ETS Research Spotlight

Assessment and Gaming

Foreword

Digital copies of this report can be downloaded free of charge from: **www.ets.org/research/spotlight8**

To comment on any article in this issue, write to: **RDWeb@ets.org**

Computer games and mobile apps have become ubiquitous, and are quickly spreading to the world of education, where educators and entrepreneurs are exploring new ways to use electronic games as educational tools.

We should therefore not be surprised at the growing interest in game-based assessments, but how do we make sure that games that claim to be educational live up to their promise? How do we build games that both support education by conveying and encouraging learning, and allow educators and educational policymakers to learn about what the students have learned?

The featured research synopsis in this issue of *ETS Research Spotlight* summarizes an essay titled *Three Things Game Designers Need to Know About Assessment*, which appeared as a chapter in a book on game-based assessment. The lead author of the essay was Robert J. Mislevy, ETS's Frederic M. Lord Chair in Measurement and Statistics. Mislevy and his co-authors based the essay on their experiences using evidence-centered design as part of a team of researchers from the Cisco Networking Academy to develop educational games that Cisco® could use when training networking students.

This work began in the late 1980s and early 1990s, using a simulation prototype called NetPASS, and eventually resulted in an "epistemic game" called Aspire. The essay reflects 10 years of work growing from a pioneering collaboration between researchers from ETS and Cisco on simulation-based assessment, and provides lessons for computer game designers working to develop educational games and game-based assessments. It was thanks to such rigorous research that Aspire became a successful tool for education, and one of the main lessons of this article is that we have to take learning and assessment seriously when developing educational games and new forms of assessment.

This essay is part of a growing body of work on game-based assessment, a field that ETS and its researchers are taking seriously. We are fortunate to be able to draw on a broad range of sciences and tap into seasoned experts such as Robert Mislevy and his colleagues. For anybody venturing into this new field, this article provides a natural starting point.

Ida Lawrence,
Senior Vice President,
Research & Development

Three Things Game Designers Need to Know About Assessment

Introduction

Three Things Game Designers Need to Know About Assessment is an essay that discusses an emerging field where computer game development and educational assessment are coming together. The essay also discusses the role of evidence-centered design (ECD) for building games that are potentially both engaging and able to capture reliable and valid evidence of learning.

The three things game designers need to know are:

1. Assessment and game design are based on the same principles of learning.

2. Assessment design isn't fundamentally about tasks and scores, but about the structure of reasoning — reasoning from what students say and do, to understand what they know and can do.

3. Designers of game-based assessments must consider key constraints of both assessment and game design right from the start.

The essay is by Robert J. Mislevy of ETS and co-authors John Behrens, Kristen DiCerbo, Dennis Frezzo, and Patti West.[1] It appears as Chapter 5 of *Assessment in Game-Based Learning: Foundations, Innovations, and Perspectives,* published by Springer and edited by Dirk Ifenthaler, Deniz Eservel, and Xun Ge. The book brings together international experts to address both theoretical and practical aspects of learning in a computer-based games setting, including key issues in research, methodology, assessment, and technology.

The 488-page volume has three parts. *Part one* provides the background for game-based learning and assessment and discusses what game designers need to know in order to build games that also are valid and efficient tools of assessment. *Part two* explores new ways of evaluating learning in

> **Editor's note:** *The full reference list appeared in the original work, which was:*
>
> Mislevy, R. J., Behrens, J. T., Dicerbo, K. E., Frezzo, D. C., & West, P. (2012). Three things game designers need to know about assessment. In D. Ifenthaler, D. Eservel, & X. Ge (Eds.), Assessment in game-based learning: Foundations, innovations, and perspectives (pp. 59–81). New York, NY: Springer New York.

---

[1] John Behrens, Kristen DiCerbo, Dennis Frezzo, and Patti West worked at Cisco when the article was written. Frezzo continues to work for Cisco, while Behrens and DiCerbo now work for Pearson®. West is currently at Santa Fe College.

computer-based gaming environments that can provide both individualized assessment and feedback to learners. *Part three* explores the latest empirical research findings and successful examples of game-based assessment.

The authors of *Three Things Game Designers Need to Know About Assessment* believe that combining these fields opens new opportunities for educational assessment, but also warn that the process comes with significant challenges. The essay highlights the need to bring together diverse expertise in order to meet these challenges, and recommends using evidence-centered design to create tasks that are not only interactive and engaging, but also provide useful evidence of what test takers know and can do.

The relation between assessment and game design

For the authors, both assessment and game design seek to engage people by placing them in situations where they must apply knowledge to face challenges. The authors stress that, while game designers focus on engaging the player, assessment designers focus on getting information that reveals what people have learned. These are different — but not necessarily mutually exclusive — constraints.

Building on the principles of learning, the authors note that humans excel at pattern recognition, whether it is interpretation of sounds, cultural knowledge, or interpersonal behavior. Humans can become experts at certain tasks or experts in a field through reflective practice and by learning from feedback. The authors propose that such expert skills can be developed using a computer game's simulated environment, which places an individual in challenging situations and then provides feedback in response to choices made during the game play. An individual participant learns while pushing his or her capabilities to the limit in a highly motivated state.

While game designers focus on engaging the player, assessment designers focus on getting information that reveals what people have learned. These are different — but not necessarily mutually exclusive — constraints.

The authors see Cisco Networking Academy's Aspire as an example of an "epistemic" educational game,[2] which allows a player to develop certain professional knowledge and skills. The players of Aspire act as networking consultants in a virtual world designed to assess their ability to not only design and troubleshoot networks, but also work with customers and make business decisions.

Interaction plays a key role in games and game-based assessments, and it can be used as a feedback loop, according to the authors. A computerized adaptive test can present the test taker with a complex interaction loop, where each response leads to new tasks and new levels of difficulty. Games function in a similar way, allowing the player to move along different paths and levels depending on the skills he or she reveals. The authors

[2] The term "epistemic games" comes from David Shaffer at the University of Wisconsin at Madison, and refers to games in which a player develops real-life professional skills, knowledge, identities, values, and epistemology (SKIVE).

argue that while both games and computer-based assessments are compatible and build on learning principles, game-based assessment comes with design constraints that reflect the goal of measuring the players' knowledge and skills.

The structure of reasoning is central to assessment design

In order to function well as an assessment, the authors argue, a game must be designed so that it captures information about the gamers' thought processes. Traditionally, tests have been thought of as collections of questions, or items. The test-taker responses result in assessment data that applied statisticians, or psychometricians, analyze using a range of statistical models and functions. Getting good assessment information from a game, Mislevy et al. point out, may require models that allow for appropriate use of data not only from the results of test-taker actions, but also from the processes they used to achieve those results.

Some game-based assessments are evaluated using relatively simple models that count the number of correct answers, but the authors note that the level of complexity and the challenges to measuring the performance increases in simulated game environments where action sequences can be tracked and time stamped. This requires more than just adding up a numerical score. For the assessment developer, this raises questions about targets of inferences, sources of evidence, and ways of evaluating the activity, which the writers see as essential for determining what the player learns during the game.

Next, the authors introduce the ECD approach and explain how ECD-based models can be used to design assessments based on more complex interactive activities like Aspire. They assert the need to consider reliability and validity in order to extend the toolkits we need when drawing inferences from game-based assessments. As game-based assessments become more complex, Mislevy et al. point out, they can present more of a challenge here because test takers face different situations as they move through the game and may choose different paths depending on their responses.

The authors argue that while both games and computer-based assessments are compatible and build on learning principles, game-based assessment comes with design constraints that reflect the goal of measuring the players' knowledge and skills.

Getting good assessment information from a game, Mislevy et al. point out, may require models that allow for appropriate use of data not only from the results of test-taker actions, but also from the processes they used to achieve those results.

The authors believe that these challenges to managing evidence can be addressed by adapting some existing psychometric methodologies and point to two examples:

- diagnostic classification models (analyze multidimensional feedback with many categorical variables representing skills underlying a test)

- Bayesian inference networks (analyze uncertainty and outcomes by combining common-sense knowledge and observational evidence in a graphical statistical model)

The authors write that game designers need not master these models, but that they should be familiar enough with them to use them effectively. The models are important for developing games because they allow the designers to see how different actions, situations, and modes of play affect both the evidence gathered and the level of engagement.

When it comes to validity, the authors see a more complex challenge for game-based assessments: It is not clear, for example, that the problem-solving skills acquired playing a game like World of Warcraft® transfer into skills that are useful in the real world. According to Mislevy et al., the "question of transfer is a critical aspect of validity in game-based assessment."

Following Messick (1994[3]), they stress the need for game designers to leave the right features in and leave overly distracting features out of the system. The focus of assessment designers in this context should be on student actions that reflect the underlying principles of the system about which the students are learning.
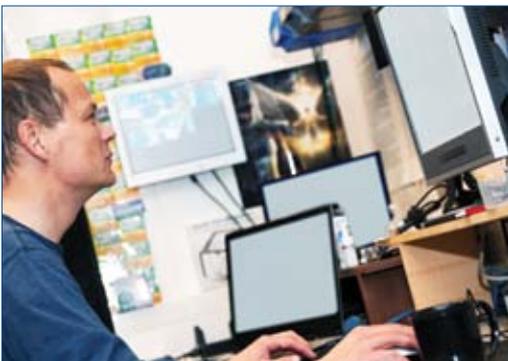
Designers of game-based assessments must consider key constraints of both assessment and game design right from the start.

Key constraints should be addressed early on

The authors argue that assessment and game designers need to address early on the key constraints that they must meet for the game environments to be useful as educational assessments. The product should be able to both engage students and provide useful assessment information.

The authors argue that "to first design a great game, then try to figure out 'how to score it,' is not a good way to design a game-based assessment" (p. 75). A number of problems will occur if the developers have not tried to address key constraints early on in the process. Instead, the assessment designers should begin by considering both key assessment principles and those for

---

[3] Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23(2), 13–23. Messick states, "The issue, as always in test validation, becomes the nature of the evidence accrued to counter the two major threats to construct validity, namely, construct underrepresentation (which jeopardizes authenticity) and construct-irrelevant variance (which jeopardizes directness)" (p. 14).

design of games and computer simulations. The evidence argument should be brought in early on so that the developers can focus on the kinds of knowledge and skills we want the students to display in the assessment.

Discussion and the way forward

When the authors look to the future, they suggest a path forward along three lines of work. The first entails learning from previous work to develop computer simulations and game-based assessments. The second is theory, and seeks to discover and make the underlying concepts and principles explicit while creating a shared language that experts coming from different backgrounds can share. The third calls for creating reusable elements that improve the efficiency of building game-based assessments.

You May Also Be Interested In …

Listed below are some ETS resources and publications related to research on gaming and assessment.

RESOURCES

In response to fundamental shifts in societal views about teaching, learning, and assessment, ETS has created a new Center for Advanced Psychometrics (CAP). The Center will focus on developing and applying psychometric models to support the development, analysis, and scoring of a new generation of assessments.

The center will be part of ETS Research & Development's Statistical Analysis, Data Analysis, and Psychometric Research area. It will eventually tackle issues such as:

- data collected from simulations and intelligent tutoring systems

- dynamic models

- statistical models for complex assessment tasks of educational constructs that integrate cognitive science with psychometrics

- psychometric modeling for novel constructs such as collaborative problem solving

- data mining techniques applied to assessment data

Learn more at **www.ets.org/newsroom**.

PUBLICATIONS

Citation records for these and other publications are available in the ETS ReSEARCHER database at **http://search.ets.org/researcher**. *Where available, the database contains the full text of ETS-published works and links to the publisher's abstract for works that ETS staff members published in independent venues, such as journals and books.*

2012

Exploring the Role of Games in Educational Assessment

D. Zapata-Rivera & M. I. Bauer (2012)

Chapter in *Technology-Based Assessments for 21st Century Skills: Theoretical and Practical Implications from Modern Research,* pp. 149–171

Editors: M. C. Mayrath, J. Clarke-Midura, D. H. Robinson, & G. Schraw

Publisher: Information Age Publishing

2009

Melding the Power of Serious Games and Embedded Assessment to Monitor and Foster Learning: Flow and Grow

V. J. Shute, M. Ventura, M. I. Bauer, & D. Zapata-Rivera (2009)

Chapter in *Serious Games: Mechanisms and Effects*, pp. 295–321

Editors: U. Ritterfeld, M. J. Cody, & P. Vorderer

Publisher: Routledge

Combining Learning and Assessment in Assessment-Based Gaming Environments: A Case Study from a New York City School

D. Zapata-Rivera, W. H. VanWinkle, B. Doyle, A. Buteux, & M. I. Bauer

Interactive Technology and Smart Education, Vol. 6, No. 3, pp. 173–188

2008

Monitoring and Fostering Learning Through Games and Embedded Assessments

V. J. Shute, M. Ventura, M. I. Bauer, & D. Zapata-Rivera

ETS Research Report No. RR-08-69

About ETS

At ETS, we advance quality and equity in education for people worldwide by creating assessments based on rigorous research. ETS serves individuals, educational institutions and government agencies by providing customized solutions for teacher certification, English language learning, and elementary, secondary and post-secondary education, as well as conducting education research, analysis and policy studies. Founded as a nonprofit in 1947, ETS develops, administers and scores more than 50 million tests annually — including the *TOEFL®* and *TOEIC®* tests, the *GRE®* tests and *The Praxis Series™* assessments — in more than 180 countries, at over 9,000 locations worldwide.

Scan this code to read the publication
on your mobile device.

21913

Listening. Learning. Leading.®