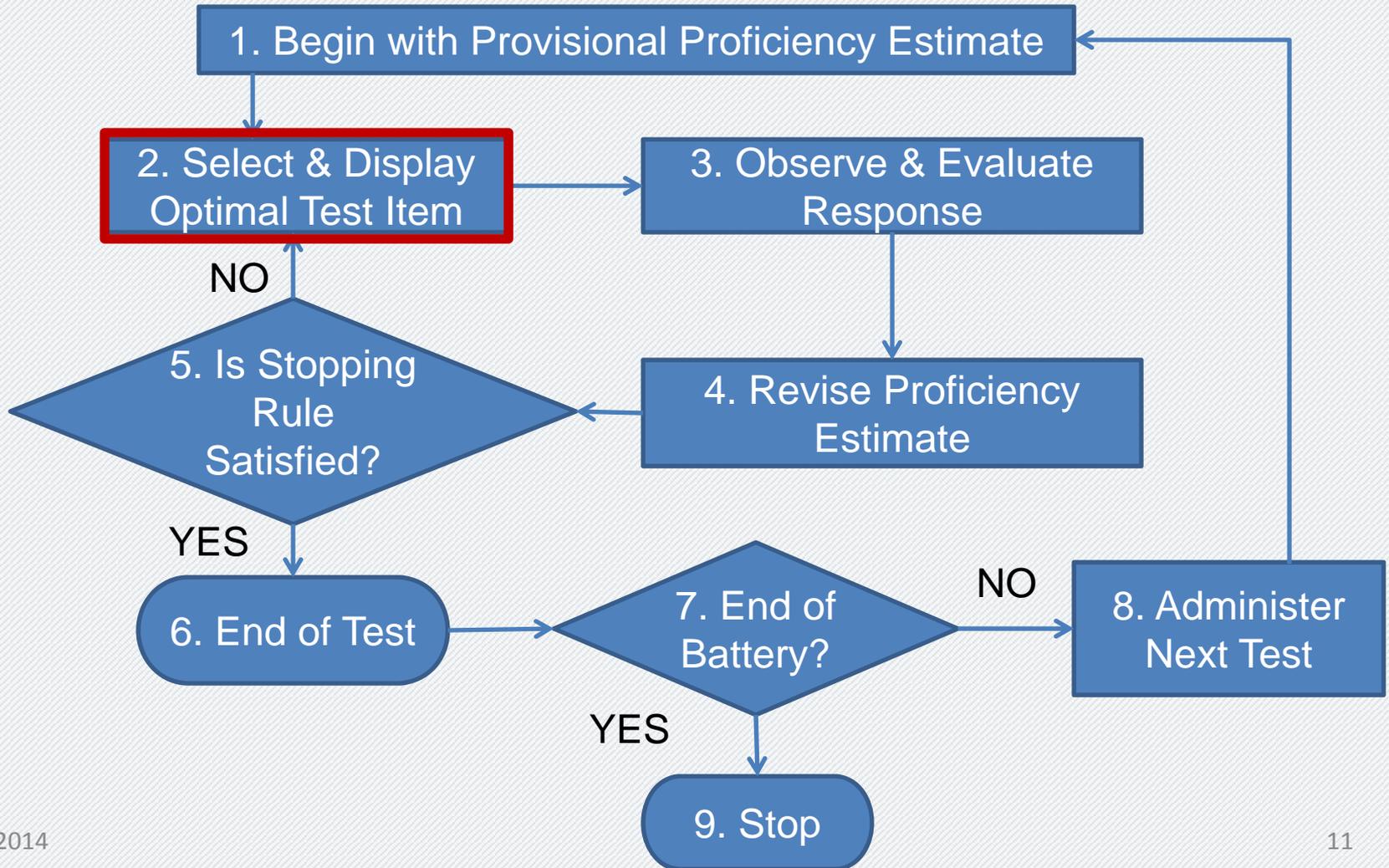
Goal of This Presentation

I. To introduce the Item Bank Specification (IBS) metric (work in progress)

- Goals and merits
- Methods

II. To discuss the implications of the IBS metric

- on accumulating evidence for building a validity argument framework
- on fulfilling emerging business needs, especially Computerized Adaptive Testing

Item Bank Specification (IBS) Metric Goals and Merits (1)

- **By quantifying test item characteristics**

  1) Verify that test items measure specific and appropriate abilities at each proficiency level

  2) Evaluate and improve internal consistency and reliability of ILR experts' item judgments during item construction and review

  3) Investigate relationship between item characteristics and empirical item difficulty

Item Bank Specification (IBS) Metric Goals and Merits (2)

- **By quantifying test item characteristics**

  4) Economize the item replenishment process

  5) Achieve detailed content balancing during automatic test form construction process, i.e., Computerized Adaptive Testing (CAT)

DLPT5 ITEM BANK SPECIFICATION (IBS) METRIC METHODS – OVERVIEW

Item Bank Specification (IBS) Metric (work in progress)

- **Two methods** incorporated:

1) **A rubric** – to specify test construct based on Text, Task, and Ability along the ILR base levels from 1 to 4

2) **Metadata inventories** – to specify each item's characteristics in terms of variables

Item Bank Specification (IBS) Metric – Rubric

ILR	Text	Task	Ability
1	Lay out major characteristics of target language stimulus (written passage or audio)	Capture comprehension task each test question requires examinees to tackle	Describe examinees' expected ability needed to process given text and task - Focus on language knowledge
2			
3			
4			

- Express general language proficiency based on the interplay among Text, Task, and Ability (DLPT5 Framework, 2009; DLPT5 Test specifications, 2006)
- Focus on ILR base levels from 1 to 4

Item Bank Specification (IBS) Metric – Metadata Inventories

Item ID	VAR related to TEXT						VAR related to TASK						VAR related to Ability					
	V1	V2	…				…			…			…			…		
XX5L0001																		
XX5L0002																		
XX5L0003																		
XX5L0004																		
XX5L0005																		
XX5L0006																		

- Exhibit each item's characteristics through a combination of variables
- Variables: theoretical correlates as well as technical specs
- Enable mapping between item characteristics and ILR level assignment for each item

DLPT5 ITEM BANK SPECIFICATION (IBS) METRIC METHODS – DETAILS

Item Bank Specification (IBS) Metric – Rubric

ILR	Text	Task	Ability
1			
2			
3			
4			

The IBS rubric – Text
(target language stimulus)

ILR	Topical domain	Mode	Type	Lexical range	Syntactic characteristics	Organizational characteristics	Length
1							
2							
3							
4							

- Based on:
  - DLPT5 Framework (2009), DLPT5 test specifications (2006)
  - ILR skill level descriptions for reading and listening
  - Documents on text typology and mode, e.g., Child (1987)

The IBS rubric – Ability

- Bachman & Palmer's (2010) framework: Language Ability = Language Knowledge + *Strategic Competence*
  - ➢ *a set of metacognitive strategies*
  - ➢ *Purpura (1999) also includes cognitive strategies*
- Buck's (2001) view on second-language testing for adult learners:
  - ➢ More emphasis on testing Language Knowledge (LK) rather than testing Strategic Competence
  - ➢ LK includes procedural and declarative knowledge

The IBS rubric – Ability
(expected language knowledge)

ILR	Grammatical aspects: Vocabulary Morphology Phonology	Grammatical aspects: Syntax Semantics	Discourse aspects:	Pragmatic aspects:	Sociolinguistic aspects:
			e.g., Cohesion, Rhetorical/ Conversation structure	e.g., Illocutionary force, Presupposition & entailment, Inference, Implicature	Language knowledge related to particular socio-cultural settings (e.g., register, dialects, idiomatic expressions, cultural references and figure of speech)
1					
2					
3					
4					

- Focuses on examinees' expected language knowledge
- Based on:
  - DLPT5 Framework (2009), DLPT5 test specifications (2006)
  - ILR skill level descriptions for reading and listening
  - E.g., Alderson (2000), Bachman & Palmer (2010), Buck (2001), Fromkin et al. (2011), O'Grady (2010), Yule (1996)

The IBS rubric – Task

- **Language use**: a cognitive task by default
  - Abundant evidence from both behavioral and brain research (e.g., Spivey, 2008)

- **Language proficiency test**: Focus on procedural knowledge of language
  - e.g., Bachman & Palmer (2010), Buck (2001)

- **Assessment task**: cognitive task to handle information conveyed through language

Anderson et al.'s (2001) revised Bloom's taxonomy on educational objectives

The knowledge dimension

The cognitive process dimension

	Remember	Understand	Apply	Analyze	Evaluate	Create
Factual k.						
Conceptual k.						
Procedural k.						
Meta-cognitive k.						

Focus on knowledge transfer

Focus on knowledge retention
(i.e., retrieving relevant knowledge
from long term memory)

The IBS rubric – Task
(expected cognitive task)

ILR	Major cognitive process	Specific cognitive process	Operational definition

- Based on Anderson et al.'s (2001) revised Bloom's taxonomy on educational objectives
- Frame what examinees are asked to do for each test question in terms of cognitive processes required by the assessment task

The IBS rubric – TASK

ILR	Major cognitive process	Specific cognitive process	Operational definition
1 2 3 4	Understand	Paraphrasing Summarizing Exemplifying Classifying Comparing Explaining	
3 4	Analyze	Differentiating Organizing Attributing Inferring*	
3 4	Evaluate	Checking Critiquing	

The IBS Metadata Inventories (1)

Examples of variables related to Text:

- Final Learning Objectives (FLO) content areas
- Topic   Mode   Text type
- Length in Target Language   Length in English Rendering
- Text font   Character type
- Lexical range   Syntactic complexity
- Organizational characteristics
- Authenticity
- Speech setting   Speech register   Speech rate
- Number of speakers   Gender of speakers
- Region of language spoken   Dialect
- Background noise

The IBS Metadata Inventories (2)

Example	VAR related to TEXT (Reading)					
Item ID	FLO	Mode	Text Type	Length	Lexical range	Syntactic complexity
XX5L0001	3	2	1	115		
XX5L0002	1	4	3	398		
XX5L0003	2	1	1	60		
XX5L0004						
XX5L0005						
XX5L0006						

- Variables of different types (e.g., nominal, ordinal, ratio, continuous)
- A combination of variables can signal the most appropriate ILR level for each item

The IBS Metadata Inventories (3)

- Work in progress in terms of defining theoretical correlates of language complexity and ability (esp., language knowledge)

- Examples for further consideration:
  - The notion of the **necessary information:**
    "the information in the text that the test-taker must understand in order to be sure the task has been done correctly" (Buck, 2001: p.129)
  - Examples of associated variables:
    - Location of the necessary information
    - Repetition of the necessary information
    - Lexical characteristics of the necessary information
    - Syntactic characteristics of the necessary information

General Discussion

- Need to gather theoretical correlates of item difficulty parameters that are clear and operational (e.g., Graesser et al., 2004)
  - Launch of the IBS Task Force

- Needs for generality to cover language comprehension regardless of modality (reading vs. listening) or response type (e.g., multiple choice vs. constructed response)

- Needs for specificity to handle modality and response types

The IBS Metric Implications (1)

- By quantifying item characteristics, we can answer various questions about the composition of an item pool.
  - E.g., What % of Language X Lower Range DLPT5 multiple choice items assess examinees' ability to summarize information presented in the text?

- We can easily identify what's sufficient or insufficient in the item pools.
  - Economize the item replenishment process

The IBS Metric Implications (2)

- We can better examine the relationship between items' intended ILR levels and empirical item difficulty.

- We can evaluate and improve internal consistency and reliability of ILR experts' item judgments during item construction and review.

The IBS Metric Implications (3)

- We can provide detailed content balancing for Computerized Adaptive Testing.

Conclusion

- In support of the US Defense Foreign Language Program, the IBS metric will enable:
  - close examination of the ILR-based item characteristics toward building a validity argument framework to support test score interpretations and uses
  - construction of Computerized Adaptive Testing with precise content balancing

References (1)

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., et al. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives.* New York: Longman.

Bachman, L., & Palmer, A. (2010). Language Assessment in Practice. New York: Oxford University Press.

Buck, G. (2001). *Assessing listening.* Cambridge: Cambridge University Press.

Chapelle, C. A., Enright, M. K., & Jamieson, J. M. (2008). Building a validity argument for the Test of English as a Foreign Language. New York: Routledge.

Child, J. R. (1987). Language proficiency levels and the typology of texts. (H. Byrnes, & M. Canale, Eds.) *Defining and developing proficiency: Guidelines, implementations and concepts*, 97-106.

Downing, S. M., & Haladyna, T. M. (2006). *Handbook of test development.* New York: Routledge.

Fromkin, V., Rodman, R., & Hyams, N. (2011). An Introduction to Language. Boston: Wadsworth.

References (2)

Graesser, A., McNamara, D.S., Louwerse, M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavioral Research Methods, Instruments, and Computers*, 36, 193-202.

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, Vol.50 (1), pp. 1–73.

O'Grady, W. (2004). *The syntax files: An introductory survey of basic syntactic concepts and phenomena* (Vol. 10). Honolulu: University of Hawaii at Manoa.

Purpura, J. E. (1999). *Modeling the relationships between test takers' reported cognitive and metacognitive strategy use and performance on language tests.* Cambridge: University of Cambridge Local Examinations Syndicate and Cambridge University Press.

Retrieved September 12, 2014, from Interagency Language Roundtable: http://www.govtilr.org/

Spivey, M. J. (2008). *The continuity of mind.* New York: Oxford University Press.

References (3)

TED-RE. (n.d.). The text modes. Monterey: DLIFLC Test Review & Education, Monterey, California.

TED-TD. (2006, February 9). DLPT5 Lower Range Multiple-Choice Test Specifications. DLIFLC Test Development, Monterey, California.

TED-TD. (2009, September). Defense Language Proficiency Testing System 5 Framework. Monterey, California.

Yule, G. (1996). *Pragmatics.* Oxford: Oxford University Press.

Wainer, H. (1990). *Computerized Adaptive Testing: A Primer.* Hillsdale: Lawrence Erlbaum Associates, Inc.

Thank you!

Q & A