

When ELL Students Take Both Title I and Title III Assessments

Scott Marion
Center for Assessment
Reidy Interactive Lecture Series
September 27-28, 2007



Acknowledgement

- Thanks to Brian Gong and Rich Hill for helpful advice and direct contributions to this presentation. Of course, any remaining shortcomings are my own.



Some questions

- Can we learn anything about the validity of either assessment when students take both ELP and content assessments?
 - If so, what do we think we can learn?
 - How might we go about trying to learn these things?
- How does the relationship between the two tests/constructs interact with accountability?
- An important caveat—I'm not an ELL/ELP expert...



We Can Always Learn

- Pre-academic language proficiency
 - We could gather convergent and discriminant validity evidence to evaluate the degree of overlap between the two constructs measured
 - Useful source of evidence, but less interesting (at least to me) when the intended construct (social language vs. academic content knowledge) is clearly different for each of the tests



The Assessment Triangle (Again!)

- **What** is the intended construct and how do we measure it?
 - Proficiency
 - Progress
- **Who** are the students and **how do they learn?**
 - Heterogeneity
- How do we **interpret** the two sets of test results?

- Focus on #1 and a bit on #3.



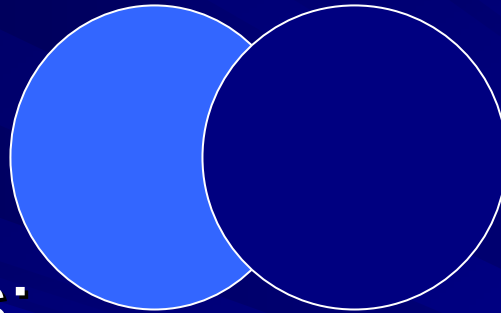
What is the intended construct?

- ELP—The English language knowledge and skills students need to access and achieve in the academic disciplines
- Title I—The knowledge and skills (i.e. content standards) determined to be important for students to know and be able to do at a given grade level
- Are these two constructs the same?
 - I'll say *no* (remember my caveat), but that's not the end of the story.
- Further, what would be the academic equivalents of the ELP construct of speaking and listening? Some states include speaking and listening in ELA standards, but what about the other content areas?

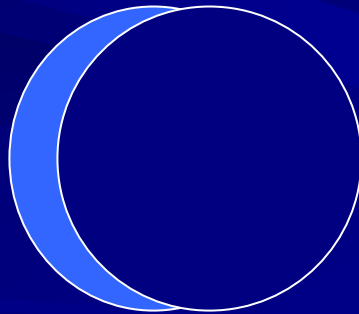


Do the constructs overlap?

■ Is it like this:



■ Or like this:

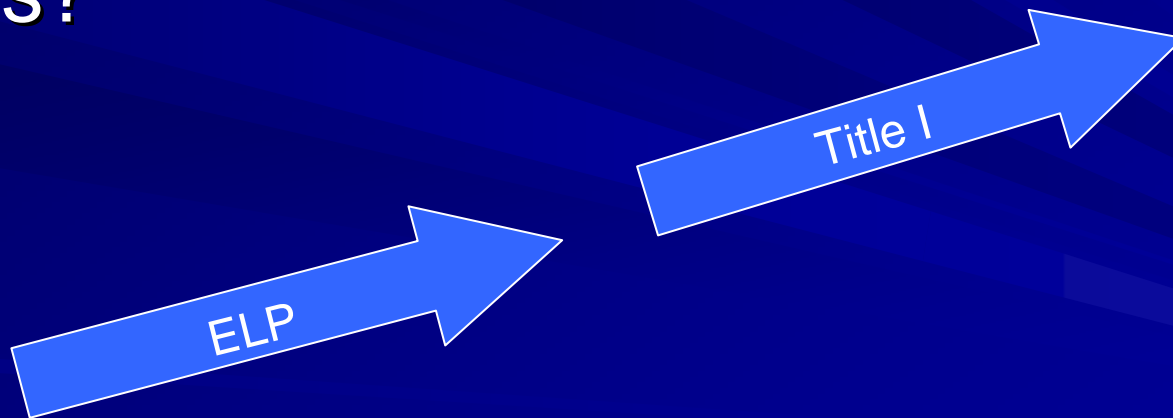


■ Or is this even the correct representation?



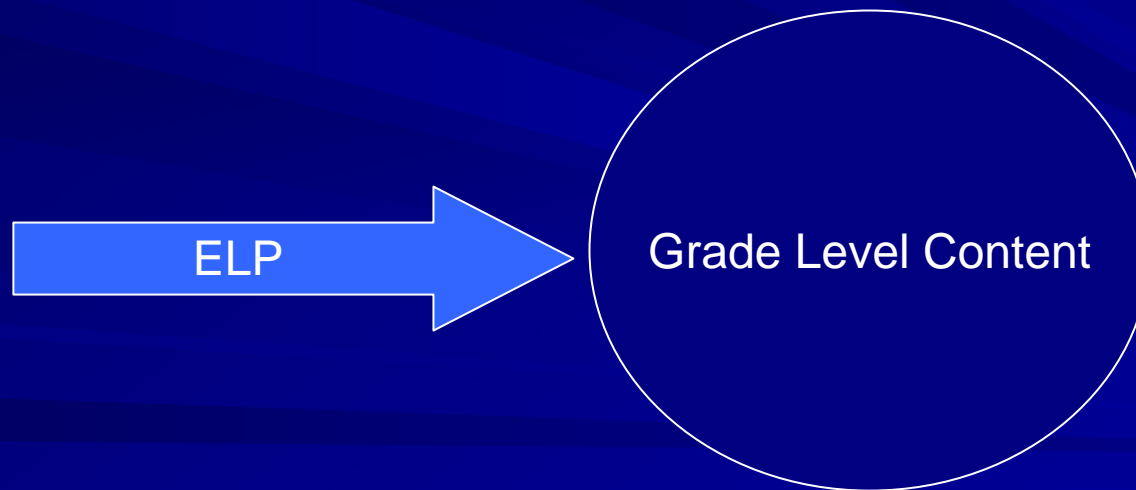
Are the constructs sequential or continuous?

- The previous slide implies that the two domains are related and at the same “level” (for lack of a better term), but might the relationship be more sequential like this?



Do ELP skills simply serve as access to grade level content?

- Instead of sequential, what if the ELP content simply provides an entry point into grade level content, but the relationship doesn't extend beyond that?



Progression and Proficiency

- Grade levels serve as a structure for both progression and proficiency on the Title I assessments.
- Proficiency is often related to at least grade spans on the ELP assessments, but it is not clear that progression is as closely tied to grade level as is the case with the Title I tests



Progression and Proficiency

- Once we have a sense of how the constructs interact, we need to figure out how the defined proficiency levels on the ELP assessment interact with performance on the Title I assessments.
- For example, does “proficient” on the ELP assessment mean anything in relationship to the student’s likelihood of scoring proficient (or any other level) on the Title I grade level assessment?
 - What should be the nature of this link?
 - Should this link be similar across reading, math, and science?



Two Different Questions

- Let's keep in mind that the Title I and Title III tests are designed to answer two quite different questions:
 - Title I: How much does the student know relative to the content standards (knowledge and skills)?
 - Title III: How much “school or academic” language does the student possess to enable them to access and achieve in the discipline?
 - Another way of thinking about the Title III question is: How much are the student's language skills an obstacle to demonstrating achievement in the content areas?



So What?

- While the particular relationships and representations of the two constructs are important, I am concerned about how we go about determining the nature of this relationship
 - In other words, how are these constructs and relationships between tests being validated?



Evaluation Considerations

- Convene ELL and content area experts as part of a judgmental process to articulate the relationship between the two constructs
 - Which picture fits best or is an entirely new picture needed?
 - Do these relationships—within content area—change across grade levels?
 - These judgments should help contextualize the results of other analyses



Evaluation Considerations

- Variety of approaches that can be used to help evaluate the extent of the relationship between the two tests
 - “Think alouds” to help get at the nature of the response process and more importantly the differences in response process across the two tests
 - Some pretty conventional convergent/discriminant approaches (e.g., multi-trait/multi-method)
 - Rich Hill’s correlational evaluation
 - Both of these approaches help describe the relationship among the observed responses, but do not help us figure out what the relationship should be
 - Confirmatory factor analysis can help evaluate the internal structure of each test and if students complete both assessments, we could evaluate the overlap in the hypothesized structures
 - Predictive analyses allow us to judge the score on one assessment associated with a score on another assessment
 - What should we expect for the predictive relationship? It should be considerably greater than zero, but it shouldn’t be 1.0 either.



Evaluation Considerations

- We have to think about the converse too—
 - While our construct analyses might lead us to conclude that there is far from a perfect relationship between the two constructs...but
 - It is tough to explain how ELL students can score proficient on the grade level ELA tests, yet not score proficient on the ELP test...failing the “smell test”



Some Practical Advice

- Some assumptions (perhaps oversimplified):
 - If a student gets an item correct, we assume:
 - s/he knows the content AND
 - s/he does not have a language obstacle
 - If a student gets a test item incorrect, we could conclude:
 - s/he does not know the content OR
 - s/he does not have the language skills to allow her/him to demonstrate the content knowledge
- Of course there could be confounding factors, but let's just think of the general case for now...



Practical Advice

- More assumptions (relative to content area tests):
 - English language competence is relatively constant at a given point in time
 - Academic knowledge in domains varies across subskills and items
 - The English language demand varies across test items
 - For the initial analyses, we argue that the language demand is fixed with the item and not variable across persons. We recognize the possibility of interaction, but similar to how we approach cognitive demand in alignment studies, we assume the depth of knowledge resides with the item



Practical approaches

- Adhering to the assumptions outlined on the previous slides, we envision several lines of research to address some key questions.
- The multi-trait/multi-method types of analyses referenced previously should be structured to help us disentangle the:
 - language proficiency of the student,
 - the content knowledge of the student, and
 - the language demands associated with items and subdomains
- We need the data from the concurrent administrations of both the Title I and Title III tests to address these issues.



Validity Evaluation Plan

- So when double testing occurs, it makes sense to make use of these data in validity evaluations for both assessments, particularly for the ELP assessment
- Again, specifying the hypothesized relationship among the various assessments must be articulated in the validity studies plan
- The results of the empirical analyses must be integrated with the judgmental results into a coherent validity argument



How does this relate to accountability?

- For the n^{th} time, validity always comes down to purposes and uses
- More importantly, assessment validity is a necessary, but not sufficient requirement for accountability validity
- The burden on the ELP assessments are more onerous than for grade level assessments—schools are accountable for both status and progress
 - Most states want to report and some even want to hold schools accountable for subdomains (e.g., reading, speaking)



Accountability

- How and where the various AMAOs are set has implications for the validity of the accountability system
- Is there anything we can learn from the double test administrations about the validity of the accountability AMAOs?
 - School level relationships among the success of meeting AMAOs and performance and improvement on the content area assessments



Accountability

- Another consequential burden faced by ELP tests that we don't face with grade level tests...
 - Kids who score too high get exited out of the program before they might be ready
 - This is certainly a good validity check, but it is one additional burden on validating ELP assessment and accountability systems



For more information

■ Contact smarion@nciea.org

