

Innovative Computerized Test Items: A Review

Heath Schechinger, M.Ed.

The Center for Educational Testing and Evaluation

The University of Kansas

Definition

Innovative computerized test items: A test item that uses technologies that use features and functions of a computer to deliver assessments that do things not easily done in traditional paper-and-pencil format.¹

History

- Initial wave focused on comparing innovative item formats to their identical traditional formats.^{2 3 4 5 6}
- There have been numerous types innovative items developed (see ⁷ and ⁸), but only a few comparison studies that have been conducted (see ⁹ and ¹⁰).
- Another wave of research has been on effective ways to write traditional items ^{7 11 12}, and develop taxonomies ^{13 14 15 16 17 1 18}.
- Many researchers have emphasized the need to address questions about the essential psychometric properties of innovative items ^{19 20 21 9 22 8}, but the research has yet to be conducted ¹⁰.**

Classification

There are many ways to classify innovative item types. The most recent model ¹, consists of seven dimensions (below) that can be organized on a continuum that ranges from least to most innovative.

- 1.Assessment Structure:** The type of item presented to the test taker and how responses are collected (e.g., from constructed response to classic multiple choice).
- 2.Response Action:** What an examinee must physically do to answer a particular item (i.e., clicking a mouse, typing on a keyboard, or speaking into a microphone).
- 3.Media Inclusion:** Use of non-text media test items, such as graphics, audio, video and animation.
- 4.Level of Interactivity:** Level of interaction or adaptation to the individual examinee.
- 5.Complexity:** Degree of complexity is determined by the amount of information test takers must process to respond to an item.
- 6.Fidelity:** Fidelity pertains to a test's *reality*. Items closely resembling a real-world expression are considered to have high fidelity.
- 7.Scoring:** Tests that simply use computers to record responses and calculate the number of correct responses are no longer considered innovative. Rather, the term is reserved for the computerized tests requiring sophisticated scoring algorithms.

Formats

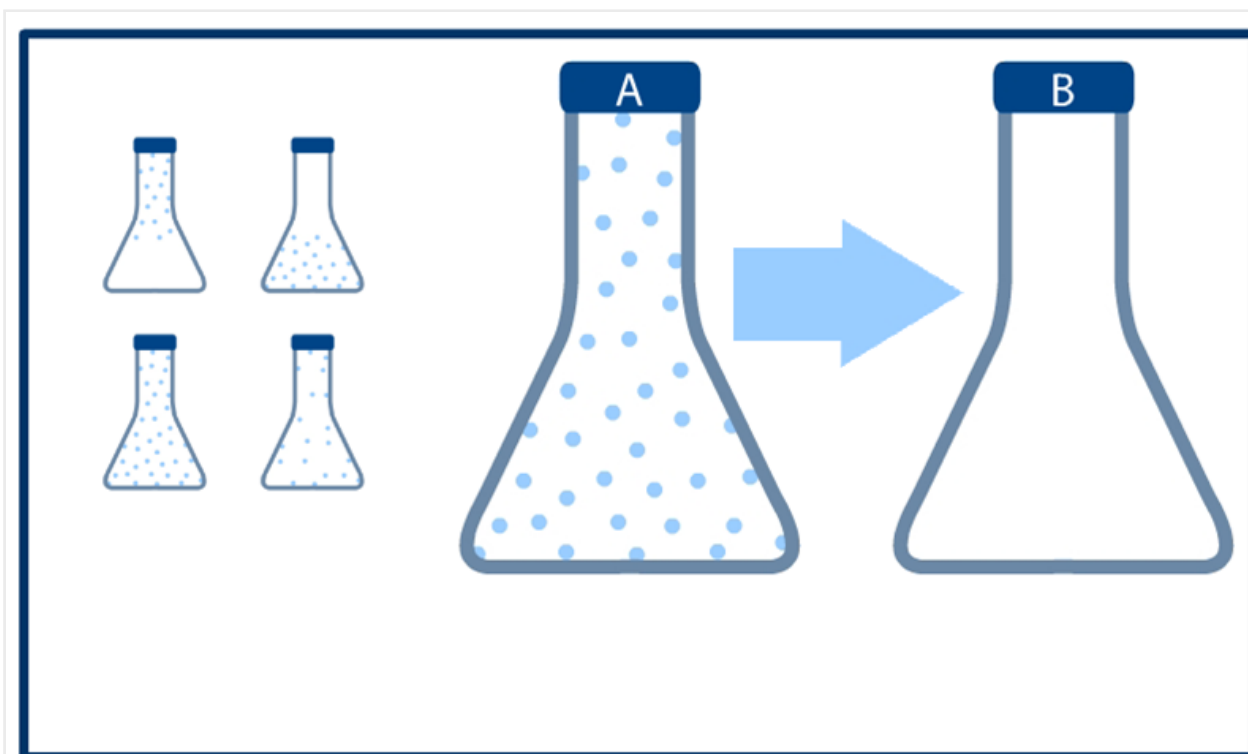
There are numerous types of innovative item formats. The following is the most recent comprehensive model, presented by Sireci & Zenisky in 2006:

1. Multiple Choice
2. Extended Multiple Choice
3. Multiple Selection
4. Specifying Relationships
5. Drag-and-Connect
6. Ordering Information
7. Select and Classify
8. Inserting Text
9. Corrections and Substitutions
10. Completion
11. Graphical Modeling
12. Formulating Hypotheses
13. Computer-Based Essay
14. Problem-Solving Vignette

Multiple Choice

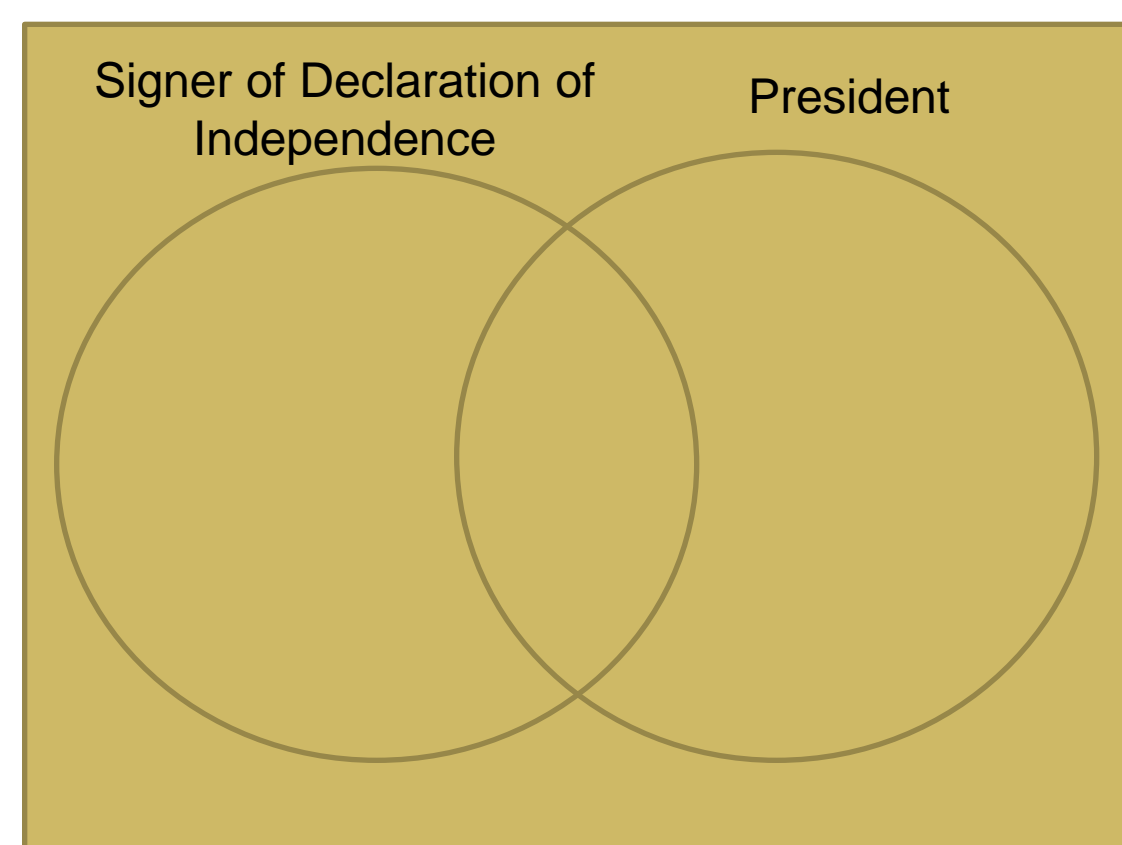
What do you think happens when half the air is let out?

- Drag and drop onto Flask B



Drag-and-Connect

Correctly place these U.S. Founding Fathers into the Venn diagram.



George Washington
John Hancock
Thomas Jefferson
James Madison
John Adams
Samuel Adams

Benefits

There is much enthusiasm for innovative item formats, and for good reason. They have been found to:

- Enhance validity ⁷
- Broaden the construct measured ²³
- More accurately emulate real-world situations (^{24 25 8}) through using graphics (²⁶), sound (²⁷) and video (²⁸)
- Improve motivation ²⁴
- Elicit positive reactions from students ^{25 29}

However, despite all the promise innovative items provide, it is critical that the new technologies are *improving* a test and not simply being used for the sake of innovation.

Conclusions

- There is a significant need for research addressing questions about the essential psychometric properties of innovative items.
- It cannot be assumed the increased 'realness' innovative items offer elicits appropriate reliability and validity.
- Not all innovative item-type format are the same- each needs to be researched independently on a variety of psychometric constructs.

•There have been very few studies comparing the psychometric properties of innovative item-formats compared to traditional formats (such as multiple choice).

•Until adequate research has been completed, it is difficult to conclude whether or not it is worth investing the extra capital to use an innovative item format.

•It is suggested that future innovative item research focus on measurement efficiency, reliability and validity, dimensionality, face validity, development costs, computer familiarity, scoring, and test security.

References:

1. Parshall, C. G., Harmes, J. C., Davey, T., & Pashley, P. (2010). Innovative items for computerized testing. In W. J. van der Linden & C. A. W. Glas (Eds.), *Computerized adaptive testing: Theory and practice* (2nd ed.). Norwell, MA: Kluwer Academic Publishers.
2. Greaud, V., & Green, B. F. (1986). Equivalence of conventional and computer presentation of speed tests. *Applied Psychological Measurement*, 10, 23-34. 1986-26215-001
3. Green, B. F. (1988). Construct validity of computer-based tests. In H. Wainer & H. Braun (Eds.), *Test validity* (pp. 77-96). Hillsdale, NJ: Lawrence Erlbaum Associates.
4. McBride, J. R., & Martin, J. T. (1983). Reliability and validity of adaptive ability tests in a military setting. In D. J. Weiss (Ed.), *New horizons in testing: Latent trait test theory and computerized adaptive testing* (pp. 223-236). New York: Academic Press.
5. Sympson, J. B., Weiss, D. J., & Ree, M. J. (1983). A validity comparison of adaptive testing in a military technical training environment (AFHLR-TR-81-40). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.
6. Sympson, J. B., Weiss, D. J., & Ree, M. J. (1984). Predictive validity of computerized adaptive testing in a military training environment. A paper presented at the annual convention of the American Educational Research Association, New Orleans, LA. (AFHLR-TR-81-40). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.
7. Sireci, S. G., & Zenisky, A. L. (2006). *Innovative item formats in computer-based testing: In pursuit of improved construct representation*. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
8. Zenisky, A. L., & Sireci, S. G. (2002). Technological innovations in large-scale assessment. *Applied Measurement in Education*, 15(4), 337-362. doi:10.1207/S15324818AME1504_02
9. Jodoin, M. G. (2003). Measurement efficiency of innovative item formats in computer-based testing. *Journal of Educational Measurement*, 40(1), 1-15. doi:10.1111/j.1745-3984.2003.tb01093.x
10. Gutierrez, S. (2009). Examining the psychometric properties of a multimedia innovative item format: Comparison of innovative and non-innovative versions of a situational judgment test. Ph.D. dissertation, James Madison University, United States -- Virginia. Retrieved July 28, 2011, from Dissertations & Theses: Full Text.(Publication No. AAT 3352759)
11. Haladyna, T. M., & Downing, S. M. (2004). Construct-irrelevant variance in high-stakes testing. *Educational Measurement: Issues and Practice*, 23(1), 17-27.
12. Haladyna, T. M., & Downing, S. M. (1989). A taxonomy of multiple-choice item-writing rules. *Applied Measurement in Education*, 2(1), 37-50.
13. Bennett, R. E., Ward, W. C., Rock, D. A., & LaHart, C. (1990). *Toward a framework for constructed-response items* (RR-90-7). Princeton, NJ: Educational Testing Service.
14. Harmes, J. C., & Parshall, C. G. (2005, November). *Situated tasks and simulated environments: A look into the future for innovative computerized assessment*. Paper presented at the annual meeting of the Florida Educational Research Association, Miami, FL.
15. Koch, D. A. (1993). Testing goes graphical. *Journal of Interactive Instruction Development*, 5, 14-21.
16. Parshall, C. G., Davey, T., & Pashley, P. (2000). Innovative item types for computerized testing. In W. J. van der Linden and C. Glas (Eds.), *Computer-adaptive testing: Theory and practice* (pp. 129-148). Boston: Kluwer Academic Publishers.
17. Parshall, C. G., Stewart, R., & Ritter, J. (1996, April). Innovations: Graphics, sound and alternative response modes. Paper presented at the annual meeting of the American Educational Research Association, New York, NY.
18. Scalise, K., & Gifford, B. (2006). Computer-based assessment in E-learning: A framework for constructing "Intermediate Constraint" questions and tasks for technology platforms. *Journal of Technology, Learning, and Assessment*, 4(6).
19. Bennett, R. E., Goodman, M., Hessinger, J., Kahn, H., Liggett, J., Marshall, G., et al. (1999). Using multimedia in large-scale computer-based testing programs. *Computers in Human Behavior*, 15, 283-294.
20. Drasgow, F., & Olson-Buchanan, J. B. (Eds.). (1999). Innovations in computerized assessment. Mahwah, NJ: Erlbaum.
21. Huff, K. L., & Sireci, S. G. (2001). Validity issues in computer-based testing. *Educational Measurement: Issues and Practice*, 20(3), 16-25.
22. van der Linden, Wim J. (2000). Optimal assembly of tests with item sets. *Applied Psychological Measurement*, 24(3), 225-240. doi:10.1177/01466210022031697
23. Bennett, R. E., Morley, M., & Quardt, D. (2000). Three response types for broadening the conception of mathematical problem solving in computerized tests. *Applied Psychological Measurement*, 24(4), 294-309. doi:10.1177/01466210022031769
24. Mezan, T. H., Avedon, M. J., Paese, M., & Smith, D. E. (1994). The effects of applicants' reactions to cognitive ability tests and an assessment center. *Personnel Psychology*, 47, 715-738.
25. Richman-Hirsch, W. L., Olson-Buchanan, J. B., & Drasgow, F. (2000). Examining the impact of administration medium on examinee perceptions and attitudes. *Journal of Applied Psychology*, 85, 880-887.
26. Ackerman, T. A., Evans, J., Park, K. S., Tasmassia, C., & Turner, R. (1999). Computer assessment using visual stimuli: A test of dermatological skin disorders. In F. Drasgow & J. B. Olson-Buchanan (Eds.), *Innovations in computerized assessment* (pp. 137-150). Mahwah, NJ: Erlbaum.
27. Vispoel, W. P. (1999). Creating computerized adaptive tests of musical aptitude: Problems, solutions, and future directions. In F. Drasgow & J. B. Olson-Buchanan (Eds.), *Innovations in computerized assessment* (pp. 151-176). Mahwah, NJ: Erlbaum.
28. Olson-Buchanan, J. B., Drasgow, F., Moberg, P. J., Mead, A. D., Keenan, P. A., & Donovan, M. A. (1998). Interactive video assessment of conflict resolution skills. *Personnel Psychology*, 51, 1-24. 1998-00850-001
29. Shottland, A., Alliger, G. M., & Sales, T. (1998). Face validity in the context of personnel selection: A multimedia approach. *International Journal of Selection and Assessment*, 6, 124-130.