



PLENARY 2

Current Developments in ELT Testing

Dr Thom Hudson
University of Hawaii

Abstract

Three current developments in second and foreign language testing that reflect a common concern are the development of scales of proficiency, task-based performance assessment, and criterion referenced testing. Scales such as the Common European Framework of Reference for Languages developed by the Council of Europe, and the ACTFL Guidelines developed by the American Council on the Teaching of Foreign Languages reflect a desire to have language ability linked to descriptions of ability rather than to purely numerical rankings. This presentation will describe certain developments in each of these areas and indicate how the interrelationships are related to concerns of validity in test score use. Task-based performance assessment strives to move beyond indirect indications of a language learner's ability in order to reveal what the learners can do with the knowledge they have acquired beyond a simple display of knowledge. Criterion referenced testing is designed to assess the learners' knowledge of a well defined domain of knowledge. This presentation will try to situate these three developments as concerned with construct validity.

Summary of the Paper

1. Validity

Messick (1989) redefined validity as follows: *“an integrated evaluative judgement of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment.”* (p.13)

Figure 1: Messick’s Facets of validity (1989)

	Test Interpretation	Test use
Evidential basis	1 Construct validity	2 Construct validity & Relevance/Utility
Consequential basis	3 Value implications	4 Social consequences

2. Scales

Table 1: FSI descriptors for Level 2 speaking

<p>Speaking 2</p> <p><u>Able to satisfy routine social demands and limited work requirements.</u> Can handle with confidence but not with facility most social situations including introductions and casual conversations about current events, as well as work, family, and autobiographical information can handle limited work requirements, needing help in handling any complications or difficulties; can get the gist of most conversations on non-technical subjects (i.e. topics which require no specialized knowledge, and has a speaking vocabulary sufficient to respond simply with some circumlocutions; accent, though often quite faulty, is intelligible; can usually handle elementary constructions quite accurately but does not have thorough or confident control of grammar.</p> <p><u>Examples:</u> While these interactions will vary widely from individual to individual, the individual can typically ask and answer predictable questions in the workplace and give straightforward instructions to subordinates. Additionally, the individual can participate in personal and accommodation-type interactions with elaboration and facility; that is, can give and understand complicated, detailed, and extensive directions and make non-routine changes in travel and accommodation arrangements. Simple structures and basic grammatical relations are typically controlled; however, there are areas of weaknesses. In the commonly taught languages, these may be simple markings such as plurals, articles, linking words, and negatives or more complex structures such as tense/aspects usage, case morphology, passive constructions, word order, and embedding.</p>
--

FSI (n.d.) (p.14)

Table 2: ACTFL Advanced descriptor

<p>Advanced: Sufficient comprehension to understand conversations about routine social conventions and limited school or work requirements. Able to understand face-to-face speech in the standard language, delivered at a normal rate with some repetition and rewording by a native speaker not used to dealing with foreigners. Understands everyday topics, common personal and family news, well-known current events, and routine matters involving school or work descriptions and narration about current, past and future events; and essential points of discussion or speech at an elementary level on topics in special fields of interest.</p>
--

ACTFL (1989) (P.221)

Table 3: Sample of Bachman and Palmer decontextualised scale

Grammatical competence		
Rating	Range	Accuracy
0	No systematic evidence of morphological and syntactic structures	Controls over a few or no structures; errors of all or most possible types
1	Limited range of both morphologic and syntactic structures, but with some systematic evidence	Control of few or no structures; errors of all or most possible types
2	Limited range of both morphologic and syntactic structures, but with some systematic evidence	Control of some structures used but with many error types
3	Large, but not complete, range of both morphologic and syntactic structures	Control of some structures used but with many error types
4	Large, but not complete, range of both morphologic and syntactic structures	Control of most structures used, with few error types
5	Complete range of morphologic and syntactic structures	Control of most structures used, with few error types
6	Complete range of morphologic and syntactic structures	No systematic errors

Pragmatic competence		
Rating	Vocabulary	Cohesion
0	Extremely limited vocabulary (A few words and formulaic phrases. Not possible to discuss any topic, due to limited vocabulary)	No cohesion (Utterances completely disjointed, or discourse too short to judge)
1	Small vocabulary (Difficulty in talking with examinee because of vocabulary limitations)	Very little cohesion (Relationships between utterances not adequately marked; frequent confusing relationships among ideas)
2	Vocabulary of moderate size (Frequently misses or searches for words)	Moderate cohesion (Relationships between utterances not adequately marked; frequent confusing relationships among ideas)
3	Large vocabulary (Seldom misses or searches for words)	Good cohesion (Relationships between Utterances well-marked)
4	Extensive vocabulary (Rarely, if ever, misses or searches for words. Almost always uses appropriate words)	Excellent cohesion (Uses a variety of appropriate devices; hardly ever confusing relationships among ideas)

Adapted from Bachman (1990)

Table 4: Interagency Language Roundtable Levels and Selected Contexts – SPEAKING

Speaking 0	No proficiency	
Speaking 0+	Memorised proficiency	Dealing with requests
Speaking 1	Elementary proficiency	Getting facts over telephone
Speaking 1+	Elementary proficiency, plus	Making arrangements
Speaking 2	Limited working proficiency	Eliciting and informed opinion
Speaking 2+	Limited working proficiency, plus	Defending a point of view
Speaking 3	General professional proficiency	
Speaking 3+	General professional proficiency, plus	
Speaking 4	Advanced professional proficiency	
Speaking 4+	Advanced professional proficiency, plus	
Speaking 5	Functionally native proficiency	

Table 5: The ILR to ACTFL concordance

ILR	ACTFL
Levels 0-0+	Novice – Low Novice-Mid Novice-High
Level 1	Intermediate-low Intermediate-mid
Levels 1+	Intermediate-high
Level 2	Advanced
Level 2+	Advanced plus
Levels 3-5	Superior

Table 6: Organisation of Canadian Language Benchmark components

AN OVERVIEW				
Benchmark	Proficiency level	Speaking & listening competencies	Reading competencies	Writing competencies
STAGE I: BASIC PROFICIENCY				
1	Initial	Creating/interpreting oral discourse in routine non-demanding contexts of language use in: <ul style="list-style-type: none"> • Social interactions • Instructions • Suasion (getting things done) • Information 	Interpreting simple texts: <ul style="list-style-type: none"> • Social interaction texts • Instructions • Business/service texts 	Creating simple texts: <ul style="list-style-type: none"> • Social interactions • Recording information • Business/service messages • Presenting information
2	Developing			
3	Adequate			
4	Fluent			
STAGE II: INTERMEDIATE PROFICIENCY				
5	Initial	Creating/interpreting oral discourse in moderately demanding contexts of language use in: <ul style="list-style-type: none"> • Social 	Interpreting moderately complex texts: <ul style="list-style-type: none"> • Social interaction texts 	Creating moderately complex texts: <ul style="list-style-type: none"> • Social interactions • Recording
6	Developing			
7	Adequate			

8	Fluent	interactions <ul style="list-style-type: none"> • Instructions • Suasion (getting things done) • Information 	<ul style="list-style-type: none"> • Instructions • Business/service texts • Information texts 	information <ul style="list-style-type: none"> • Business/service messages • Presenting information
STAGE III: ADVANCED PROFICIENCY				
9	Initial	Creating/interpreting oral discourse in very demanding contexts of language use in: <ul style="list-style-type: none"> • Social interactions • Instructions • Suasion (getting things done) • Information 	Interpreting complex and very complex texts: <ul style="list-style-type: none"> • Social interaction texts • Instructions • Business/service texts • Information texts 	Creating complex and very complex texts: <ul style="list-style-type: none"> • Social interactions • Recording information • Business/service messages • Presenting information/ ideas
10	Developing			
11	Adequate			
12	Fluent			

(adapted from Pawlidowska-Smith, 2002)

Table 7: Performance conditions - Canadian Language Benchmarks (Pawlidowska-Smith, 2002)

Speaking : Stage II Benchmark 6
<p>Performance conditions:</p> <ul style="list-style-type: none"> • Interaction is face-to-face, or on the phone, with familiar and unfamiliar individuals and small informal groups. • Rate of speech is slow to normal • Context is familiar, or clear and predictable • Context is moderately demanding (e.g., real world environment, limited support from speakers) • Circumstances range from informal to more formal • Setting or content is familiar, clear and predictable • Topic is concrete and familiar • Presentation is informal or formal • Use of pictures or other visuals • Presentation is five to seven minutes long <p>Interactions one-on-one</p>

- Interactions are face-to-face or on the phone
- Interactions is formal or semi-formal
- Learner can partially prepare the exchange

Interactions in a group

- Interaction occurs in a familiar group of three to five people
- Topic or issue is familiar, non-personal, concrete
- Interaction is informal or semi-formal

Table 8: An overview of Speaking Benchmarks-Global performance descriptors

First Benchmark for each stage (Pawlidowska-Smith, G. (2002) www.language.ca)

B.1 Learners can speak very little, mostly responding to basic questions about personal information and immediate needs in familiar situations. Speaks in isolated words or strings of 2 to 3 words. Demonstrates almost no control of basic grammar structures and verb tenses. Demonstrates very limited vocabulary. No evidence of connected discourse. Makes long pauses, often repeats the other person's words. Depends on gestures in expressing meaning and may also switch to first language at times. Pronunciation difficulties may significantly impede communication. Needs considerable assistance.

Sample tasks: Hello, how are you? My name is X. Please come in, wait. Please sit down. Excuse me, Bob. Help me please. Answer questions about basic personal information in short interviews with teachers, other learners and counsellors.

B5. Learners can participate with some effort in routine social conversations and can talk about needs and familiar topics of personal relevance. Can use a variety of simple structures and some complex ones, with occasional reductions. Grammar and pronunciation errors are frequent and sometimes impede communications. Demonstrates a range of common everyday vocabulary and a limited number of idioms. May avoid topics with unfamiliar vocabulary. Demonstrates discourse that is connected (and, but, first, next, then, because) and reasonably fluent, but hesitations and pauses are frequent. Can use the phone to communicate simple personal information; communication without visual support is still very difficult for him/her.

Sample Tasks: Respond to small talk comments. Express and respond to compliments and congratulations. Extend an invitation for a coffee, dinner, party. Direct a person to a place with or without maps, diagrams, sketches. Request permission to leave work early or take a day off.

B9. Learners can independently, through oral discourse, obtain, provide, and exchange key information for important tasks (work, academic, personal) and complex routine and a few non-routine situations in some demanding contexts of language use. Can actively and effectively participate in 30-minute formal exchanges about complex, abstract, conceptual, and detailed information and ideas to analyse, to problem-solve, and to make decisions. Can make 15- to 30-minute prepared formal presentations. Can interact to coordinate tasks with others, to advise or persuade (e.g., to sell or recommend a product or service), to reassure others, and to deal with complaints in one-to-one situations. Grammar, vocabulary, or pronunciation errors very rarely impede communication. Prepared discourse is mostly accurate in form, but may often be rigid in its structures/organization and delivery style.

Sample tasks: Convey appropriately respect, friendliness, distance and indifference in a variety of conversations in a variety of contexts. Give complex instructions on familiar first aid and emergency procedures in the workplace. Discuss concerns about your child's progress in school with the child's teacher and school principal.

Table 9. Common European Framework – Global Scale

	C2	Can understand with ease virtually everything heard or read. Can summarise information from different spoken and written sources, reconstructing arguments and accounts in a coherent presentation. Can express him/herself spontaneously, very fluently and precisely differentiating finer shades of meaning even in more complex situations.
	C1	Can understand a wide range of demanding, longer texts, and recognize implicit meaning. Can express him/herself fluently and spontaneously without much obvious searching for expressions. Can use language flexibly and effectively for social, academic and professional purposes. Can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organizational patterns, connectors and cohesive devices.

	<i>B2</i>	Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialization. Can interact with a degree of fluency and spontaneity that makes regular interactions with native speakers quite possible without strain for either party. Can produce clear, detailed text, a wide range of subjects and explain a viewpoint on a topical issue giving advantages and disadvantages of various options.
	<i>B1</i>	Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise while travelling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes and ambitions and briefly give reasons and explanations for opinions and plans.
	<i>A2</i>	Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g. very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need.
	<i>A1</i>	Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce him/herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has. Can interact in simple way provided the other person talks slowly and clearly and is prepared to help.

Council of Europe (<http://culture2.coe.int/portfolio/documents/052183136text.pdf>)

Table 10: Example descriptors for the CEF (from North, 2000)

C1:

Can communicate spontaneously, often showing remarkable fluency and ease of expressions in even longer complex stretches of speech.

Can relate own contribution skilfully to those of other speakers?

Can use circumlocution and paraphrase to cover gaps in vocabulary and structures.



ELTC SYMPOSIUM ON ASSESSMENT FOR LANGUAGE LEARNING:
ISSUES AND POSSIBILITIES.

22 – 24 APRIL 2003

Can carry out an effective fluent interview, departing spontaneously from prepared questions, following up and probing interesting replies.

Can follow the essentials of lectures, talks and reports and other forms of academic / professional presentation which are propositionally and linguistically complex.

Can develop an argument systematically with appropriate highlighting of significant points, and relevant supporting detail.

A2:

Can write simple notes to friends

Can ask and answer questions about personal details, such as where they live, people they know, and things they have

Can reply in an interview to simple direct questions spoken very slowly and clearly in direct non-idiomatic speech about personal details

Can indicate time by such phrases as next week, last Friday, in November, three o'clock

Can understand instructions addressed carefully and slowly to him/her and follow short, simple directions.

Figure 2: Sample tasks (Norris, et al., 1998)

ITEM 1

Situation: Your friend John has broken a bone in his hand. He cannot write (see photo of John). You told him that you would help him with the writing. Now, he wants you to fill out a change of address form for him. Study the form provided. Be prepared to listen for the information requested on the form. John said he would leave the information on your answering machine.

Task: Play the message from John. Listen for information from the change of address form. Fill in the form for John. You may listen to the message as many times as you need to get the correct information.

Time: You have 10 minutes to complete this task.

Product: Completed change of address form.

ITEM 2

Situation: At your workplace, many of your co-workers look healthy. Your boss wants you to write an office memo about the relationship between diet and heart disease.



ELTC SYMPOSIUM ON ASSESSMENT FOR LANGUAGE LEARNING:
ISSUES AND POSSIBILITIES.

22 – 24 APRIL 2003

Task (part 1): Read the memo from your boss. Pay close attention to the questions she wants you to answer.

Task (part2): Now listen to the information from a health information service. Find answers to your boss's questions about the relationship between diet and heart disease. Take notes on any important information you hear. You may listen more than once.

Task (part 3): After you find information about the relationship between diet and heart disease, write a brief office memo on the office stationery provided. Your memo should summarise answers to your boss's questions, based on what you now know. Write the memo to your co-workers.

Time: You have 15 minutes to complete this task.

Product: Written notes, written memo

IV. Criterion-referenced language testing

Definitions: (Brown & Hudson, 2002)

Norm-referenced test

any test that is primarily designed to disperse the performances of students in a normal distribution based on their general abilities, or proficiencies, for purposes of categorizing the students into levels or comparing students' performances to the performances of the others who formed the normative group.

Criterion –referenced test

Criterion-referenced measures indicate the content of the behavioural repertory, and the correspondence between what an individual does and the underlying continuum of achievement. Measures which assess students' achievement in terms of a certain criterion standard thus provide information as to the degree of competence attained by a particular student which is independent of reference to the performance of others. (Glaser, 1963, p.519).



Issues/Concerns raised by participants

1. Will multiple measures fit in with (CRT)?

Multiple measures will fit in with CRT especially in task-based assessment where multiple tasks may be built. However, it takes time to formulate the item specifications. It is important to remember that the measures allow for decisions to be made.

2. Is there any effort to bring in the learners' perspective other than the test users' perspective in scales development?

'Can do' statements can be formulated from the learners' point of view. Learners can perform self-assessment based on what they have done with the language.

3. What kinds of qualities are we trying to assess and whether they are functional or not?

How do we know?

You know by relating to the relevance and the context the learners are going to be in.

4. Which is better CRT or NRT to allow students' reflection? To what extent do teachers use test scores for reflection?

CRT is more helpful in describing students' performance. Whether teachers use test scores to reflect depends on the teacher's training and background and whether she has the time.

References

ACTFL. (1989). The ACTFL provisional proficiency guidelines. In T.V. Higgs (Ed.). *Teaching for proficiency, the organizing principle*. Lincolnwood, Ill.,: National Textbook Company. 219-226

Bachman, L.F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.

Bachman, L.F. & Palmer, A.S. (1996). *Language testing in practice*. Oxford: Oxford University Press.

Council of Europe (2001). *Common European framework of reference for language : Learning, teaching, assessment*. Cambridge: Cambridge University Press.



ELTC SYMPOSIUM ON ASSESSMENT FOR LANGUAGE LEARNING:
ISSUES AND POSSIBILITIES.

22 – 24 APRIL 2003

FSI (n.d.) Absolute language proficiency ratings. In M.L. Adams & J.R. Firths (Eds). *Testing kit: French and Spanish*. Washington, D.C.: Department of State. 13-17.

Messick, S. (1989) Validity. In R.L. Linn (Ed.), *Educational measurement*. 3rd.ed. New York: American Council on Educational and Macmillan, 13-103.

North, B. (2000). *The development of a common framework scale of language proficiency*. New York: Peter Lang.

Pawlidowska-Smith, G. (2000) Canadian language benchmarks. English as a second language – adults. Centre for Canadian Language Benckmarks ([www. Language.ca](http://www.Language.ca)) accessed March 2003.

Pawlidowska-Smith, G. (2000) Canadian language benchmarks 2000. Theoretical framework. Centre for Canadian Language Benckmarks ([www. Language.ca](http://www.Language.ca)) accessed March 2003.