A clinical-holistic approach to educational assessment for learning:
Using multiple data sources and multiple methods.

Joanna S. Gorin
Using Assessment Data for Instruction and Learning: A Complex Decision Making Process
Three Claims...none of which are new.

1. Human decision making is highly susceptible to biases and errors

2. Educational assessment is analogous to clinical diagnosis – both are complex decision-making processes.

3. Formal decision making structures incorporating multiple measures will improve the validity and utility of our decisions.
Educational Assessment Use-Argument: Evidence-Centered Design

**Claim about a test taker**

**Test Score**

**Targeted Knowledge, Skill Ability**

**GROUNDS**
Student responded correctly to a multiple choice item that requires knowledge of core idea A.

**DATA**
Given that... the student bubbled in the keyed response.

**WARRANT**
Since... a student who has mastered core idea A is more likely to select the correct response.

**QUALIFICATION for Mastered Inference**
Unless... the student has seen the item before. the problem can be solved without knowing the content. the student guessed.

**QUALIFICATION for Not Mastered Inference**
Unless... the problem requires prior knowledge the student does not have.

**BACKING**
On account of... conditional probabilities, conditioned on A, are independent. content alignment study shows item requires content Y.. item statistics show good discrimination.
Evidence-Based Decision Making in Medicine

Step 1: Identify the problem or opportunity.
Step 2: Gather internal evidence or data about the problem, and evaluate its relevance and validity.
Step 3: Gather external evidence about the problem from published research.
Step 4: Gather views from stakeholders affected by decision and consider ethical implications.
Step 5: Integrate and critically appraise all data and then make a decision.

So what goes wrong?

• There is too much evidence
• There is no enough good evidence.
• The evidence doesn’t quite apply.
• People are trying to mislead you.

You are trying to mislead you.
• The side effects outweigh the cure.
• Stories are more persuasive anyway.
How Do We Make Decisions?

**System 1**

- Rapid pattern recognition and cognitive crystallization;
- Quick intuitive assembly of interpretations
  - Fast
  - Efficient
  - Effortless
  - Automatic
  - Unconscious
  - Everyday decisions
  - Error prone

**System 2**

- Piecing together outputs of fast cognition and provoking further fast-cognition operations
  - Slow
  - Conscious
  - Effortful
  - Complex Decisions
  - Reliable
  - Deliberate reasoning
Cognitive Bias in Clinical Practice

- Availability Heuristic
- Representativeness Heuristic
- Anchoring Bias
- Confirmation Bias
- Overconfidence Bias
- Search Satisficing
- Framing Bias
- Hindsight Bias
- Escalation Bias

All of these Biases are Associated with System 1... and apply equally to educational assessment.
Supporting System 2 Thinking with Decision Trees

• Collect information from different sources to use.
• Present a range of possible choices.
• Use intuitive design and reduce distractions.
• Use structures and supports to document decision paths.
• Evaluate the merits/plausibility of each decision path.
## Decision Making in Clinical Assessment

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Symptom 1</th>
<th>Symptom 2</th>
<th>Symptom 3</th>
<th>Symptom 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosis X</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Diagnosis Y</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Diagnosis Z</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Client</th>
<th>Symptom 1</th>
<th>Symptom 2</th>
<th>Symptom 3</th>
<th>Symptom 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Client 1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Client 2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Client 3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Client 4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

But that is just the first set of measurements...

- Physical & Visual Examination
- Analysis of Bodily Fluids
- Diagnostic Imaging
- Genetic Testing
- Cellular & Chemical Analysis
- Measurement of Bodily Functions
- Biopsy
- Response to Treatment
Multiple Measures for Psychosocial Assessment
Decision Flow for Psychoeducational Assessment
Bayes Net for Diagnosis of Dementia, Alzheimer’s Disease and Mild Cognitive Impairment

Seixas, et al., 2014
Multiple Measures in Educational Assessment

• “Persons who interpret, use, and communicate assessment results have a professional responsibility to use multiple sources and types of relevant information about persons or programs whenever possible in making educational decisions.”
  

• What to consider:
  
  • Construct validity
  
  • Decision validity
  
  • Purpose of the assessment
More Data vs. Multiple Measures

• Source of Data
• Method of Data Collection
• Timing of Data Collection
• Level of Information
  • Individuals
  • Assessment Tasks
  • Assessment Context
  • Sociocultural Context
• Type of Data
  • Evidence of Product
  • Evidence of Process

• “[P]eople who are considering buying a house look at the house's age, condition, location, style, features, and construction, as well as the price of nearby homes. Doctors diagnosing an illness use multiple assessments: the patient's medical history, lab tests, answers to questions about how the patient feels, and so on. The question is, Why do education policymakers and practitioners sometimes opt to make important decisions based on only one indicator?”
  
  Brookhart, 2009
Multiple Measures in terms of a Sociocognitive Model of Assessment Argument

- Goals, self-efficacy, task utility-value
- Opportunity-to-learn
- Cultural norms
- Linguistic background
- Values and attitudes
- Health
- Emotional state
- Prior knowledge and experience

Figure adapted from Mislevy, 2017
Multiple Measures in terms of a Sociocognitive Model of Assessment Argument

- Claim about student
  - Warrant
  - Data concerning student performance
    - Data concerning task situation
    - Social/cultural context in which the assessment argument is framed – what is important, what is expected, what is assumed...
    - Other information about student vis a vis assessment situation

Student acting in assessment situation

Linguistic, cultural and substantive milieu that a student has accumulated through experience which can be brought to bear on the assessment activities.

Multiple Measures:
- Linguistic background (e.g., native language, daily language usage/interaction)
- Cultural context (e.g., formality, hierarchy, expectations, gender roles)
- Classroom factors (e.g., teacher quality, curricular decisions, physical environment)

Figure adapted from Mislevy, 2017
The Importance of Assessment Purpose When Using Multiple Measures

<table>
<thead>
<tr>
<th>Three Ways to Define Multiple Measures</th>
<th>Conjunctive</th>
<th>Compensatory</th>
<th>Complementary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measures of different constructs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Different measures of the same construct</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiple opportunities to pass the same test</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

To use multiple measures appropriately, start by understanding their purposes.

Brookhart, 2009
## Assessment Purposes as Claims

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Tests results can be used to:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Licensure or Certification</td>
<td>Verify whether someone has the necessary knowledge and skills to be a qualified practitioner or to be given advanced standing in an occupation or profession</td>
</tr>
<tr>
<td>Admissions</td>
<td>Inform decisions about which people should be selected for entrance to an educational institution</td>
</tr>
<tr>
<td>Placement</td>
<td>Determine which courses or level of a course a student should take</td>
</tr>
<tr>
<td>Employment</td>
<td>Inform decisions on the hiring, placement and promotion of potential and current employees</td>
</tr>
<tr>
<td>Curriculum-based End-of-Course Testing</td>
<td>Determine whether students have mastered the objectives of the course taken</td>
</tr>
<tr>
<td>Exit Testing</td>
<td>Find out whether students have learned the amount necessary to graduate from a level of education</td>
</tr>
<tr>
<td>Policy Tools</td>
<td>Provide data to policymakers that helps them make decisions regarding funding, class size, curriculum adjustments, teacher development and more</td>
</tr>
<tr>
<td>Course Credit</td>
<td>Indicate whether a student should receive credit for a course he or she didn’t take through demonstration of course content knowledge</td>
</tr>
<tr>
<td>Accountability</td>
<td>Hold various levels of the education system responsible for test results that indicate if students learn what they should have learned</td>
</tr>
</tbody>
</table>
Game-based Learning & Assessment:
Off-Task Behaviors as Self-Regulation

Dynamic In-Game Progress Variables

- Overall Learning
- Overall Progress
- Off-Task Behavior
- Valence
- Emotion
- Performance Focus
- Learning Focus

Static Survey-Based Personality Variables

Adapted from Fig. 7 Sabourin, et al, (2013)
Student Math Performance: What Could We Measure in Advance To Help?

Adapted from Figure 5 in Smail, 2017
Challenges and Opportunities

- Increasing amount and types of data to be collected.
- Limited knowledge of how to interpret data.
- Increasing computational sophistication and capability.
- Limited insight into algorithms.
- Increasing knowledge about human cognition.
- Increasing complexity in problems/score uses of interest.
- Increasing access to learners in-situo.
- Challenges for privacy.
What does this mean for assessment research and practice?

• Provide tools and structures to help use System 1 and System 2 thinking when they are most appropriate.
• Expand our definition of evidence to include a broader range of data types, data sources and data structures.
• Develop methods and models to support score-use and interpretation based on multiple measures.
• Understand the consequences for score users based on decisions about:
  • Constructs valued
  • Data sources and types
  • Weights and decision rules
  • Interpretations and decisions