

Can we trust teacher judgments? A validity study of a standards-based classroom assessment of English proficiency

Lorena Llosa
Department of Teaching and Learning
New York University

May 16, 2007

CAESL 2007, San Jose, CA

The Study

Llosa, L. (in press). Validating a standards-based classroom assessment of English proficiency: A multitrait-multimethod approach. *Language Testing*.

Llosa, L. (2005). Assessing English learners: A qualitative investigation of teachers' interpretations of the California ELD Standards. *The CATESOL Journal*, 17(1), 7-18.

Background

- Use of standards-based classroom assessments in a high-stakes accountability system
- Advantages
 - Assess a greater number of standards; broad range of student performance over time; formative function
- Issues
 - Variability in teachers' interpretations of language ability
 - Other factors that may threaten the validity and reliability

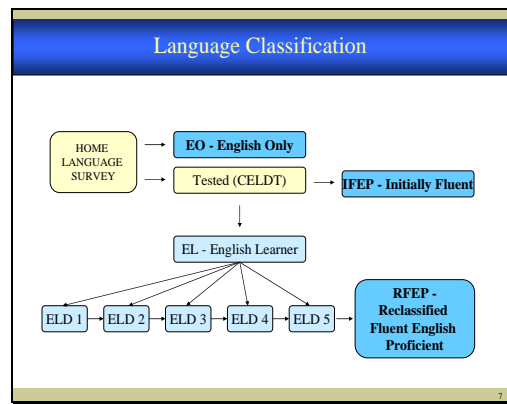
(Brindley 1998 & 2001; Rea-Dickins & Gardner, 2000)

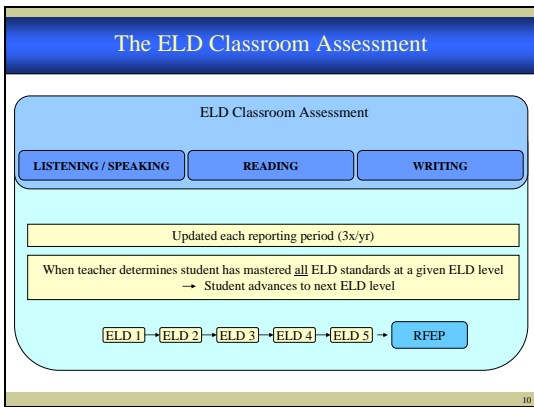
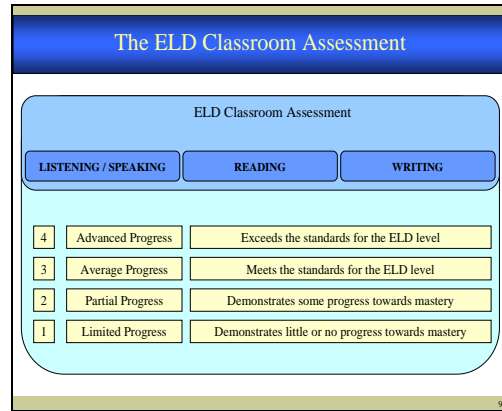
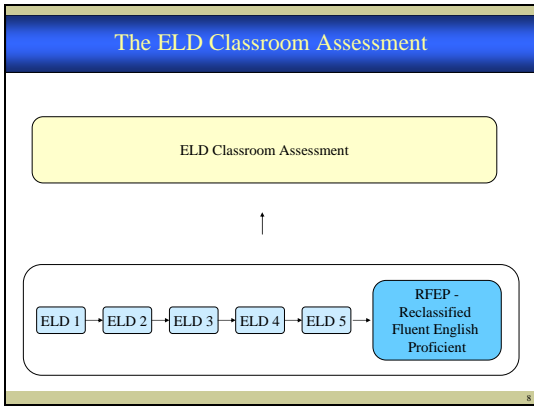
Context for the Study

- No Child Left Behind Act (2001)
- Emphasis on assessment and accountability
- English learners
 - Must demonstrate adequate yearly progress in content assessments and assessments of English proficiency

Context for the Study

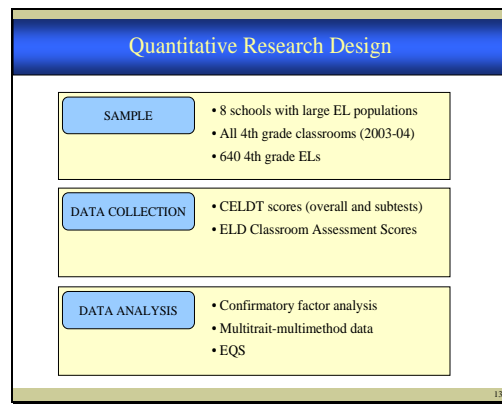
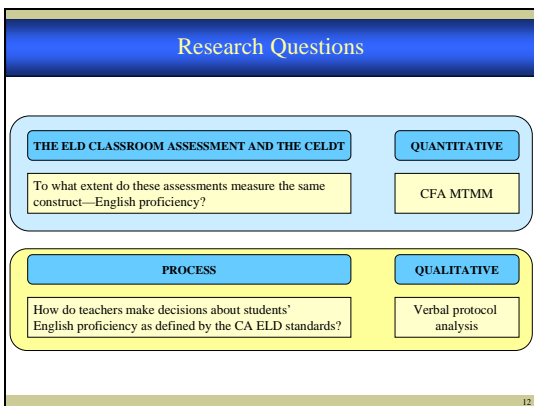
- Large urban school district in California
 - 44% English Learners
- Assessments of English proficiency
 - California English Language Development Test (CELDT)
 - ELD Classroom Assessment





Purpose of the Study

The purpose of this study was to determine the extent to which we can interpret scores on the ELD Classroom Assessment as indicators of English language ability.



Multitrait Multimethod

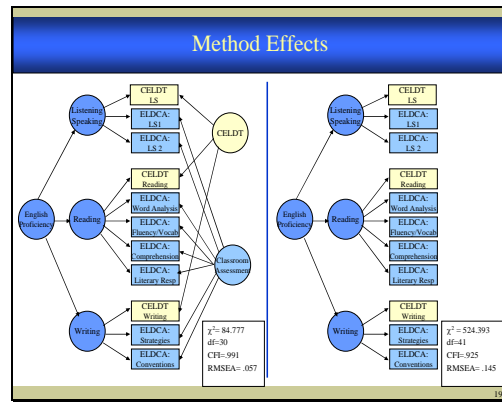
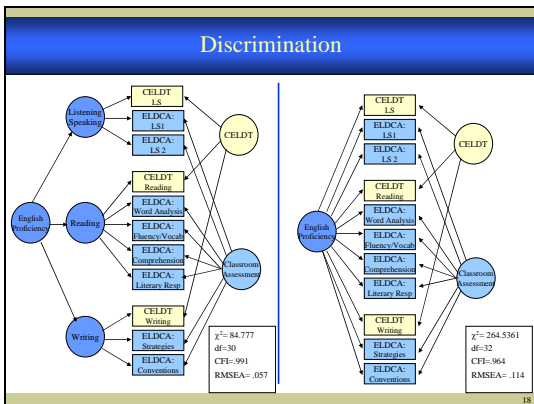
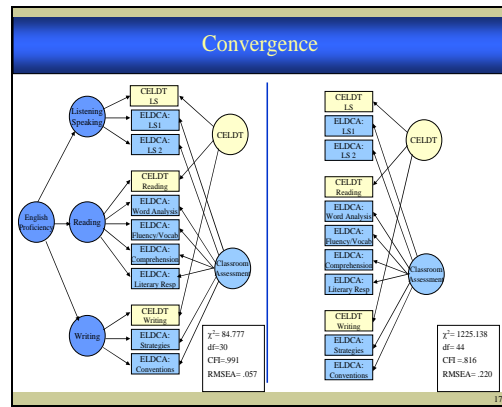
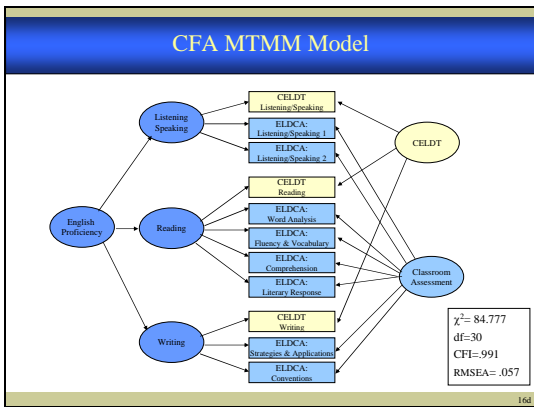
- Approach for assessing construct validity of measures (Campbell & Fiske, 1959)
- Convergence: the degree to which concepts or traits that should be related theoretically are interrelated in reality.
- Discrimination: the degree to which traits that should not be related theoretically are, in fact, not related.

14

CFA MTMM

- Widaman's (1985) approach to CFA MTMM
- Create a method factor for each measurement method and a trait factor for each measured construct.
- A set of models are compared to examine convergence and discrimination and method effects.

15

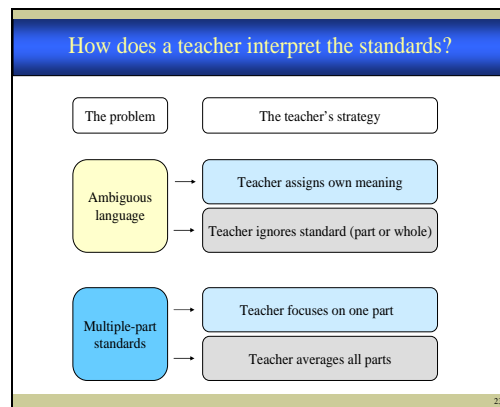
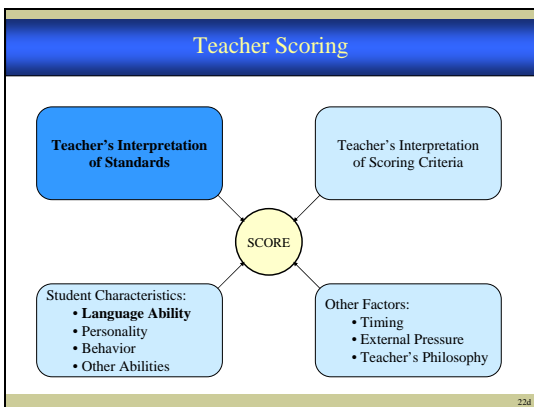


So far . . .

- The CFA MTMM analysis provides evidence that the ELD Classroom Assessment measures English proficiency as defined by the California ELD Standards.
 - Convergence: measures assess the traits
 - Discrimination: measures assess three distinct traits
- However, there is substantial method effect associated with both assessments

Research Design—Qualitative component

SAMPLE	<ul style="list-style-type: none"> 5 schools with large EL populations 10 4th grade teachers Varying backgrounds
DATA COLLECTION	<ul style="list-style-type: none"> 1-1.5 hour meetings <ul style="list-style-type: none"> Verbal protocols Semi-structured interviews Audiotaped
DATA ANALYSIS	<ul style="list-style-type: none"> Qualitative analysis of transcripts Responses grouped by standard Identified relevant themes/categories



How does a teacher interpret the standards?

Ambiguous language	<p>Apply knowledge of language to derive meaning / comprehension from literary texts.</p> <p><i>Again, we went over this again today – going over our vocabulary in class. And she was very proficient at knowing that if a '-tion' is at the end of a word, that means that it's a noun. So if I know it's a noun I can start to understand what that word means. And so we actually went over that today and she did a great job at that.</i></p>
Teacher assigns own meaning	

How does a teacher interpret the standards?

Ambiguous language	<p>Apply knowledge of language to derive meaning / comprehension from literary texts.</p> <p><i>I'm never sure what they mean by apply knowledge of language, but he can definitely find meaning and comprehension in the text so for these kinds of things I just use my best judgment. So I'd give him a 4.</i></p>
Teacher ignores part of the standard	

How does a teacher interpret the standards?

Ambiguous language

Teacher assigns own meaning

Use common English morphemes in oral and silent reading.

*These are so hard because how do I know he's applying knowledge of morphemes especially in such a big class. Unless he says: "Oh, I'm noticing that this word has an ending and that this ending is changing the meaning of the word." It's really hard to tell. So what I do, if he's **doing well in reading**, I say a 3. He is **deriving meaning** from what he's reading.*

28

How does a teacher interpret the standards?

Ambiguous language

Teacher ignores standard

Use common English morphemes in oral and silent reading.

I really don't know for sure, they say in oral and silent reading. Oh, that's a strange one. I would say a 3 because I can't be sure. I don't think anyone can be sure.

29

How does a teacher interpret the standards?

Multiple-part standards

Teacher focuses on one part

Independently create cohesive paragraphs that develop a central idea with consistent use of standard English grammatical form. (Some rules may not be in evidence.)

A 2. He's not really independent. If I sit and I work with him, he does a lot. He can write a paragraph, but I have to really keep him on task and keep him focused. So independently he can't do it. It would be a 2.

30

How does a teacher interpret the standards?

Multiple-part standards

Teacher focuses on one part

Independently create cohesive paragraphs that develop a central idea with consistent use of standard English grammatical form. (Some rules may not be in evidence.)

[W]hat I read from this is that they're mostly concerned that he can write a paragraph using standard English because writing a paragraph, a cohesive paragraph that develops the central idea, is difficult for any 4th grader. So I would give him a 3 because I think they're mostly focusing on the fact that he can use standard English.

32

How does a teacher interpret the standards?

Multiple-part standards

Teacher averages all parts

Actively participate in social conversations with peers and adults on familiar topics by asking and answering questions and soliciting information.

I would give him a 3. He's usually a 2. Well, he would be a 4 with his peers. He has no problem with his peers. With adults, it would probably be a 2. But when it's a familiar topic he does great. Because ... [H]e had attended space camp, and he wants to be an astronaut. So when we had the story about that, he was sharing every day. I would've given him a 4 that week. But overall, I would say a 3.

33

Qualitative Findings

- Teachers do not interpret standards consistently
- ELD Classroom Assessment scores can mean different things for different students.

34

Findings

- The CFA MTMM analysis indicates that the ELD Classroom Assessment measures English proficiency as defined by the California ELD Standards
- The method effect and the qualitative analysis, however, reveal that teachers do not interpret the ELD standards or the scoring criteria consistently when scoring the ELD Classroom Assessment.

How can these seemingly contradictory findings be reconciled?

35

Explanations for contradictory findings

- Teacher scoring in practice vs. teacher scoring for the purpose of the study
- Scores on individual standards are not used to make decisions

37

Why are teachers not interpreting the standards consistently?

- Nature of the standards
- Lack of professional development
- Lack of alignment between ELD Classroom Assessment and instruction

38

Implications

- For standards-based education
 - Standards are at the core of the system.
 - Unless there is a common understanding of the standards, the system cannot work as intended.

39

So, can we trust teacher judgments?

- Yes and No
- Effective use of classroom assessments
 - Professional development
 - Alignment between assessment and curriculum

"It is possible for teachers operating in a systematic way over time to develop highly accurate assessments of student learning" when standardized assessments are consistent with classroom assessments, are guided by the same standards, and are evaluated in terms of shared criteria for judging high quality work (Shepard, 2000, p. 36).

40