



Innovative Item Types: How to Evaluate the Bang for the Buck

Brian Bontempo, Ph.D.
Microsoft Corporation

Continuum of Innovation



Complexity of Innovation

- ◆ Non-Text Stimuli
- ◆ Polytomous Scoring
- ◆ Multiple Selected Response
- ◆ Simple Constructed Response
- ◆ Case Based Items
- ◆ Interactivity
- ◆ Complex Constructed Response
- ◆ Automate Judgment Scoring
- ◆ Interactivity



Shoot the moon



If you fail to shoot the moon in the game of Hearts, the consequences are monumental.



Goal



- ◆ Provide a framework whereby an organization can *evaluate* a testing program and make sound decisions about the development and implementation of Innovative Item Types.



Synthesis



- ◆ Innovative Item Types
- ◆ Evaluation
- ◆ Cognitive and Performance Assessment

Process of Evaluation

- ◆ Involve Stakeholders
- ◆ Establish the goal in using IIT
- ◆ Ascertain IIT mobility
- ◆ Identify possible directions
- ◆ Gauge the impact of the directions on the Stakeholders
- ◆ Collect & analyze some data
- ◆ Share your results

Involve Stakeholders

- ◆ Item Developers – Minimize Development Effort
- ◆ Managers – Minimize Development Cost
- ◆ Psychometricians – Maximize Test Predictive Capabilities
- ◆ Test Delivery Vendor – Minimize Test Publishing Efforts
- ◆ Test-Takers – Minimize Test Administration Effort
- ◆ Test-Users – Maximize Test Utility

Why Implement IIT?

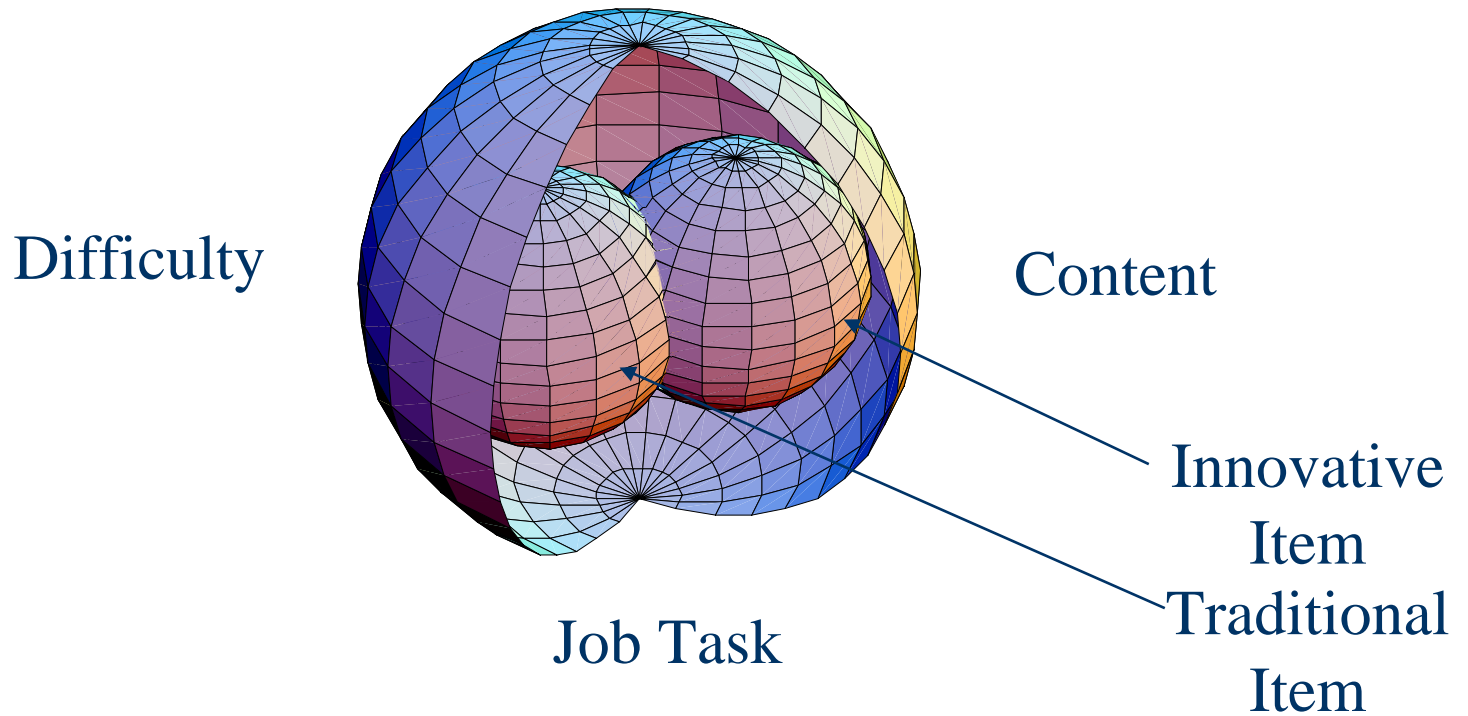
- ◆ Improve the quantity of information gathered
- ◆ Improve the quality of information gathered
- ◆ Improve the efficiency of gathering information
- ◆ Improve the efficiency of processing information
- ◆ Improve the fidelity of the information



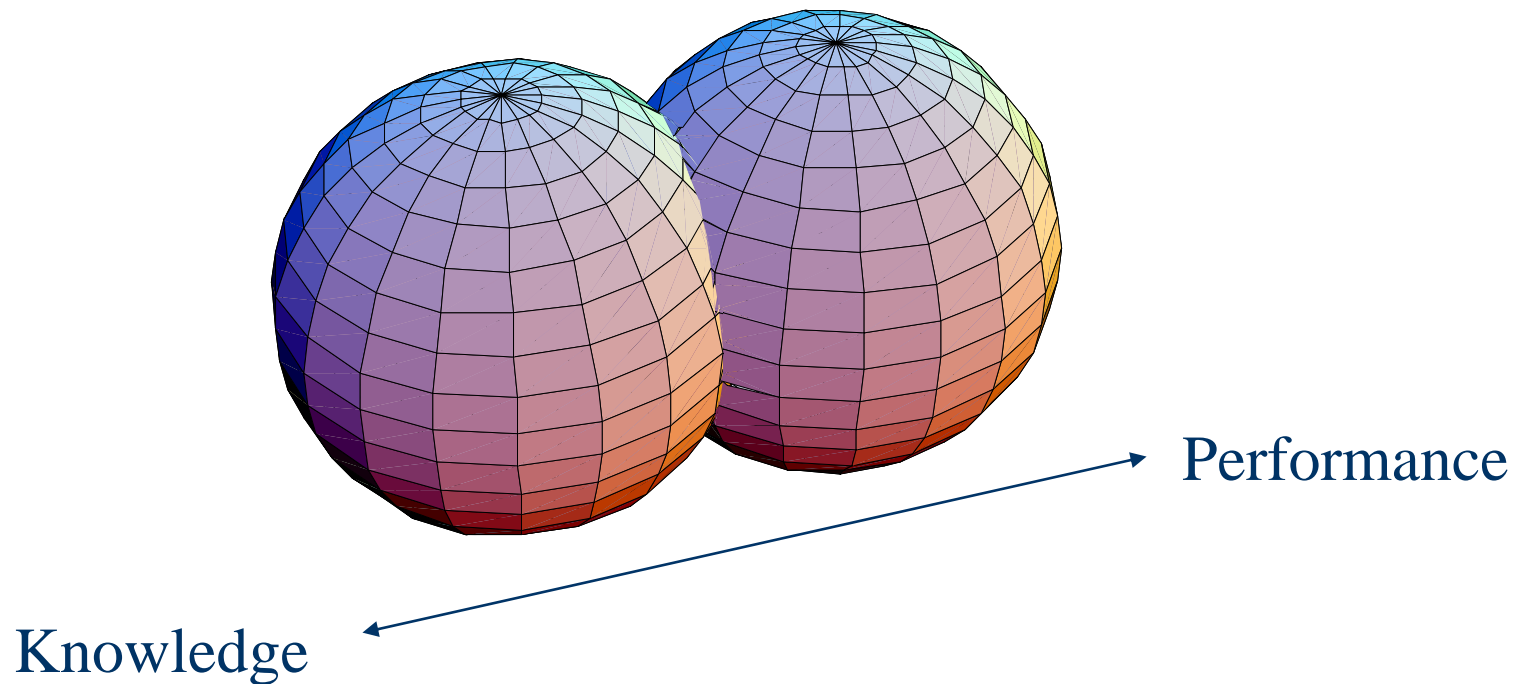
Why Implement IIT?

**Gather a different type
of information**

Domain



Performance without Knowledge can be incomplete



Ascertain IIT Mobility

- ◆ Do you know how information is obtained by professionals on the job?
- ◆ Do you have confidence in your scoring rubric development process?
- ◆ Do you have enough access to SMEs to serve as scoring judges?
- ◆ Does your profession have a clearly defined, widely accepted conceptual model for practice?

Ascertain IIT Mobility – Stakeholders Perspectives

- ◆ How quickly do tests need to be developed?
- ◆ How skilled are the item developers?
- ◆ How flexible is the test administration time?
- ◆ How skilled are your test-takers with the interface?
- ◆ What are your test delivery limitations?
 - What level of consistency is required of the test-delivery system?
 - Test delivery vendor specifications

Identify Possible Directions

- ◆ Add Non-Text Stimuli
 - Add Tables, Charts, Audio/Video Clips
 - Access reference material
 - Access internet help
- ◆ Expand Item Scoring Rubric
 - Dichotomous * Polytomous
- ◆ Expand Number of Responses / Item
 - Single * Multiple Response Option
- ◆ Add New Item Response Formats
 - Selected * Constructed Response

More Solutions

- ◆ Increase Task Complexity
 - Fill-in-the-Blank ✱ Essay
- ◆ Change Item Scoring Method
 - Predefined Rubric ✱ Judgment Based Review ✱
Automated Judgments
- ◆ Increase Interactivity
 - Case-Based Items ✱ Branching Cases ✱
Simulation

Gauge the Impact of the Solutions on the Stakeholders

- ◆ How much extra time to develop items?
- ◆ How much extra cost to develop items?
- ◆ How much is reliability changed?
- ◆ How operationally reliable are the items?
- ◆ How much extra time to answer the items?
- ◆ How much extra fidelity is achieved?



Did you think of this?

- ◆ Cheating
- ◆ Gaming
- ◆ Confounding variables
 - Reading comprehension
 - User computer literacy
 - Consistency of administrative experience

Collect Pilot Data

- ◆ ID – Development Time
- ◆ M – Development Cost
- ◆ Psy – Item Performance Characteristics by Innovation
- ◆ TDV – System Reliability
- ◆ TT – Customer Satisfaction with Item Type
- ◆ TU – Validity Study - Critical Incidents



Share Results



Bring your results to ATP!



Bang for the Buck

- ◆ Innovations should provide deliverables along the way
- ◆ Innovation deliverables should improve the quality and development of later innovations
- ◆ Assess the added value with each innovation

Bibliography

- ◆ Parshall, C. (In press) Innovative Item Types. Computer Based Testing: Building the Foundations for Future Assessments. NJ: Lawrence Earlbaum Associates.
- ◆ Drasgow, F. and Olson-Buchanan, J.B. (1999). Innovations in Computerized Assessment. NJ: Lawrence Earlbaum Associates.
- ◆ Haladyna, T.M. (1994) Developing and Validating Multiple-Choice Items. NJ: Lawrence Earlbaum Associates.
- ◆ Osterlind, S. J. (1998) Constructing Test Items: Multiple Choice, Constructed-Response, Performance and Other Formats. Boston : Kluwer Academic Publishers.



History of Innovative Assessment



- ◆ Performance Assessment
- ◆ Portfolio



What's unique about today's innovative item types?

- ◆ Interactivity
- ◆ Automated Scoring



Issues – Item Developers

- ◆ Time to Develop
- ◆ Ease of Content Creation
- ◆ Ease of Item Construction
- ◆ Item Accuracy



Issues – Test Taker

- ◆ Validity – Do the test tasks seem appropriate?
- ◆ Fidelity – Are the test tasks real?
- ◆ Cheating – Are other people cheating?



Issues – Managers



- ◆ Cost of Development
- ◆ Impact on Test Volumes
- ◆ Customer Satisfaction

Issues – Psychometricians

- ◆ Performance of Items
- ◆ Amount of Information Item Provides
- ◆ Test Quality
- ◆ Congruence between item type and cognitive task
- ◆ Dependency – Case, Item, Task, or Step
- ◆ Weighting – Cases, Items, Tasks, or Steps
- ◆ Precision and Accuracy of Scoring Rubric