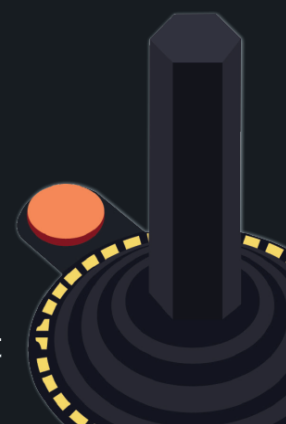


game-based assessment

an interdisciplinary workshop integrating
organizations, education and assessment



August 22-23, 2019 at the Graduate Hotel Minneapolis

This conference is sponsored by the National Science Foundation (USA). They have covered the rental space, setup costs, breakfast, lunch, coffee breaks, and travel grants for presenters and grad students. Please thank your neighborhood NSF program officer!

Please note all conference events on August 22-23 occur on the Mezzanine level. To access it, please use the prominent double escalators located in the main lobby of the hotel. The conference registration desk will be visible from the top of the escalator.

Workshop Agenda

Wednesday, August 21

5:00 pm – 8:00 pm	<i>Early Registration, Welcome Happy Hour (pay-your-own), and Golf/Zombie Simulator</i>	Ground Level, Topgolf Suite
-------------------	---	--------------------------------

Thursday, August 22

		<u>Mezzanine Level</u>
8:00 am – 9:00 am	<i>Registration</i>	Top of Escalator
8:00 am – 9:00 am	<i>Breakfast</i>	Meridian Foyer
9:00 am – 9:30 am	Welcome and Goals Discussion	Meridian 3 4
9:30 am – 10:30 am	Opening Keynote Richard Landers, University of Minnesota	Meridian 3 4
10:30 am – 11:00 am	<i>Morning Coffee Break</i>	Meridian Foyer

11:00 am – 12:30 pm	Organizational Sciences Research Symposium Andrea Sinclair, HumRRO Sarena Bhatia, Kincentric Sebastian Loh, Southern Illinois University Elizabeth Short, QCHI Tim Warszta, West Coast Univ of Applied Sciences	Meridian 3 4
12:30 pm – 1:30 pm	<i>Lunch</i>	Meridian 1 2
1:30 pm – 3:00 pm	Educational Research Symposium Karrie Godwin, Kent State University Russell Almond, Florida State University Mina Johnson-Glenberg, Arizona State University Tanner Jackson, Educational Testing Service Richard Halverson, University of Wisconsin-Madison	Meridian 3 4
3:00 pm – 3:30 pm	<i>Afternoon Coffee Break</i>	Meridian Foyer
3:30 pm – 5:00 pm	Vendor Research Symposium Jason Blaik, Revelian Kristen DiCerbo, Pearson Lara Montefiori, Arctic Shores Kelly Trindel, Pymetrics Alina Siemsen, Aon's Assessment Solutions Jim Wexler, Persona Labs	Meridian 3 4
5:00 pm – 6:00 pm	<i>Closing Remarks and Networking Hour</i>	Meridian Foyer
6:30 pm – 9:30 pm	<i>Conference Dinner</i>	TBD

Friday, August 23

		<u>Mezzanine Level</u>
8:00 am – 9:00 am	<i>Breakfast</i>	Meridian Foyer
9:00 am – 9:30 am	Welcome Back and Announcements	Meridian 3 4
9:30 am – 10:30 am	Invited Keynote Sidney D'Mello, University of Colorado Boulder	Meridian 3 4
10:30 am – 11:00 am	<i>Morning Coffee Break</i>	Meridian Foyer

11:00 am – 12:30 pm	<p>Graduate Student Top Research Symposium</p> <p>Elena Auer, University of Minnesota</p> <p>Cassandra Eng, Carnegie Mellon University</p> <p>Chaima Jemmali, Northeastern University</p> <p>Laura Levy, Georgia Institute of Technology</p> <p>Jenna McChesney, North Carolina State University</p>	Meridian 3 4
12:30 pm – 1:30 pm	<i>Lunch</i>	Meridian 1 2
1:30 pm – 3:00 pm	Research Priorities and Planning Discussion	Meridian 3 4
3:00 pm – 3:30 pm	<i>Afternoon Coffee Break</i>	Meridian Foyer
3:30 pm – 4:30 pm	<p>Closing Keynote</p> <p>Fred Oswald, Rice University</p>	Meridian 3 4
4:30 pm – 5:00 pm	Closing Remarks	Meridian 3 4

Presentation Summaries

Thursday 11:00AM Organizational Science Symposium

Developing games to measure personality—it's not all fun and games!

Andrea Sinclair, HumRRO

This presentation will discuss how a team of practitioners, academics, and game-developers worked, over a three-year period, to develop computer-administered, game-like performance tests (GLPTs) to measure personality (facets of conscientiousness) in a way that reduces the impact of faking on personality measures. We'll discuss the challenges and pitfalls encountered, as well as lessons learned along the way. We'll provide an overview of the research findings and conclude with potential next steps.

Getting into the Game: Applicant Reactions to Game-Based Assessments

Sarena Bhatia, Kincentric

The purpose of this study was to compare how applicants respond to use of a game-based assessment (GBA) in comparison to a traditional assessment in a selection context. This study utilizes a 2 (GBA, traditional selection assessment) by 2 (made an offer, not made an offer) experimental time-lagged design. Participants were asked to imagine they were applying for a fictional role, and then were randomly assigned to one of the assessments. The next day, they were emailed either an acceptance or rejection letter. Those in the GBA condition were significantly more likely to experience higher rates of flow. Taking the GBA was positively related to perceived job-relatedness (when mediated by flow), justice perceptions, perceived person organization fit, organizational attractiveness and positive intentions to accept the offer. Selection decision was a moderator in that those who were made an offer and perceived the assessment to be job relevant had the highest perceptions of justice. This study is one of the first to introduce and measure the role flow to the assessment literature, and to shed light on the relationship between GBAs and many important applicant reactions outcomes.

Expert Similarity Index (ESI): A Serious Games Analytics Performance Index for Organizational Training Outcomes

Sebastian Loh, Southern Illinois University-Carbondale

The design of serious games strongly affects what types of data can be collected *in situ*, which in turn, affects the analytics produced. Serious games can create values for organization performance through *repeatable events* (i.e., repetitions) aimed at affording metacognition, such as promoting practice-to-mastery, building memory, learning from mistakes, improving strategy, and decision-making. These repetitions also form the necessary data for repeated measures statistics, trainee profiles, and predictive analytics. Expert-novice differences in game-based training are easily distinguishable. Novices should act/ behave increasingly similar to the experts as the training (and game) progresses. Using experts' in-game actions as an achievement baseline, we can accurately quantify how similar a person's actions are to that of the experts' by way of the Expert Similarity Index (ESI). The ESI metric (0 to 1) has a wide range

of use for the calculation of individual and organizational performance analytics with serious games. As an index, the ESI is highly functional in estimating the cost-benefit ratio of training, creating trainee profiles (e.g., scoring and ranking individual growth potentials), and determining *training prescriptions*—the amount and frequency of training—to manage organizational performance outcomes.

Massive Multiplayer Online Experiences: Using Recreational Gameplay Behaviors to Inform Assessment Design

Elizabeth Short, QCHI

Studying emergent gameplay behaviors from recreational gameplay may help to inform our understanding of the design and operation of game-based assessments. Massively Multiplayer Online (MMO) games, in particular, offer a plethora of opportunities to observe and objectively record recreational gameplay behaviors. This may both provide a form of assessment in itself as well as inform the design of more targeted future assessments. The inherently social nature of MMO gameplay combined with the vast array of experiences these games offer makes them an ideal sandbox to explore how personality and individual differences shape emergent gameplay behaviors. This symposium will explore what designers of game-based assessment can learn from examining how players interact with videogames, in particular MMOs, to evaluate non-cognitive predictors of behavior. This information may be useful for informing the design of game-based assessments for non-cognitive predictors such as personality and other individual differences. This presentation will include a discussion of how personality theory can be used to better understand and address the challenges associated with designing game-based assessments that will serve as both reliable and valid predictors of organizational behavior.

The Impact of Game-Design Elements on Participants' Reactions in Gamified Assessment

Tim Warszta, West Coast University of Applied Sciences

Purpose: Gamified assessment is increasingly used in recruitment and selection. The purpose of this study is to investigate the impact of different design elements on participants' reactions towards a gamified assessment in self-assessment and selection settings. **Design/Methodology:** A policy-capturing design experiment was conducted. In a 2x2x2x2 design sixteen vignettes were formulated testing the impact of the four independent variables cover story (business vs. fantasy), levels (levels vs. no levels), graphical elements (pictures vs. no pictures), and instant feedback (feedback vs. no feedback) on the dependent variables process fairness (PF) and acceptance in self-assessment (ASA). In addition, the impact of participants' gaming experience was captured. **Results:** Strongest impact on PF and ASA was found for the cover story (a business context was perceived more favorable) while levels and graphical elements explained additional variance. Gaming experience correlated positively with PF and ASA. **Limitations:** Results were obtained in a policy-capturing design and should be replicated in field studies. **Practical Implications:** A cover story that uses a business setting, levels, and pictures could be used to improve participants' reactions towards gamified assessments. **Originality/Value:** To my

knowledge, this is the first empirical study that analyses GBA elements systematically in an experimental design.

Thursday 1:30PM Educational Research Symposium

Creating a Game-based Assessment to Measure Children's Attention

Karrie Godwin, Kent State University

Cognitive assessment batteries can provide valuable information for researchers as well as practitioners. However, if participants fail to complete the entire assessment, the utility of the assessment battery is hampered. Strategies to reduce attrition are important, particularly when working with young children, as children's task engagement can wane rapidly. Game-based assessments provide one potential solution to this problem. In this presentation, I discuss research in which we use a cognitive assessment to index children's selective sustained attention and predict children's learning outcomes. We then redesign this assessment as an engaging video game, *Monster Mischief*. Through an experimental study we explore whether the video game maintains its validity as an assessment while simultaneously increasing children's enjoyment and motivation. Future directions will also be discussed.

A Four-Process Implementation of Game-Based Scoring

Russell Almond, Florida State University

Almond, Steinberg and Mislavy (2002) introduced a four-process architecture for adaptive assessments. The four processes were the Presentation Process, the Evidence Identification (EI) process, the Evidence Accumulation (EA) process, and the Activity Selection (AS) process. For the game *_Physics Playground_*, we built a scoring engine implementing these four processes. The presentation process (game engine) was written in Unity and logged events through Learning Locker. The work product from playing a game level was a collection of event statements in xAPI format. The EI process used a newly developed event processing language, *_EI-Event_*. The EI process converted the events to a collection of observed outcome variables for each game level. The EA process estimated the players' proficiencies on Physics skills using a Bayes net (Netica). Its output was proficiency estimates for the player. The AS process was some simple adaptivity rules using the estimated proficiencies embedded in the game engine. Communication between processes was done through a document database used to queue messages with a simple web interface. This scoring model was field tested in a moderately sized study in spring of 2019. Many of the components are open source and available from <https://pluto.coe.fsu.edu/Proc4>.

How to Assess the Educational Quality of XR Content

Mina Johnson-Glenberg, Arizona State University

Dr. Johnson-Glenberg both creates immersive Augmented/Virtual Reality (XR) content, and she assesses the content's efficacy. She has published guidelines on how to create embodied, STEM content in VR (Johnson-Glenberg, 2018 in *Frontiers in Robotics and AI*), and is now focused on creating a rubric to quantify the pedagogical "quality" of educational content in immersive VR. She encourages designers to implement creative and active uses of assessment that can be

seamlessly embedded in sims/games. She will present the 20-item quality rubric her lab has created and is actively seeking feedback from the audience. Examples of free STEM games can be found at her spin-out company site www.embodied-games.com.

Awkward Annie: Impacts of Playing on the Edge of Social Norms

Tanner Jackson, ETS

Effective interpersonal and cross-cultural communication relies on pragmatics – knowing what to say to whom, and under what circumstances. Nevertheless, pragmatics is generally absent from formal second language instruction. The current effort describes a game designed to assess people’s pragmatic ability. In the game, Awkward Annie, players are asked to intentionally select the most inappropriate things to say within conversations (i.e., be inappropriate and see what happens). Thus, players are able to escape from reality by being inappropriate. This presentation will discuss an empirical exploration and evaluation of specific game-based assessment design choices and how they impact aspects of user performance, preference, and affect.

Design and Assessment Guidelines to Make Compelling Games for Learning

Richard Halverson, University of Wisconsin-Madison

Games provide a variety of models for successfully engaging players. Games for learning, however, must engage players as well as teach desired content. Meeting these two goals well requires an iterative, data-rich collaboration process between developers, content experts, and learning specialists. In this talk, I will discuss our experiences at the University of Wisconsin-Madison to build games for learning in biology, environmental science, mindfulness and professional development. I propose design and assessment guidelines for coordinating the expertise and data necessary to build compelling games for learning, and discuss how to use assessment to avoid the ways in which the development process can go wrong.

Thursday 3:30PM Vendor Research Symposium

Game-Based Assessment for High-Stakes Personnel Selection

Jason Blaik, Revelian

This presentation will provide insight into the development and validation of two game-based assessments designed as psychometric tools for use in personnel selection, i.e. Cognify and Emotify. Cognify is a measure of cognitive ability which targets broad categories of the Cattell, Horn and Carroll theory of cognitive abilities (Schneider & McGrew, 2018). Emotify represents a measure of emotional intelligence underpinned by Mayer & Salovey’s (1997) four branch model of EI. Factors driving the development of these GBAs will be initially discussed, followed by a review of the development process and the design principles guiding such work. A brief treatment of the validation efforts and psychometric properties of each measure will be presented, including estimates of reliability and evidence of face, construct and criterion validity. The application of modern prediction methods, e.g. regularised regression, decision trees, to the scoring of these assessments will be shared, along with recent experiences in automated item generation and translation, as examples for the field more broadly.

Demonstration of various components of these assessments will feature throughout this presentation and some of the many lessons learnt from their development happily shared.

Lessons Learned in Game-Based Assessment Design

Kristen DiCerbo, Pearson

A decade of work designing and researching game-based assessments at two large companies has led to many lessons learned regarding design processes and ultimately to an understanding of current challenges to the scaling of game-based assessment. Regarding design processes, we have learned about the make-up of design teams, the merging of assessment design and game design processes, and the need to think about educational systems. We have identified five key elements required for game-based assessment design. In addition, we have uncovered numerous challenges that continue to stand in the way of scaling good game-based assessment. The fall into three categories: communication challenges, validity/reliability/fairness challenges, and time and budget challenges. I will discuss these challenges along with potential approaches to addressing them.

Five years of Arctic Shores: Evidence, Insight, and Future directions

Lara Montefiori, Arctic Shores

This session will cover a broad spectrum of topics, with evidence and insight collected in the first five years of Arctic Shores. This will include an overview of the scientific rationale underlying the assessment, client case studies, evidence from University-based research, thoughts about possible future developments in terms of assessment itself but also in terms of advancement in the scientific understanding, taxonomies, and regulations of the field.

Game-Based Machine Learning Technology to Optimize for Fairness and Validity in Employment Selection

Kelly Trindel, Pymetrics

Historically, psychological assessment for employment selection is challenging and has left significant room for improvement. Test-takers can intuit the intent behind survey items, and can cheat on objective questionnaires. Many assessments are directional and general, meaning that the definition of “success” is the same for all people, for all roles. In addition, many assessments suffer from fairness issues, disadvantaging candidates based on their demographic background. Here, we present findings from the first five years of pymetrics’ implementations assessing over 1 million people. We illustrate the benefits of game-based assessments that wed decades of empirical research with modern machine learning techniques to create a custom assessment approach optimized for fairness and validity. We showcase methods for testing criterion-related validity and fairness estimation and remediation before an assessment even goes live. Finally, we share case studies on the results of game-based ML technology as relevant to real-world job candidates.

smartPredict – Development Insights and Study Results of Aon’s Gamified Assessment Series

Alina Siemsen, Aon's Assessment Solutions

Game-based Assessments come in various forms and appearances (Siemsen, 2019; Siemsen & Warszta, 2018), resulting in the need of looking at those assessments more granularly. Beginning with establishing a shared understanding of “Gamified Assessments” as opposed to “Game-based Assessments”, this presentation will give insights on the development rationale and journey of Aon’s gamified cognitive assessment series “smartPredict”. It highlights how it combines sound psychological research, interdisciplinary teams and game design elements, and shares learnings that have been made on the way. Following the guidelines of numerous organizations such as the Standards for Educational and Psychological Testing, the SIOE Principles for the Validation and Use of Personnel Selection Procedures, and the Uniform Guidelines on Employee Selection Procedures, assessments that are used for personnel selection need to fulfill certain quality standards. This presentation aims at giving an overview of our research on the smartPredict assessments, describing the studies and our findings on topics such as validity, mobile-desktop equivalence, adverse impact, and applicant reactions. The presentation will close with an outlook and open questions, which may serve as the basis for following discussions

The "Gee" in GBAs

Jim Wexler, Persona Labs

Providers offer many different instruments, but our industry has a common thread: games. Gamification is meant to playfully induce behavior -- to get people to participate in a way that delights them. This feeds a virtuous circle: as users give up more data, the experiences become more insightful and relevant, delivering more value to both the users and to talent-seekers. Do today's GBAs achieve this? How can we improve these experiences to better serve the next generation of employees, clients and constituents for whom game dynamics, interfaces and interactions are second nature? How does gamification benefit talent, and those who hire them? We'll review research on how gamification's processes parallel the creation of a complex and ever-evolving digitally-enabled enterprise, and look at the fundamental driving design goals for effective game experiences as a touchstone into understanding human motivation: Autonomy: Having a sense of being in control of their own destiny. Mastery: A sense of continuous progress. Purpose: A sense that this is about more than just me, of being connected to a larger purpose and community.

Friday 11:00PM Graduate Student Top Research Symposium

Leveraging Trace Data in Game-Based Assessments

Elena Auer, University of Minnesota

Game-based assessments (GBAs) typically produce two types of data that are particularly relevant to construct measurement. One of these types is a game score, which is a planned a priori measurement of targeted constructs resulting from gameplay. The second type, trace data, refers to the thousands of micro-behavioral data points like mouse clicks and time stamps that are also captured during a GBA. Such data can be restructured and analyzed to potentially improve upon game score construct measurement. This presentation discusses an empirical study that examined if trace data could be untapped potential for improving construct

measurement and predicting criteria. Five hundred and thirty undergraduate student participants completed both traditional cognitive ability tests and a set of mini-games designed to measure general cognitive ability. First, features were engineered from log files and used as input in a series of machine learning models to predict participant scores on traditional cognitive ability tests. Generally, trace data seemed similarly effective to composite game score in predicting out-of-sample traditional cognitive ability test scores. Second, we assessed the incremental validity of trace data over the game score when predicting GPA. Results suggest that trace data can, to some degree, incrementally predict GPA over the game score.

Education and the Brain, a Bridge Not Too Far: Game-Based Assessment and Neuroplasticity

Cassandra Eng, Carnegie Mellon University

Engagement through player adaptability and changing task demands in a virtual environment are features that make cognitive game-based assessment (GBA) ideal for children due to low attention spans and the scarcity of developmentally sensitive assessments. This study provides novel insight into the design process and validity of implementing game-based features into a standard cognitive assessment, the association between GBA performance and academic achievement outcomes, and the underlying neural mechanisms associated with GBA-induced learning. Brain connectivity utilizing noninvasive functional near infrared spectroscopy (fNIRS), performance on a game-based cognitive assessment and widely-used cognitive task, enjoyment measures, and standardized academic assessments were administered to preschool-aged children. Children exhibited higher enjoyment with the addition of specific game features to the cognitive assessment, while still preserving task validity. Furthermore, participating in a physically-active adaptation of the GBA showed increased neuroplasticity—the building of new brain connections to adapt to challenging environments—when children learned novel skills via active game play. Incorporating neurophysiological measures into GBA may identify the mechanisms for behavioral changes in learning, and developing similar GBAs may encourage young children to perceive cognitive assessments not as tedious, but fun, brain-developing activities.

Insights on Debugging Processes of Beginner Programmers in an Educational Puzzle Game

Chaima Jemmali, Northeastern University

Educational games have emerged as a solution to increase engagement and learning outcomes in different subject areas, such as programming education. In such a task, the importance of easing the learning process for individuals with no prior programming experience is crucial. Debugging is a cornerstone of programming and has been shown to be especially problematic for beginners.

Beyond Results: Design Considerations for Scientifically Valid Games that Empower Users After Play

Laura Levy, Georgia Institute of Technology

Research into the use of both commercial and custom video games to assess individual differences, like personality, of players has yielded promising results. Virtual environments allow researchers to collect a variety of player behaviors and actions that correlate strongly with

inherent traits. What is less understood is how an assessment game's mechanics might affect a player's inputs that determine the assessment's scientific validity. In an online study, we developed a custom game and logging framework to assess the reliability and validity of transferring a traditional personality questionnaire into a game environment. The game was played by 212 participants in one of three experimental conditions representing different levels of game mechanics. Using results from a traditional personality assessment as our ground truth, we compared player responses and play behavior in the game. We found that responses between the traditional assessment and game-based assessment in all conditions were consistent, indicating that the game mechanics did not interfere or alter significantly a player's ability or decision to make personality-based responses. Additionally, we found several gameplay behaviors that can be used as predictors of individual differences. This informed the design of a subsequent context-aware, interactive feedback game to help students engage in self-betterment behaviors.

Partnering to Investigate Big 5 and Game Play

Jenna McChesney, North Carolina State University

Models of personality, such as the Big 5 (Costa & McCrae, 1992), are important predictors of job performance and other organizational outcomes (e.g., job satisfaction, training performance, teamwork, leadership; Hertz & Donovan, 2000; Judge et al., 1999). Weidner and Short (2019) argue that game-based assessments (GBAs) may be an effective tool for mitigating some of the limitations (Spector, 2012) of traditional methods of measuring personality (e.g., self-report). Pymetrics, Inc. is an organization that combines neuroscience game-based assessments and advanced machine learning techniques to build predictive models for personnel selection. When applicants play pymetrics games, they are assessed on 50+ cognitive, social, and emotional characteristics, including planning, risk-taking, and willingness to trust others. Partnering with pymetrics, we investigated the correlation between the scores generated from pymetrics games and the Big 5 personality traits with the hopes of better understanding how game play relates to other important organizational frameworks and outcomes. During this presentation, findings of the current study will be discussed and the process and nature of the partnership with pymetrics will be shared. Our hope is to spark conversation and feedback around the current study and share how graduate students and GBA organizations can partner to conduct meaningful research.