



Transformative Trends in High Stakes Testing

David Foster | Caveon

John Weiner | Lifelong Learner Holdings

#CLEAR2023AEC



Meet the panelists



David Foster, CEO, Caveon



John Weiner, Chief Science
Officer, LLH



Today's discussion

What is a transformative trend or innovation that you think is important for high stakes credentialing testing?

- *How has high stakes testing changed as a result of this innovation?*
- *What is the impact or problems solved?*
- *What new issues are raised?*
- *What guidance is available regarding best practices in this area?*



6 Trends



Test design
innovations



Online/Remote
proctored testing



Game-based
assessment



Big data



Noncognitive
competencies



Generative AI /
AIG



1. Test Design/Development Innovations





1. Test Design/Development Innovations

- **With technology we can create more items**
 - Large numbers of items can be automatically created to support test security efforts.
- **Tests are able to move beyond traditional testing (i.e., fixed tests and standard multiple-choice items)**
 - Computerized adaptive tests and LOFT designs are supported.
 - Alternative item types are proliferating.



1. Test Design/Development Innovations



How has high-stakes testing changed as a result?

- **Better test security** – Less item exposure with automated item production and stingy test designs.
- **Integration of learning & assessment** – Enhanced items and tests can provide educators with more information about students in real time.
- **Better measurement** – Items and tests can more validly assess the targeted skills.





1. Test Design/Development Innovations

What new issues are raised?

- **How do we systematically update systems and procedures, and train personnel?**
- **Do new types of response data (i.e., response latency) need new analyses tools?**
- **Should our standards be revised?**





1. Test Design/Development Innovations

What guidance is available?

- ***ITC/ATP Guidelines for Technology-based Assessment***
(2022)
 - Ch 1: Test Development
 - Ch 2: Test Design and Assembly



2. Online/Remote Proctored Testing



2. Online/RPT



- **Internet-based testing (IBT) emerged in the 1990's**
 - UIBT emerged in low stakes employment testing in early 2000's
 - RPT emerged in high stakes testing in mid-2000's
 - RPT gained traction in credentialing in 2010's, **accelerated by Covid19**
- **Mixed mode testing is now common – RPT & Test center**
 - A variety of approaches to online proctoring and use of technology



2. Online/RPT



How has high stakes testing changed as a result?

- **Emergence of new proctoring systems, technologies, vendors**
- **Security methods are evolving to address RPT**
 - Planning, prevention, detection/forensic analysis
- **Impact often positive for test takers and organizations**
 - Supports geographically dispersed test takers, social distancing
 - Enables flexible scheduling, continuous testing
 - Eliminates travel time and costs for test takers, brick-and-mortar test center costs



2. Online/RPT



What new issues are raised?

- **Threats to security** – cheating, theft, content exposure (not new, but amplified)
- **Threats to validity** – CIV
 - Comparability across modalities
 - Technology disruptions, non-standard environments
 - Test taker experience, impact on performance, fairness
- **Technology access, fairness**
- **Legal risks & considerations** – privacy, content protection



2. Online/RPT



What guidance is available?

- **ITC/ATP *Guidelines for Technology-based Assessment*** (2022)
 - Ch 3: Test Delivery Environments: web-based delivery, offline, local and mobile delivery, locked-down browsers, interoperability, test disruptions
 - Ch 8: Test Security: risks, threats, strategies
- **ICE report: *Remote proctoring test delivery: A report on options and considerations*** (2015)
- **JATT special issue** (vol. 23, 2022): *Online remote proctored delivery of high stakes tests: issues and research*



3. Game-Based Assessment



3. Game-Based Assessment



- **For thousands of years, games have been able to--almost effortlessly--assess life-relevant skills.**
 - Most games require skills of planning, problem solving, cooperation, learning, communication, and leadership, to name a few.
 - Games are especially relevant when a game mirrors actual job tasks (solving crimes, gambling decisions, golfing, driving, etc.)
- **All existing high-stakes exams can incorporate one or more elements of game features to make the tests better, more fun, and interactive.**



3. Game-Based Assessment



How has high-stakes testing changed as a result?

- **Most high-stakes exams are still decidedly un-game-like, but research is being conducted and the acceptability of gamifying tests is growing.**
- **High-stakes tests have been built and used for more than 20 years that incorporate simulations and real-life scenarios with multimedia.** – These tests immerse test takers into approximations of real-life job situations, which is also common with games.



3. Game-Based Assessment



What new issues are raised?

- **Which elements of games can be incorporated into test designs and which are acceptable to test takers?**
- **Can games be adapted and used as high-stakes assessments as opposed to gamifying a test?**
- **Which elements of game play are more important than others in supporting high-stakes assessments?**



3. Game-Based Assessment



What guidance is available?

- **A GoogleScholar search for “gamification of assessment” will yield a large number of current research articles on the topic.**
- ***ITC/ATP Guidelines for Technology-based Assessment* (2022)**
 - Ch 1: Test Development, Guidelines 1.18 – 1.32



4. “Big Data”



4. Big data



- **Microprocessors were invented in the 1970's**, since then computing power has doubled every 18 months (Moore's law)
- **Better/cheaper technology** enables systems to house massive databases to support software and automation
 - 1 TB of storage = 85 million documents (10% of the library of congress)*
 - Cost = \$4.16 per month

*Infosecurity-magazine.com



4. Big data



How has high stakes testing changed as a result?

- **Content workflow automation** - item banking, test assembly, content management systems
- **Enhanced analytics** - integrated security, psychometric, and technology system data
- **Integration of learning & assessment** data for enhanced reporting, feedback, recommender systems
- **AI / ML applications** – provides the fuel for ML training algorithms for wide range of applications (more on AI later)



4. Big data



What new issues are raised?

- **Integrity** threats – data accuracy, completeness, technology failure, software bugs
- **Security & privacy** threats – access, breach, hacking, compliance
- **Management** – retention & removal policies
- **Validity** – 2-edged sword, many benefits, but more is not always better at the expense of meaningful measurement



4. Big data



What guidance is available?

- **ITC/ATP *Guidelines for Technology-based Assessment*** (2022)
 - *Ch 6: Data Management*: Governance, maintenance, integrity and security, integrating assessment data with learning systems
 - *Ch 9: Data Privacy*: protection of PI, compliance with privacy laws
- **ATP *privacy bulletins*** (2019-2022)



5. Behavioral Competencies



5. Behavioral Competencies



- **With the help of technology, high-stakes exams are incorporating ways to measure job skills directly.**
 - Many high-stakes testing programs (e.g., IT certifications, medical licensing, and many others) have included “performance” components to their assessments, measuring relevant job skills more directly.
- **Other assessment methods, such as remote on-the-job observation, AI review of work product, and automated scoring of essays, are alternative assessment methods which are also improved by technology.**



5. Behavioral Competencies



How has high-stakes testing changed as a result?

- **Technology has encouraged the design and development of “performance” assessments.** Actual job tasks are completed and scored.



5. Behavioral Competencies



What new issues are raised?

- **How should such assessments be scored?**
- **How can such technology-based tests be kept up to date when the job skills change often?**
- **Are such assessments comparable to traditional tests in terms of providing reliability, validity, security, and fairness evidence?**



5. Behavioral Competencies



What guidance is available?

- **There is a large body of literature describing such performance assessment.**
- ***ITC/ATP Guidelines for Technology-based Assessment* (2022)**
 - Ch 1: Test Development, Guidelines 1.18 – 1.32



6. Generative AI / AIG



6. Generative AI / AIG



- AI is not new, but has reached a tipping point with technology advances
 - **Computer algorithms** have been used for decades in science and business applications (e.g., actuaries)
 - **Machine learning** has been advancing to **leverage Big Data sets** to train and improve predictive models (e.g., chatbots)
 - **Generative AI** is a new level of capability that is disrupting work and learning models, able to perform human tasks





6. Generative AI / AIG

- Example AI capabilities

Interpretive	reading, listening, translating
Generative	writing, speaking, drawing, music and art composition
Computational	math, statistical analysis
Reasoning	predicting, decision-making, recommending
Sensory-motor	visual scanning, sorting, assembling, etc.



6. Generative AI / AIG



How has high stakes testing changed as a result?

- Test content generation – scenarios, complex questions, creative approaches, comprehensive
- Text and speech scoring
- Integrated learning and assessment – recommender systems



6. Generative AI / AIG



What new issues are raised?

- We can generate massive content, then what? Do we still need calibration to assure that test takers receive equivalent forms?
- Validity – Data scientists focus on prediction, not meaningful measurement
- Fairness – training data sets and algorithms may be biased
- Ethical use (human rights, explainable, accountable, secure & safe)
- Laws & regulations – NYC, EU, various states



6. Generative AI / AIG

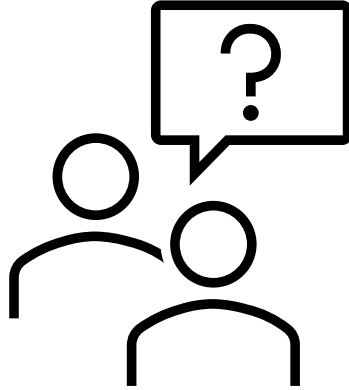


What guidance is available?

- *SIOP Considerations and Recommendations for AI-Based Assessment (2023)*
- *ITC / ATP Guidelines for TBA – evolving area*
- *Various Ethical AI frameworks (OECD, IEEE, NIST, etc.)*
- Wild west - no best practices published yet



Q&A Discussion



References

ITC/ATP Guidelines for Technology-Based Assessment (2022) <https://atpu.memberclicks.net/atp-white-papers>

Weiner, J. (2022). *The Increasing Importance of Noncognitive Competencies: Opportunities for Credentialing Assessment*. *Credentialing Insights – Online Journal for I.C.E.*
<https://www.credentialinginsights.org/Article/the-increasing-importance-of-noncognitive-competencies-opportunities-for-credentialing-assessment-1>





Speaker Contact Information

John Weiner

John.weiner@lifelonglearner.com

David Foster

David.foster@caveon.com

#CLEAR2023AEC





Thank You

#CLEAR2023AEC

