Approaches to Developing Assessments of 21st Century Skills

Deirdre J. Knapp, Ph.D.
Human Resources Research Organization (HumRRO)

Presented to:
Workshop on Assessment of 21st Century Skills
January 12-13, 2011

HumRRO
Human Resources Research Organization

Presentation Overview

- Identifying what to assess
- Assessment strategies
- Selecting assessment strategies
- Evaluating assessments
- Developing assessments
- Take-away points

Identifying What to Assess

- Depends on need…
  - not just what you have (or could easily get) a tool to measure
  - not on what someone simply thinks up without evidence to support the choice(s)
- Strategies for defining needs depends on context
  - In employment settings, usually rely on a job analysis study

Assessment Strategies: A Laundry List

- Fact-based multiple-choice tests
- Situational judgment tests
- Self-report multiple-choice instruments
- Forced-choice and ranking exercises
- Essay tests
- Oral exams and interviews
- Live simulations (e.g., role-plays)
- Computer-based simulations
- Actual performance through observations or portfolios

Language to Help Think about Options

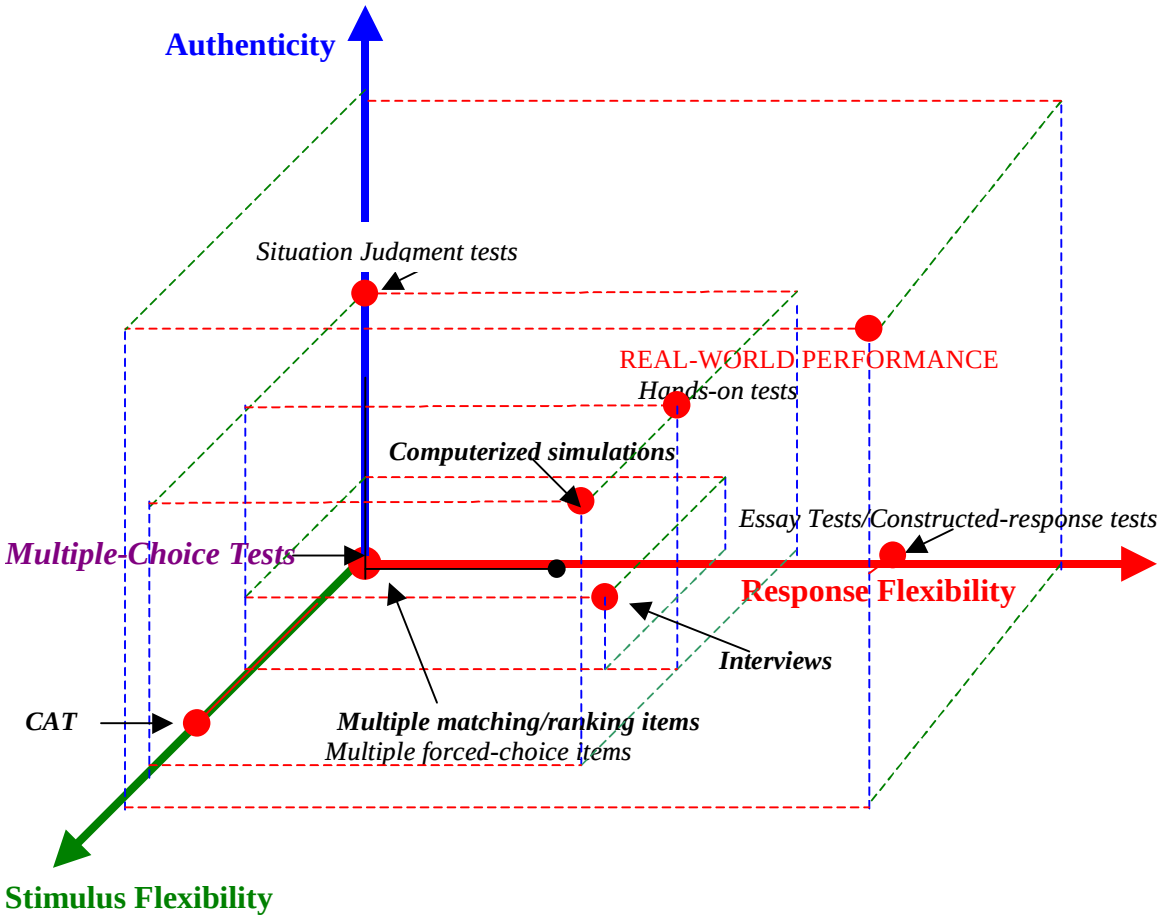
For example…

- Selected vs. constructed response
- Dichomotous vs. polytomous items
- Items vs. testlets vs. problems vs. scoring opportunities vs. events
- Evidence Centered Design Theory vs. content validation or construct validation

A 3-D Model of Measurement Strategies

- Helpful to bring structure to the dizzying array of choices and lack of common language for describing alternative measurement strategies
- Three primary dimensions
  - Stimulus flexibility: To what extent do the problems vary based on the examinee's responses to the assessment
  - Response flexibility: To what extent are response options structured?
  - Authenticity: To what extent does the assessment simulate real-world experience?

Visual Model



Selecting Assessment Strategies

- What are you trying to measure, and how well do you need to measure it?
- Anticipated reliability and validity
- Administrative constraints/capabilities
- Costs
- Test security requirements

Context makes all the difference

What You Are Trying to Measure

- Some of the possibilities
  - Abilities or aptitude
  - Procedural knowledge
  - Skill
  - Motivation
  - Other individual differences (e.g., interests, personality)
- Incorporate relevant questions into requirements analysis and discussions with stakeholders regarding testing goals
- Consider measurement contamination (e.g., over-reliance on reading ability)
- Consider what you may not be measuring that may be of interest (motivation to perform over time)

Administrative Considerations

- Ease of standardizing test administration and scoring
- Examinee volume
- Testing frequency
- Personnel/expertise available to support program
- Test delivery options
  - Paper and pencil
  - Computer-based
  - Live exercises
  - Individual versus group test administration

Costs

- Test design and development
  - Including requirements analysis
- Test delivery/administration
- Scoring and score reporting
- Maintenance costs
- Costs may also include validation studies and other activities such as program design and evaluation

Other Considerations

- Consider stakes of the testing outcome – the higher the stakes, the greater the likelihood of cheating or misrepresentation by examinees
- Even low stakes tests may be subject to performance distortion (e.g., low motivated examinees not doing their best)

Evaluating Assessments

- Validity
  - Extent to which assessment scores measure what they are purported to measure
  - Extent to which decisions based on test scores are accurate
  - Evaluate based on evidence, both rational and empirical (e.g., test design and development processes, comparison of outcomes to theoretically related variables)
- Reliability
  - Extent to which assessment yields consistent measurements and associated decisions
  - Evaluate empirically using test data (e.g., test-retest scores, examining consistency across raters)

*Assessment Development**

- Use some type of requirements analysis to identify what should be measured; be clear on the purpose of the assessment process
- Operational definition of the construct(s) to be measured
  - Observable/measurable articulation of the construct
  - Particularly challenging for the types of constructs identified as "21st century skills" which tend to be ill-defined and overlap with other more clearly understood constructs (e.g., general cognitive ability, spatial skills, job knowledge)
  - For example, critical thinking, cultural sensitivity, self-management
- Develop associated test specifications
  - For example, content specifications, test length, test item types, dimensions to be rated, etc.

**If using off-the-shelf assessments, evaluate them with regard to how the developer handled and has documented all of these steps.*

Assessment Development (Cont)

- Develop items/exercises/problems and associated scoring protocols
- Try-out on the target population
- Evaluate and refine
- Develop examinee feedback system
- If applicable, establish passing performance standard
  - This is hard to do well, so avoid if not required for purposes of the assessment program
- Conduct post-implementation evaluation

Availability of Psychometric Tools

- Computer-based technology seems to have gotten way ahead of the capabilities of available psychometric tools – it's a bit of a dirty laundry thing
- Classical test theory doesn't hold all the answers
  - Item diagnosis (e.g., *p*-values, item-total correlations)
  - Reliability estimation
- Item Response Theory and other tools are not cure-alls either

Bottom line is that some assessments can be evaluated more accurately than others. It's helpful to have an unbiased testing expert in your corner

Take-Away Points

- Select solutions that address your problems and situation
- Assessment drives behavior
  - If you assess the wrong thing or the right thing in the wrong way, the consequences will be undesirable
- Good intentions ≠ good assessment
- Make use of all available resources – you do not have to be the assessment expert but you should be an informed consumer