Considering Instructional Sensitivity in the Validity Evaluation Process

Ellen Forte, edCount, LLC
Marianne Perie, University of Kansas

Instructional Sensitivity Conference
The Achievement and Assessment Institute
The University of Kansas
November 13-15, 2013
Claim 1. The content standards have been designed to facilitate learning in a subject matter by building on prior knowledge through known subject and cognitive development.

Claim 2. The assessment has been designed to yield scores that reflect students’ knowledge and skills in relation to academic standards.

Claim 3. The assessment is administered and scored as intended.

Claim 4. The assessment performance levels accurately reflect the level of knowledge and skills each student has.

Claim 5. Assessment scores and performance levels reflect meaningful differences in students’ proficiency.

Claim 6. Change in individual assessment scores from one year to the next reflects the level of learning during the school year.

Claim 7. Assessment scores can be used to evaluate student proficiency in the subject matter.

Claim 8. Changes in assessment scores can be used to evaluate teacher effectiveness.

Claim 9. Distribution across proficiency levels and changes in that distribution can be used to evaluate school and district quality.

Students leave the grade/school having gained sufficient knowledge and skills required for the next phase of education or training.

Teachers are recognized for the effectiveness and given feedback to improve.

Resources are appropriately allocated to schools and districts to maintain or increase their quality.
Breaking Claim 2 into Required Assumptions

Claim 2. The assessment has been designed to yield scores that reflect students’ knowledge and skills in relation to academic standards.

Assumption 1. The assessment blueprint reflects the intended domains of knowledge to be assessed.

Assumption 2. The item specifications capture the intended knowledge and skills.

Assumption 3. The items are written to allow students to provide evidence of their knowledge and skills.

Assumption 4. The type of task is appropriate to the knowledge/skill it is intended to measure.

Assumption 5. Rubrics appropriately define the domains of performance and distinguish among levels of performance.

= Alignment
Tricky Questions in Evaluating Instructional Sensitivity (IS)

• Is instructional sensitivity a characteristic of an item? A set of items?

• Does instructional sensitivity vary across students or contexts?

• Are instructionally sensitive items and items that guide instruction the same?

• How do we know what has been taught? Is it the same as what has been learned?
Claim 1. The content standards have been designed to facilitate learning in a subject matter by building on prior knowledge through known subject and cognitive development.

Claim 2. The assessment has been designed to yield scores that reflect students’ knowledge and skills in relation to academic standards.

Claim 3. The assessment is administered and scored as intended.

Claim 4. The assessment performance levels reflect meaningful differences in students’ proficiency.

Claim 5. Assessment scores and performance levels accurately reflect the level of knowledge and skills each student has.

Claim 6. Change in individual assessment scores from one year to the next reflects the level of learning during the school year.

Claim 7. Assessment scores can be used to evaluate student proficiency in the subject matter.

Claim 8. Changes in assessment scores can be used to evaluate teacher effectiveness.

Claim 9. Distribution across proficiency levels and changes in that distribution can be used to evaluate school and district quality.

Students leave the grade/school having gained sufficient knowledge and skills required for the next phase of education or training.

Teachers are recognized for the effectiveness and given feedback to improve.

Resources are appropriately allocated to schools and districts to maintain or increase their quality.
Claim 1. The content standards have been designed to facilitate learning in a subject matter by building on prior knowledge through known subject and cognitive development.

Claim 2. The assessment has been designed to yield scores that reflect students’ knowledge and skills in relation to academic standards.

Claim 3. The assessment is administered and scored as intended.

Claim 4. The assessment performance levels accurately reflect meaningful differences in students’ proficiency.

Claim 5. Assessment scores and performance levels accurately reflect the level of knowledge and skills each student has.

Claim 6. Change in individual assessment scores from one year to the next reflects the level of learning during the school year.

Claim 7. Assessment scores can be used to evaluate student proficiency in the subject matter.

Claim 8. Changes in assessment scores can be used to evaluate teacher effectiveness.

Claim 9. Distribution across proficiency levels and changes in that distribution can be used to evaluate school and district quality.

Big IS Idea #1

Big IS Idea #2

Precursors & Context

Assessment System

Assessment Outcome

Score Uses

Goals

Generic Interpretive Argument

Teachers are recognized for the effectiveness and given feedback to improve.

Students leave the grade/school having gained sufficient knowledge and skills required for the next phase of education or training.

Resources are appropriately allocated to schools and districts to maintain or increase their quality.
Claim 2. The assessment has been designed to yield scores that reflect students’ knowledge and skills in relation to academic standards.

Claim 1. The content standards have been designed to facilitate learning in a subject matter by building on prior knowledge through known subject and cognitive development.

Claim 5. Assessment scores and performance levels accurately reflect the level of knowledge and skills each student has.

Claim 6. Change in individual assessment scores from one year to the next reflects the level of learning during the school year.

Big IS Idea #1

Test scores are meant to generalize to a target domain.

If testing information is meant to reflect the knowledge and skills students have acquired in instructional settings, then one must be able to distinguish between knowledge and skills that were acquired in association with exposure to instruction from those that were pre-existing or acquired via other experiences or maturation.
Big IS Idea #2

Test scores are used to evaluate (students), teachers, schools, and programs.

If testing information is to be used to judge teacher, school, or program effectiveness, then one must establish evidence that that information reflects teacher, school, or program effectiveness and not mostly other variables.
<table>
<thead>
<tr>
<th>When tested</th>
<th>Target not taught</th>
<th>Target taught</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior to instructional segment</td>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td>Following instructional segment</td>
<td>B</td>
<td>D</td>
</tr>
</tbody>
</table>

If an item/item set is instructionally sensitive then D-C > B-A and hopefully D-B > C-A and D-B ≈ D-A

What now?
Key Components of IS Evaluation

• Theory of domain knowledge and skill acquisition
• Micro-genetic studies
• Evidence-centered design principles
• Cognitive labs
• Instruction capture and quasi-experiments
Claim 1. The content standards have been designed to facilitate learning in a subject matter by building on prior knowledge through known subject and cognitive development.

Claim 2. The assessment has been designed to yield scores that reflect students’ knowledge and skills in relation to academic standards.

Claim 3. The assessment is administered and scored as intended.

Claim 4. The assessment of groups and individuals accurately reflects the level of knowledge and skills each student has.

Claim 5. Assessment scores and performance levels accurately reflect the level of knowledge and skills each student has.

Claim 6. Change in individual assessment scores from one year to the next reflects the level of learning during the school year.

Claim 8. Changes in assessment scores can be used to evaluate teacher effectiveness.

Claim 9. Distribution across proficiency levels and changes in that distribution can be used to evaluate school and district quality.

Students leave the grade/school having gained sufficient knowledge and skills required for the next phase of education or training.

Teachers are recognized for the effectiveness and given feedback to improve.

Resources are appropriately allocated to schools and districts to maintain or increase their quality.

Evidence-centered design principles

Cognitive labs

Micro-genetic studies

Instruction capture and quasi-experiments

Theory

Precursors & Context

Assessment System

Assessment Outcome

Score Uses

Goals
Breaking Claim 2 into Required Assumptions

Claim 2. The assessment has been designed to yield scores that reflect students’ knowledge and skills in relation to academic standards

Assumption 1. The assessment blueprint reflects the intended domains of knowledge to be assessed.

Assumption 2. The item specifications capture the intended knowledge and skills.

Assumption 3. The items are written to allow students to provide evidence of their knowledge and skills.

Assumption 4. The type of task is appropriate to the knowledge/skill it is intended to measure.

Assumption 5. Rubrics appropriately define the domains of performance and distinguish among levels of performance.

= Alignment