THE CONFERENCE ON TEST SECURITY

2024

SALT LAKE CITY, UTAH

13th Annual

A SPECIAL THANK YOU

TO OUR CO-HOSTS:

duolingo english test

caveon®

Honorlock®

MEASURE LEARNING

Pearson VUE

WELCOME

TO COTS 2024

Dear Colleagues,

Welcome to the 13th annual Conference on Test Security – the only event dedicated entirely to test security. It is my pleasure to have you with us in Salt Lake City, Utah.

When COTS was first started in 2012, it was called the Conference on the Statistical Detection of Potential Test Fraud, and it focused primarily on statistics. In 2014, the scope of the conference broadened to encourage dialogue on all test security capabilities and enhancements. Now in its thirteenth year, COTS has rightfully gained a reputation for facilitating an atmosphere that encourages open and honest test security discussions amongst professionals across the testing industry.

While the ever-constant threats to the security of our tests continue, it is your zeal for wisdom and innovation that keeps those dangers at bay. It is my sincerest hope that this year's conference will be of the utmost benefit to you as we learn and network together. With over 50 sessions and attendees from across the world, we hope you'll find information and resources that are helpful for you and for your program's needs.

If you're new to COTS, we hope that you'll join us for the Cocktails and Conversations event in the Uintah Room (2nd floor) on the evening of Thursday, October 17th. Here, you'll find the opportunity to converse one-on-one with some of the brightest minds in testing and participate in our poster presentations. Dubbed as "the favorite event of COTS," you won't want to miss meeting with leaders in the field on this level.

Finally, a special thank you to all our sponsors for making this year's conference possible. We appreciate their unyielding support. It is because of them that we can keep our registration costs low and successfully gather together every year, and forward our joint goal to protect the validity of test results and exam integrity.

We hope that you enjoy your time in Salt Lake City, Utah at the 2024 Conference on Test Security.

David Foster
CEO & President, Caveon Test Security
2024 Host of the Conference on Test Security

AND THANK YOU

TO OUR FRIENDS:

kryterion by DRAKE INTERNATIONAL

NCTA NATIONAL COLLEGE TESTING ASSOCIATION

ncme NATIONAL COUNCIL ON MEASUREMENT IN EDUCATION

PROMETRIC

psi

TABLE OF CONTENTS

3 WELCOME TO COTS

2-4 THANK YOU TO OUR SPONSORS

6 CONFERENCE AGENDA

7 WEDNESDAY PROGRAM

8 THURSDAY PROGRAM

26 FRIDAY PROGRAM

33 MAPS, NOTES, & MORE

CONFERENCE AGENDA

COTS 2024

WEDNESDAY, OCTOBER 16

3 - 7 PM	**Registration & Information** *Sinclair Office, 1st Floor*
7 - 8:30 PM	**Opening Reception** *Uintah, 2nd Floor*

THURSDAY, OCTOBER 17

7 AM - 5 PM	**Registration & Info** *Sinclair Office, 1st Floor*
7 - 8:10 AM	**Pre-Keynote Tailgating Workshop** *Snowbasin, 2nd Floor*
7:45- 8:30 AM	**Breakfast** *Ballrooms A & B, 1st Floor*
8:45 - 9:45 AM	**Opening Keynote** *Ballrooms A & B, 1st Floor*
10 - 11 AM	**Sessions** *Olympus, Teton, Snowbasin, Flagstaff, 2nd Floor*
11:15 AM - 12:15 PM	**Sessions** *Olympus, Teton, Snowbasin, Flagstaff, 2nd Floor*
12:15 - 1:45 PM	**Lunch & Plenary Session** *Ballrooms A & B, 1st Floor*
2 - 3 PM	**Sessions** *Olympus, Teton, Snowbasin, Flagstaff, 2nd Floor*
3:15 - 4:15 PM	**Sessions** *Olympus, Teton, Snowbasin, Flagstaff, 2nd Floor*
4:30 - 5:30 PM	**Sessions** *Olympus, Teton, Snowbasin, Flagstaff, 2nd Floor*
5:45 - 6:45 PM	**Poster Presentations** *Ballroom Foyer, 1st Floor*

FRIDAY, OCTOBER 18

7:30 AM - 12 PM	**Registration & Info** *Sinclair Office, 1st Floor*
7:30 - 8:15 AM	**Breakfast** *Ballrooms A & B, 1st Floor*
8:15 - 9:15 AM	**Sessions** *Olympus, Teton, Snowbasin, Flagstaff, 2nd Floor*
9:30 - 10:30 AM	**Sessions** *Olympus, Teton, Snowbasin, Flagstaff, 2nd Floor*
10:45 AM - 12 PM	***Closing Keynote & Debates*** *Ballrooms A & B, 1st Floor*

WEDNESDAY, OCT 16

3 - 7 PM — *SINCLAIR OFFICE*

REGISTRATION & INFORMATION

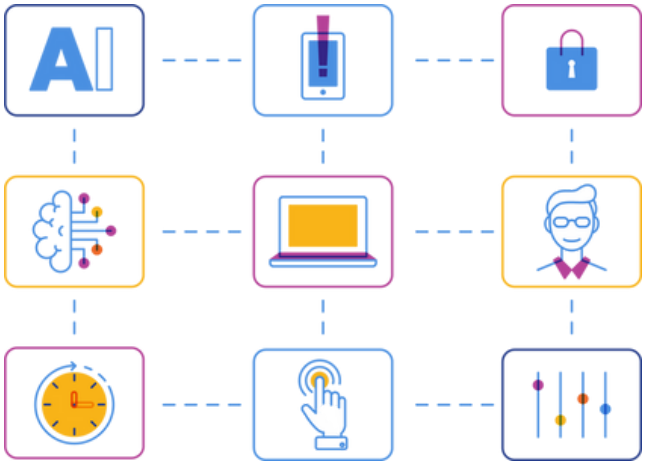
7 - 8:30 PM — *UINTAH ROOM*

OPENING RECEPTION

Honorlock

A Better Way to Secure Exams

Honorlock combines the power of live proctoring with AI to protect exam integrity and enhance the testing experience. From detecting cell phones and speech to preventing the use of ChatGPT, Honorlock's remote proctoring secures exams from all angles



Attend Honorlock's Session:
Harness AI: Strategies for Maintaining Test Integrity and Validity

When
Thursday, 10/17 at 11:00 a.m. MDT
Where
Uintah Room (2nd floor)

Visit us at Honorlock.com

THURSDAY, OCT 17

7 AM - 5 PM *SINCLAIR OFFICE*

REGISTRATION & INFORMATION

7 - 8:10 AM *SNOWBASIN ROOM*

PRE-KEYNOTE KICKOFF TAILGATING WORKSHOP

The What and the How of Technology-Enabled Cheating

Rachel Schoenig, Cornerstone Strategies | Craig Nicholson, Cambridge University Press and Assessments | Cory Clark, Meazure Learning | Rose Hastings, Duolingo English Test

Rise and shine with this pre-keynote "tailgating" event! There won't be hot dogs or beer, but we will have some interesting new tech-enabled cheating tools to share. Attendees will have the opportunity to get hands-on with some of the newest tools available on the market, including some that build upon generative AI. During this session, we will discuss and demonstrate common methods for cheating with technology, from answers created by generative AI to virtual proxies and deep fakes. With new and open-source capabilities advancing at a rapid pace, the quality and availability of these tools continues to grow. Tailgaters will learn how tech-enabled tools can be built and used relatively affordably and easily, how they are being distributed and accessed, and resources and methos for deterring and detecting them in your testing program. This early morning tailgating session is the perfect start to your COTS conference sessions and prepare you for the Keynote Kickoff!

7:45-8:30 AM *BALLROOM A*

BREAKFAST

THURSDAY, OCT 17

OPENING KEYNOTE

8:45 – 9:45 AM | BALLROOM B

BREAKING DOWN TECH-ENABLED TEST FRAUD

Craig Nicholson, Cambridge University Press and Assessments

Kickoff the 2024 conference with one of the most inquisitive minds in the industry! Craig Nicholson is Chief Information Security Officer and Director of Technical Innovation at Cambridge University Press and Assessments. His interest in technology started decades ago, when he built his first computer at age three. Today, in addition to his roles at multiple international high-stakes testing programs, Craig works with governments and tech companies around the world to ethically improve technology and information security. From breaking down today's technology tools to running his own experimental AI and hardware threat lab, Craig is on the cutting edge of what's possible with today's technology. During this moderated keynote kickoff, Craig will share the technologies that energize him, the technologies that concern him, and his vision of a future where technology plays an even bigger role in education and assessment!

A SPECIAL THANK YOU TO OUR CO-HOSTS:

Pearson VUE

caveon

MEASURE LEARNING

Honorlock

duolingo english test

THURSDAY, OCT 17

10 - 11 AM

OLYMPUS ROOM

HOW TO THRIVE IN YOUR ROOKIE SEASON

Panel Presentation

Rachel Schoenig, Cornerstone Strategies | Camille Thompson, College Board | Ray Nicosia, ETS | Faisel Alam, LSAC

If you're new to testing and exam security, or are just looking for a helpful refresher, this is the session for you!We get it, rookies. We were there once too, trying to understand how psychometricians work, figuring out what a p-value is, and wrapping our arms around item harvesting. The testing industry isn't easy for rookies. It requires an entirely new lingo and a new understanding of the broader ecosystem that the game is played in today.That's why we're here to help! Join some exam security pro's who can break down the basics of testing and exam security. We'll share what we wish we had known when we started and bring you up to speed on testing terms and foundational exam security concepts. We'll take you from drafted to "made the roster" before our time together is through. Join us for a fun and engaging session that will set you up for the all-pro team when you get back to the office.

TETON ROOM

USE OF GENERATIVE AI IN SECURE ITEM DEVELOPMENT FOR STATE ASSESSMENT PROGRAMS: RECOMMENDATIONS FOR FUTURE IMPLEMENTATION BY STATES

Panel Presentation

Walt Drane, Caveon | Sarah Quesen, WestEd & University of Pittsburgh | Sergio Araneda, Caveon | John Olson, OEMAS

The testing industry is on the brink of significant change, with Generative AI now able to transform key functions like item and test development for state assessment programs. By sharing successful examples of overcoming challenges with AI, states can adopt best practices, enhance effectiveness, and reduce costs, while also improving test security.

Refreshing item pools, traditionally costly and time-consuming, has been revolutionized by AI in the past year. Recent successful applications in high-stakes tests can provide valuable insights for state programs. In this session, leaders from state assessment and university sectors will discuss how Generative AI benefits item development and share strategies for evaluating and implementing it. Two national experts will also offer recommendations on effectively introducing AI in state assessments.

THURSDAY, OCT 17

10 - 11 AM *CONTINUED...* *SNOWBASIN ROOM*

TEST SECURITY IN LONGITUDINAL ASSESSMENT: CONSIDERATIONS, TECHNIQUES, AND INITIAL FINDINGS

Coordinated Symposium

Derek Sauder, American Board of Internal Medicine (ABIM) | Danielle Lee, ABIM | Jordan Yee Prendez, ABIM | Jim Wollack, University of Wisconsin-Madison

In both K-12 and certification assessments, there's a growing focus on programs that measure progress towards learning objectives and maintain knowledge over time. This approach, known as through-course summative assessment in K-12 and longitudinal assessment in certification, involves periodic tests leading to a comprehensive evaluation. As assessments move online, content security concerns have become crucial.

Our session will explore the challenges of longitudinal assessments, using a medical certification assessment as a case study. We will discuss maintaining test integrity and present two papers: one on detecting aberrant examinee responses using Guttman-based statistics and response time metrics, and another on identifying compromised items with an IRT-based sequential procedure. Findings from both real and simulated data will be shared. Finally, a test security expert will offer commentary and lead a discussion among participants.

FLAGSTAFF ROOM

LET'S TALK IT OUT! CONDUCTING CANDIDATE INTERVIEWS IN EXAM SECURITY INVESTIGATIONS

Panel Presentation

Harry Samit, Pearson VUE | Bryan Friess, Pearson VUE | Brent Morris, Cisco

Traditional exam security techniques do an effective job of alerting us to threats to the integrity of an exam program. Unfortunately, these measures do not always tell us what happened or who was behind the nefarious activity. It is a truism in law enforcement that to understand a criminal plot, investigators must obtain statements from those inside the conspiracy. This also applies to exam security investigations, where one 10-minute interview may be worth more than 10 hours of technical investigation. Drawing on real-world examples, this presentation will offer strategies designed to maximize the effectiveness of subject interviews in exam security investigations.

11 - 11:15 AM *BREAK*

THURSDAY, OCT 17

STANDARD PRESENTATIONS: "NEW APPROACHES IN EXAM SECURITY"

11:15 - 12:15 PM
SNOWBASIN ROOM

Psychometric Practice on Item Exposure and Testing Security in Certification Assessments

Yu Meng & Amin Sair, Board of Pharmacy Specialties

The exposure of test items poses a significant threat to the validity of certification exams, particularly in the medical and healthcare fields. Ensuring exam security, which involves detecting and preventing unauthorized exposure of test content, is vital to the integrity of the certification process, as well as safeguarding the value and reputation of the certification itself. This presentation will introduce test security at different stages and share data forensic practice and psychometric insight regarding integrity of test scores. We will discuss various data security measures, including item performance analyses, response pattern analyses, and web monitoring. Additionally, we will introduce different methods to detect test-takers who exhibit abnormal or suspicious response patterns in practice. Finally, we will share statistical models for investigating test centers where there are unusually large increases in scores on credentialing examinations. The goal of this presentation is to share operational experience about item exposure specific to certification assessments.

Innovative Strategies and Actions to Bolster Exam Integrity

Chad Fletcher, Kryterion

This presentation will discuss new methods to enhance the security of your certification program. The session will delve into proactive measures to detect proxy testers and explore innovative strategies for bolstering testing program integrity. Attendees will gain insights into cutting-edge technologies and methodologies that empower organizations to mitigate the risks associated with fraudulent test-taking practices. By leveraging advanced detection code, artificial intelligence-assisted authentication, and process improvements, certification programs can safeguard their assessments against unauthorized access and ensure the validity of test results. The presentation aims to equip testing professionals with actionable strategies to enhance test security, uphold the integrity of certification programs, and protect the value of credentials in today's evolving landscape.

Harness AI: Strategies for Maintaining Test Integrity and Validity

Paul Morales, Honorlock

Higher education has faced disruptions throughout history, but educators have always risen to the challenge and ensured student and institutional success. With the advent of AI, the educational landscape is undergoing a significant transformation, and educators are once again called upon to navigate this uncharted territory with confidence and determination. AI tools like ChatGPT and Transcript aren't just a trend. AI tools are transforming traditional approaches, not as mere trends but as essential instruments for innovation in assessment design, but new challenges exist. How can educators leverage AI tools to help create engaging and unique exam content without sacrificing test integrity and validity?

11:15 - 12:15 PM *CONTINUED...* *FLAGSTAFF ROOM*

USING CHATGPT FOR THE AUTOMATIC CREATION OF SMARTITEMS™ FOR THE PURPOSES OF TEST SECURITY AND PERSONALIZED ASSESSMENT

Demonstration

Sergio Araneda, Caveon

In this demonstration, we will show how to use ChatGPT to create SmartItems™ using various strategies. The new capabilities of Large Language Models (LLMs) provide innovative methods for generating SmartItems™ through the automatic creation of JavaScript code compatible with Scorpion, facilitating the use of randomly parallel tests as a strategy for enhancing test security and expediting the item development process. Additionally, we will demonstrate how ChatGPT can be utilized to create thematic variations of items, enabling personalized assessment within this automated framework. This approach not only enhances the efficiency of item development but also expands item pools to protect against different types of cheating and theft, while promoting inclusivity and fairness in assessments. Through this demonstration, attendees will gain practical insights into the mechanics of using LLMs like ChatGPT to write functional SmartItems™ code compatible with Scorpion, and how these technologies can be synergistically combined to create secure, culturally responsive, and personalized assessments.

OLYMPUS ROOM

FRIEND OR FOE? GENERATIVE AI IN ASSESSMENT: CONTINUING THE CONVERSATION

Panel Presentation

Isabelle Gonthier, PSI Services | David Yunger, vAltal | Marc Weinstein, Marc J. Weinstein PLLC | John Dight, Surpass

Artificial Intelligence (AI) is a subject of intense debate in the testing and assessment industry, evoking reactions ranging from enthusiasm for its potential to enhance productivity to concerns about security, data management, intellectual property, and malpractice. This landscape raises the question: Can we balance the opportunities and threats AI presents?

This expert panel will explore the challenges and prospects of deploying AI tools, particularly generative AI, in test development and delivery, as well as the associated security implications. We aim to foster critical reflection, debate merits and drawbacks, and stimulate dialogue on security issues tied to increasing AI usage, encouraging attendees to reassess their viewpoints and engage in thoughtful exchange.

THURSDAY, OCT 17

11:15 - 12:15 PM *CONTINUED...* *TETON ROOM*

IMPLEMENTATION OF NEW THROUGH-YEAR AND INTERIM/BENCHMARK ASSESSMENTS BY STATES: LATEST LESSONS LEARNED FOR MORE SECURE TESTING POLICIES AND PROCEDURES

Panel Presentation

John Olson, OEMAS | Jenny Black, Florida DOE | Dusty Shockley, Delaware DOE | Walt Drane, Caveon | Shannon Jordan, North Carolina DOE

An important development in state assessments is the adoption of new Through-Year (TYA) and Interim/Benchmark (I/BM) Assessments. By 2024, 15 states will have implemented innovative TYAs that may include I/BM or other assessments alongside summative tests. TYAs require a system of high-quality assessments serving multiple purposes for diverse audiences, with some states using these designs to provide more timely results and reduce emphasis on end-of-year summative tests. Many of these initiatives are funded by USED Competitive Grants for State Assessments (CGSA) or Innovative Assessment Development Authority (IADA) programs. However, implementing these assessments with fidelity is crucial for valid and reliable outcomes, highlighting the need for comprehensive test security policies.

In this session, three state presenters will discuss how they are implementing TYAs and I/BM assessments and integrating comprehensive security into their policies. Additionally, two national test security experts will provide insights on state efforts to improve security across TYA components and offer recommendations for enhancing security procedures.

12:15- 1:45 PM *BALLROOMS A & B*

12:15 - LUNCH SERVICE BEGINS
12:50 - LUNCHTIME PLENARY SESSION BEGINS

LUNCH PLENARY

THURS, OCT 17 | 12:30 – 1:45 PM | BALLROOMS A & B

PANEL OF EDUCATION LEADERS DISCUSSING CURRENT SECURITY ISSUES AND LATEST SOLUTIONS FOR STATE ASSESSMENT PROGRAMS

Darin Nielsen, Utah State Board of Education|Cydnee Carter, Utah State Board of Education|Maria D'Brot, FocalPoint Education| Juan D'Brot, Center for Assessment| Don Peasley, U.S. Department of Education | Walt Drane, Caveon | John Olson, OEMAS

A distinguished panel of state policymakers, assessment experts, and testing program directors from state and federal agencies, large consortia, and CCSSO will share insights into their test security initiatives for state assessments. Each panelist will give a concise overview of their activities, followed by a Q&A session moderated by two assessment experts. Key topics include policy issues, best practices, security for consortia sharing test item banks, impacts of USED peer reviews, challenges with remote testing, and enhanced security for through-year assessments.

THURSDAY, OCT 17

2 - 3 PM

OLYMPUS ROOM

TEST SECURITY MYTHS DEBUNKED: SEPARATING FACT FROM FICTION

Panel Presentation

Cicek Svensson, Caveon | Angela Bostwick, MonitorEDU | Roger Meade, Prometric
Alana Chamoun, CSBS

Opinions and assumptions about test security abound, and most of them are either out-of-date or outright incorrect. In this session we will demystify prevailing beliefs and set the record straight on key aspects of test security, specifically as it applies in the educational and customer training contexts. Additionally, we will discuss the biggest disruptor in our industry (AI) and how it influences test security in exam development and exam delivery. Led by seasoned test security and customer education professionals, this interactive session unravels misconceptions, provides evidence-backed insights, and engages in a lively discussion to challenge the status quo. Attendees will have the opportunity to participate, putting their own knowledge to the test and gaining valuable insights into what test security tactics really work.

TETON ROOM

DEALING WITH AI CHEATING THREATS AND SOLUTIONS

Demonstration

David Foster, Caveon | Nathaniel Foster, Caveon

This year, AI-based products have emerged to help people cheat on exams in real time, often marketed as "homework helpers" available as browser extensions. These tools allow AI, like ChatGPT, to either take a test or provide answers to test questions, and have been shown to perform well on high-stakes exams in various professional fields. In this session, we will demonstrate these products and present unique security solutions that effectively prevent this type of cheating. We'll cover technical hardware and software solutions that hinder AI-assisted cheating and discuss test design modifications to detect such use. The session will conclude with research on the effectiveness of these solutions and a discussion on related AI threats.

STANDARD PRESENTATIONS: "CONSIDERING PROCTORING AND EXAM SECURITY"

2-3 PM
SNOWBASIN ROOM

Measuring Accuracy and Error in Remote Proctor Decision Making

Will Belzak, Duolingo

High-quality decision making by remote proctors is critical for maintaining effective test security in online tests. No empirical research has investigated the degree to which remote proctors make accurate or erroneous decisions about suspicious test-taking behaviors (e.g., looking away excessively). To address this gap, we present a simple framework for conceptualizing accuracy and error in remote proctor decision making and propose three methods to empirically estimate these quantities. The framework relies on signal detection theory and statistical hypothesis testing, such that when applied to remote proctor decision making, we can consider the probability that a proctor decides a test taker is cheating when they are not (false positive / Type I error) and the probability that a proctor decides a test taker is not cheating when they are (false negative / Type II error), in addition to their true positive and negative complements. The three methods we propose to estimate these quantities leverage an internal system, the proctor calibration tool, developed to automatically measure reliability of proctor decision making by administering the same recorded test sessions to many proctors. Notably, each method differs in how a "correct" decision is defined. We illustrate the framework and three methods using proctoring data from a high-stakes remote assessment of English language proficiency. We argue that the integrity of remote assessments can be improved by conceptualizing and estimating accuracy and error in remote proctor decision making using these principled statistical and psychometric approaches.

State Requirements for Test Administrators, Proctors, and Accommodations Providers: Implications for Test Security

Mari Quanbeck & Sheryl Lazarus, National Center on Educational Outcomes

Test security is a concern for any statewide summative assessment, but this concern may increase when testing students from some subgroups. Many students use accessibility features and accommodations to access state summative assessments that are used for accountability purposes. Test administrators, proctors, and accommodations providers must know which accessibility features and accommodations students are using, how to address issues related to these supports on test day, and how to maintain test security while ensuring that students have access to the accommodations identified on their individualized education program (IEP). The individuals who administer or proctor assessments, as well as those who provide accommodations, need to be qualified and knowledgeable about how to appropriately administer assessments and provide accommodations. We will present the findings of an analysis of state policies related to the requirements and qualifications for test administrators, proctors, and accommodations providers on state summative assessments used for accountability purposes. This session will conclude with recommendations for ensuring that individuals involved in the administration of accommodated tests are appropriately qualified and able to administer assessments in ways that do not invalidate test results.

THURSDAY, OCT 17

2 - 3 PM *CONTINUED...* *FLAGSTAFF ROOM*

ENHANCING TEST SECURITY: REAL-TIME DECISION MAKING AND AUTOMATIC INTERVENTIONS IN HIGH-STAKES ASSESSMENTS

Panel Presentation

Isabelle Gonthier, PSI Services | Nicole Tucker, PSI Services | Ray Nicosia, ETS

This panel addresses the evolving challenges of test security in the era of AI. Experts from PSI Services and ETS will discuss advanced technologies and strategies to enhance real-time security in high-stakes assessments. The session will cover integrating Data Forensics with advanced tech for near real-time detection of irregularities, optimizing response times to prevent breaches, implementing real-time proctoring, and using AI to improve security while addressing new vulnerabilities. The discussion will also emphasize the importance of human oversight in automated processes and maintaining security without compromising the test-taker experience. Attendees will gain actionable insights for improving assessment reliability and fairness.

THURSDAY, OCT 17

3 - 3:15 PM *BREAK*

3:15 - 4:15 PM *OLYMPUS ROOM*

HOW AI IS IMPACTING EXAM SECURITY PROGRAMS

Facilitated Roundtable

Kim Brunnert, Elsevier | Jim Hussey, ACT | Claire McCauley, Cambridge English | Rachel Schoenig, Cornerstone Strategies, LLC

Setting aside the promises of AI, what are the realities today? Are you curious as to how AI is being incorporated by educational institutions and assessment programs today? Interested in understanding the benefits it provides, and the methods organizations are using to incorporate AI without jeopardizing personal information or secure testing assets? Want to explore the ways you can incorporate AI for your team or individual benefit? Join this Roundtable discussion to learn more about the efficiencies and benefits of AI for educational and assessment programs. Together, we'll share practical uses and approaches and learn from one another how AI is impacting exams and exam security.

TETON ROOM

USING TEST SECURITY DATA MORE EFFECTIVELY TO IMPROVE THE MONITORING OF SCHOOLS DURING STATEWIDE ASSESSMENTS

Panel Presentation

John Olson, OEMAS | Maggie Hicks, Alabama DOE | David Ragsdale, Massachusetts Department of Elementary and Secondary Education | Jessica Fenby, Michigan DOE | Walt Drane, Caveon

The U.S. Department of Education requires states to monitor assessments for accountability reporting and submit evidence of compliance with test security standards. States must observe test administration to ensure it's conducted securely and fairly. However, implementing a quality monitoring program and passing peer review has been challenging for some states. Some states have found resourceful ways to conduct monitoring, including on-site, desk audits, and remote monitoring. They effectively use data to identify schools needing additional oversight, such as using forensic analyses and web monitoring to flag irregularities. In this session, a panel of state security directors and national experts will discuss best practices for monitoring statewide assessments and using data to ensure secure, valid testing.

STANDARD PRESENTATIONS: "APPROACHES TO DETECTING EXAMINEE PRE-KNOWLEDGE"

3:15-4:15 PM
SNOWBASIN ROOM

On the Detection of Multiple Groups of Examinees with Preknowledge

Daihui (David) Xiao, Michigan State University | Kylie Gorney, Michigan State University

With the ongoing threat of fraudulent behaviors in testing environments, maintaining test integrity is crucial. In this study, we address the challenge of detecting examinees with preknowledge in high-stakes examinations. Although previous researchers have compared different preknowledge detection statistics on their Type I error rates and power, most have only considered the case where a single group of examinees has preknowledge of a common set of items. In practice, however, it is possible that multiple groups of examinees will have preknowledge of different sets of items. In this study , we compare the effectiveness of several multiple comparison procedures in their ability to detect examinees with preknowledge of different sets of items. Simulations are conducted in which the following factors are manipulated: the number of items, the percentage of compromised items, the nominal significance level, the number of sets of compromised items, and whether or not there is overlap between the compromised item sets. Results show that the multiple comparison procedures display reasonable power and control the Type I error rates, thus helping to reinforce the fairness and integrity of the testing process.

Detecting Examinee Pre-knowledge using Network Clusters

Rich Feinberg & Tim Helbig, National Board of Medical Examiners (NBME)

Analyzing response similarity as a measure of preknowledge detection has been demonstrated to be an effective statistical approach in building a case to invalidate an examinee's score. However, more recent innovations in machine learning techniques offer potential additional utility in helping identify coordinated networks or clusters of examinees who may be colluding together. In the present study, we plan to investigate the effectiveness of a network clustering approach to detect collusion in a realistic scenario among random examinees whose only commonality in the data is the same compromised items. Operational data from a high-stakes examination containing 2,130 examinees and 100 items will be manipulated to create conditions of shared preknowledge among a random subset of examinees. Manipulated conditions will include the number of examinees with preknowledge, the number of compromised items, and the intensity of the compromise. Findings will be informative to practitioners to the extent that similar methods can be useful as a monitoring tool. Discussion will include implementation considerations for proactively seeking anomalous clusters, policy implications, and additional analysis steps to further explore detected clusters.

(Standard Presentations Continued on Next Page)

Data Mining for Pre-knowledge with Unproctored Medical Assessments

Stuart Barnum, National Board of Osteopathic Medical Examiners (NBOME)

We develop methods and present results for forensic analysis of a series of assessments offered in an unproctored delivery mode beginning in March 2020 with the start of the COVID-19 pandemic. Our methods are based on response times for correct answers to dichotomously-scored items. Examinees with preknowledge of an item, answering correctly, may be expected to respond more quickly than similar examinees without preknowledge. We develop a pre-2020 baseline response time for each item and consider differences from the baseline for each response to each item after 2020. With the differences standardized based on the pre-2020 mean and standard deviation for each item, we develop clustering methods including Bernoulli mixture models, k-means, and factor analysis. A threshold identifying the standardized difference (decrease) in response time as useful in the model is treated as a hyperparameter. Responses less than the threshold for correct answers are coded as 1, while responses not passing the threshold, or with incorrect answers, are coded as 0. The various mixture models are thus applied to matrices of binary data, with a principled distinction between response-time differences deemed useful in the model and those not deemed useful. Clusters with disproportionate numbers of 1s present evidence of preknowledge. The methods are tested with simulations and then applied to real examination data. Results related to the clusters are then correlated with measures such as fit to the scoring model and jumps in scores from previous, similar, assessments by the same examinees.

3:15 - 4:15 PM *CONTINUED...* *FLAGSTAFF ROOM*

SIMILARITY STATISTICS ON CATS?! UNCONVENTIONAL APPLICATIONS AND INSIGHTS GAINED

Coordinated Symposium

Sarah Toton, Caveon | Marcus Scott, Caveon | Thalia Huynh, Caveon | Donna Butterbaugh, ISC2

Response similarity statistics are powerful data forensics statistics for detecting potential test security breaches. They were designed for use on fixed-form tests, and their power depends on the amount of overlap in the items delivered to examinees. Variable test designs such as CAT and LOFT have less overlap in the items that are administered, by design. Typically, similarity statistics would not be used to analyze variable test designs because the decreased item overlap decreases the statistical power to detect extreme similarity, resulting in the loss of one of the field's most powerful detection tools. In this session, we will discuss discuss real situations where similarity statistics were used to analyze data from CAT tests, including a case study from the largest cybersecurity professional certification program, ISC2. Next, we will discuss research predicting similarity outcomes from other data elements and variables, with the goal of approximating these outcomes when they cannot be computed. Finally, we will discuss implications of using similarity statistics in variable test designs and special considerations for these unconventional applications.

THURSDAY, OCT 17

4:15 - 4:30 PM *CONTINUED...* *BREAK*

4:30 - 5:30 PM *FLAGSTAFF ROOM*

HOW DELIVERY MODALITY INFLUENCES CANDIDATE BEHAVIOR AND PERFORMANCE

Panel Presentation

Cicek Svensson, Caveon | Angela Bostwick, MonitorEDU | Paul Muir, risr

This session will explore the evolving landscape of test delivery methods, including test center-based, live online proctoring, automated online proctoring, and alternative locations. With advancements in technology, it is crucial to understand how these methods impact security, convenience, cost, and the candidate experience.

Three experts in test delivery and credentialing will lead a panel discussion on the practical considerations for selecting and implementing different test delivery approaches. Topics will include the pros and cons of traditional test centers, the rise of at-home testing, and the benefits and challenges of other proctoring methods.

The discussion will address key concerns such as test anxiety, distractions, technical issues, and privacy. These factors can influence candidate outcomes and raise questions about the validity and fairness of assessments. The session will also cover the operational impacts for test sponsors, including planning, documentation, practice tests, and score reporting.

Participants will gain insights into maintaining the integrity and credibility of credentialing programs across diverse contexts. The session will present a framework for evaluating different test delivery methods, followed by an interactive discussion where attendees can ask questions and share their experiences.

THURSDAY, OCT 17

4:30 - 5:30 PM

OLYMPUS ROOM

HOW TO STAY ONE (OR MORE!) STEPS AHEAD OF ONLINE TEST CHEATING RINGS

Panel Presentation

Rachel Schoenig, Cornerstone Strategies | Harry Samit, Pearson VUE | Kim Snyder, Duolingo English Test | Rose Hastings, Duolingo English Test

Cheating rings present significant challenges in online testing environments. These rings often involve coordinated efforts among multiple individuals to share answers, use external resources, or employ sophisticated methods like screen sharing during exams. Detecting such collusion is complex as it requires identifying patterns of unusually high similarity among submissions or suspicious behavior during the test. Additionally, ensuring the integrity of online assessments demands robust authentication measures to verify the identity of test takers and prevent impersonation. Continuous adaptation is necessary to stay ahead of evolving cheating tactics, which may involve the use of advanced technology such as AI-generated responses or plagiarism detection tools. Overall, combating cheating rings in online tests requires a multifaceted approach involving technological solutions, stringent monitoring protocols, and educational interventions on academic integrity. In this panel, we discuss these challenges and how we are working to identify and mitigate them.

TETON ROOM

REMOTE TEST ADMINISTRATION FOR STATE ASSESSMENTS: SUCCESSFUL APPROACHES BEING CONDUCTED IN K-12 PROGRAMS TO MAINTAIN TEST SECURITY AND INTEGRITY

Panel Presentation

John Olson, OEMAS | Karen Tohinaka, Hawaii DOE | Cydnee Carter, Utah State Board of Education | Timothy Butcher, WV DOE | Walt Drane, Caveon

Remote Test Administration (RTA) allows K-12 students to complete state assessments—formative, interim/benchmark, summative—at home rather than in school. Due to the COVID pandemic, many states transitioned to remote testing in 2021, a trend that continued into 2022 and 2023 with various adjustments. By 2024, nine states have adopted RTAs, some for specific purposes like virtual charter schools. This shift requires states to creatively leverage technology to combat test fraud, using multiple methods for security.

In this session, three state assessment staff will share their experiences conducting RTAs, discussing the procedures implemented to maintain security. Each state approached RTA differently, and the presenters will explain their comprehensive and integrated security strategies to preserve assessment integrity.

STANDARD PRESENTATIONS: "EVOLVING METHODS IN STATISTICAL DETECTION" (PART 1)

4:30-5:30 PM
SNOWBASIN ROOM

Item compromise and examinee pre-knowledge detection using scores, distractors, and response times

Merve Sarac, College Board

Statistical evidence derived from response times, which are available in computer-based assessments, remains underutilized in identifying compromised test content. This study utilized item fit analysis based on response times, in addition to item scores and distractor selection, to detect potentially compromised items. By leveraging information from items that misfit, a detection statistic incorporating item scores and response times was employed to identify examinee preknowledge. In simulations that considered various design factors, both the false positive rate and the true positive rate were assessed for items and examinees. The incorporation of response time into item fit analysis, alongside item scores and distractor selection, enhanced the true positive rates for both items and examinees.

Enhancing test cheating detection: A Population-Based Modification of the K2 statistic

Irina Grabovsky, NBME | Chunyan Liu, NBME | Carol Morrison, NBME

Test security is essential for maintaining the validity of test scores, especially for testing organizations administering credentialing examinations. It is paramount for test practitioners to be able to apply effective methods to detect suspicious response patterns. Test cheating resulting from collusion can potentially jeopardize the entire item bank.
The K2 statistic is often used to identify test-takers with potentially fraudulent behaviors. This statistic is based on the approximation of the distribution of the number of matched incorrect answers, M, between copier and source, through the binomial distribution B(w_s,p_2^*), where w_s is the number of wrong answers of the source, and p_2^* can be estimated through a quadratic regression approach. One concern with applying this statistic is that combining the "baseline" group with the "suspect" group may negatively impact the accurate identification of potential cheaters.
This study proposes to modify the K2 calculation so that p^_2^* matrix and the empirical distribution of the number of matched incorrect answers are developed based on the baseline group alone (N=500). For each cheater-copier pair in the suspect group, the probability of getting the number of matched incorrect answers (m_cs) is determined based on the empirical distribution from the baseline group. We expect this modification will improve the power of statistics to detect systematic cheating.

This simulation study will investigate several factors: the suspect group sample size (N=100, 250), the test length (L=50, 100), the percentage of cheaters (0%, 10%, 30%, 50%), and the percentage of breached items (10%, 30%, 50%, 100%).

THURSDAY, OCT 17

5:30 - 5:45 PM — *BREAK*

5:45 - 6:45 PM — *BALLROOM FOYER*

POSTER SESSIONS
CONVERSATIONS AND COCKTAILS

FRIDAY, OCT 18

7 AM - 5 PM — *SINCLAIR OFFICE*

REGISTRATION & INFORMATION

7:30-8:15 AM — *BALLROOMS A & B*

BREAKFAST

8:15 - 9:15 AM — *OLYMPUS ROOM*

HOW TO IDENTIFY AND PRIORITIZE RISKS

Demonstration

Jake Ritz, Ascend Learning | Cory Clark, Meazure Learning

Other industries face risks on a daily transactional basis. Bad actors constantly threaten their financial performance, customer retention, regulatory status, and public perceptions. Sound familiar? Join Jake and Cory (Both Operational and IT Risk Leaders) in understanding how risks are categorized, managed, mitigated, and monitored. The assessment industry is constantly attacked by coordinated and increasingly sophisticated bad actors. How should a test program understand and mitigate the risks it is facing?

What can be learned from other industries, and how can we effectively prioritize our time and resources? Risks will forever be present, understanding how to identify and, mitigate, and manage them is critical.

FRIDAY, OCT 18

8:15 - 9:15 AM *CONTINUED...* *FLAGSTAFF ROOM*

IS IT CAKE? OR IS IT AI?

Panel Presentation

Chris Foster & Andrew Marder, Caveon

Inspired by the hit TV show "Is it Cake?", this session promises to tickle your funny bone while engaging your intellect. Welcome to a game of “Is it AI or Is it Human” where participants guess whether what they see, hear, or read is AI-generated or crafted by human hands. Get ready for some mind-bending surprises as you try to discern between the work of the artificially intelligent and that of human hands.

This session explores the timing differences between a genuine human test-taker and a cunning cheater armed with proxy AI assistance. Discover how the rhythm of response reveals the telltale signs of deception, and learn how to leverage this knowledge to safeguard the integrity of tests and assessments. Together, we can learn to harness the hidden clues that expose AI-assisted cheating and empower testing programs with the tools to detect and react to the test security threat posed by AI proxy test taking.

TETON ROOM

THE ESSENTIALS OF WEB MONITORING AND ENGAGEMENT

Panel Presentation

Michael Clifton, Cornerstone Strategies | Bryan Friess, Pearson VUE | Jim Hussey, ACT Education Corp

This presentation is intended for audiences seeking an introductory overview of web monitoring and engagement. It focuses on the essentials of building a web monitoring program, such as identifying targeted domains and applications, creating search strategies, and identifying connections between websites. It also addresses strategies for engaging with website owners, such as through relationship building, evaluating a website's terms and conditions, interacting anonymously, and secret shopping. Finally, it provides an overview of the options available to testing organizations related to enforcement, such as the DMCA procedure or the UDRP procedure. The presentation is intended for those who are new to test security or who are thinking about creating a web monitoring capability to supplement their security program.

STANDARD PRESENTATIONS: "EVOLVING METHODS IN STATISTICAL DETECTION" (PART 2)

8:15-9:15 AM
SNOWBASIN ROOM

Weighted Answer Similarity Analysis

Nicholas Trout & Kylie Gorney, Michigan State University

Wollack (1997) introduced the omega statistic as a method for detecting answer copying between pairs of examinees. For each pair, the statistic considers whether the observed number of matching answers is significantly larger than the expected number of matching answers. However, one limitation of omega is that it does not account for the particular items on which examinees have matching answers. Therefore, in this paper, we propose a weighted version of omega that takes this information into account. We compare the performance of the new and existing statistics in a simulation study where the following factors are manipulated: test length (20, 40, 80), the proportion of compromised items (0.1, 0.2, 0.4, 0.6, 0.8), recall proportion (0.9, 1), and key accuracy (0.5, 1). Results show that while both the new and existing statistics are able to control the Type I error rate, the new statistic is much more powerful, on average.

A Data Forensics Compass to Help Navigate Your Test Security Surveillance

Greg Hurtz, California State University Sacramento & PSI Services, LLC | Steven Svoboda, PSI Services, LLC

In this presentation we will describe a “compass” we have developed for recognizing normal vs. irregular test taker behavior based on indices of item response and response time patterns. With normal test taker behavior, the compass tends to produce a clear horizontal and vertical arrangement. By analogy, when test takers misbehave, the gravitational force of their behavioral irregularities starts to overwhelm the magnetic balance of this compass’s elements. When this happens, an associated test taker map helps locate individuals associated with the skewed compass, who are worthy of closer inspection. The indices and methods for defining our compass and the associated test taker map will be described, and examples will be provided in real datasets where test security was “going south.”

Searching for Unidentified Patterns of Potential Cheating in Operational Testing Programs

Kirk Becker, Pearson VUE | Chris Busath, Professional Testing, Inc

The test security research literature includes analyses aimed at identifying specific types of cheating behavior. These analyses are frequently based on experience with prior cheating behaviors, and expectations based on the likely effects of exposed content on exam performance. This research will present exploratory research identifying clusters of unusual testing behavior which may be indicative of specific types of dishonest practices. Forensic flags (e.g., average time per item, pretest/operational item performance, response overlap, etc.) for several operational testing programs will be used along with KNN and other clustering methods to identify groupings of test behaviors within programs. These groups will be examined to understand whether they may be indicative of cheating strategies.

FRIDAY, OCT 18

9:15 - 9:30 AM *BREAK*

9:30 - 10:30 AM *OLYMPUS ROOM*

BALANCING SECURITY AND ACCESSIBILITY IN REMOTE ASSESSMENTS: IMPLICATIONS FOR TESTING POLICIES

Coordinated Symposium

Will Belzak, Duolingo | Liberty Munson, Microsoft | Kirk Becker, Pearson VUE | Ashley Norris, Meazure Learning

Remote assessments are becoming more common due to their potential for greater access, convenience, and cost savings compared to in-person testing. However, they present unique challenges, particularly in balancing test security with accessibility, including access for individuals with disabilities. While remote testing can increase accessibility by eliminating the need to travel, it also raises security risks, such as increased opportunities for cheating. Testing policies like bathroom breaks or room scans can further affect this balance.

This symposium explores how to balance security and accessibility in remote assessments, covering topics like testing rules, comparisons between online and test-center proctoring, and strategies for improving accessibility without compromising security.

TETON ROOM

CONTROLLING DIFFICULT SECURITY THREATS WITH TEST DESIGN

Panel Presentation

David Foster, Caveon

Architects integrate security features directly into building designs, detailing structural strength, emergency exits, fire suppression systems, and more. Similarly, test architects—psychometricians—design tests with security in mind, leading to more accurate and useful scores. However, the rise of technological risks, including AI-driven threats, requires a shift in our approach to test design. Traditional methods are no longer sufficient, and we must innovate to meet new security challenges.

This session will explore recommended test design modifications to counter emerging threats, showcasing the security benefits through scientific experiments and models. We will also provide specific instructions and a live demonstration on how to easily add design features that prevent cheating, stop test theft, and improve test accuracy.

FRIDAY, OCT 18

STANDARD PRESENTATIONS: "POST-BREACH RESPONSE: RISK METRICS, DIGITAL INCIDENT RESPONSE, AND SANCTION TRENDS"

9:30-10:30 AM
SNOWBASIN ROOM

Score-Difference-at-Risk (SDaR): Using Risk Metrics to Anticipate the Impact of Test Security Breaches

Sergio Araneda, Caveon

This presentation introduces a novel metric designed to quantify the risk of extreme score differences due to estimation errors. Inspired by the Value-at-Risk (VaR) concept in finance, Score-Difference-at-Risk (SDaR) calculates the probability of extreme score deviations, providing a more practical link between potential negative outcomes and measurement errors. Focusing on test security, we simulate the impact of examinees having access to exposed test items. By analyzing SDaR metrics, we demonstrate how this metric can be used to assess the risks associated with test security breaches, providing insights into the potential consequences of cheating and pre-knowledge. This presentation highlights the utility of SDaR for practitioners to anticipate and address test security issues, ensuring the integrity of educational assessments.

Customer Obsession Digital Incident Reports

David Green & Julie Wilt, College Board

In this session, we will explore our organization's transition from traditional pencil-and-paper testing to a digital format, focusing on the revolutionized management of incident reports. This shift emphasizes rapid and effective resolution processes, adhering to our commitment to service excellence. We prioritize resolving all teacher-filed incident reports within ten business days and addressing every customer inquiry within 48 hours. Our main customers are students and educators, and we will discuss strategies to better serve them, enhancing our brand's trust and integrity. Key highlights of the presentation include:

- ***Customer Obsession:*** Ensuring all interactions, whether report filings or inquiries, are handled with utmost respect. Our commitment to rapid responses within 48 hours and resolving issues within ten days reinforces our dedication to service and integrity.
- ***Dedicated Response Teams:*** Specialized teams across different time zones ensure timely and efficient report handling, crucial for maintaining high test integrity and stakeholder trust.
- ***Digital Transformation Benefits:*** The digital shift has streamlined incident management with real-time monitoring and faster issue identification, enhancing operational efficiency and data security.

We will share valuable insights and lessons from our digital transformation, offering actionable strategies for enhancing test security and incident management in digital environments, showcasing our unwavering commitment to customer obsession.

(Continued Standard Presentations)

FSBPT Security Investigation Sanctions: Who are they and where are they now?

Anna Canning, Federation of State Boards of Physical Therapy (FSBPT)

The aim of this investigation is to explore trends among the Federation of State Boards of Physical Therapy's past exam security sanctions for the National Physical Therapy Examination (NPTE). The researchers are most interested in who was sanctioned, what they were sanctioned for, and how sanctions may have affected these individuals in the long term. Key findings reveal that a majority of the sanctioned individuals were testing at the Physical Therapist level, were not yet licensed, and had taken the NPTE at least once prior to their violation. While the most common offense has historically been copyright infringement, this violation has had a dramatic drop in frequency over the past decade following the implementation of new security measures. Furthermore, monetary sanctions did not appear to pose as a significant barrier toward getting licensed, while NPTE registration bans of one-to-five years resulted in fewer candidates eventually obtaining a license, especially at the Physical Therapist Assistant level. These findings can be utilized to evaluate whether the existing sanctions model is effectively serving its purpose to protect the public by revealing some of its strengths and limitations.

FLAGSTAFF ROOM

UNRAVELING THE IMPACT OF CURRENT SEARCH AND SHARING TECHNOLOGIES ON YOUR WEB MONITORING EFFORTS

Panel Presentation

Cary Straw & Jen Baldwin, Caveon

Here's the truth: your web monitoring strategies are likely outdated. The online landscape changes rapidly, making last year's tactics ineffective. To stay ahead, you must understand the technical challenges, from search engine algorithms to decentralized platforms and encryption technologies.
This session will cover the latest advancements in web, social, and search technology, offering actionable recommendations to keep your monitoring efforts effective.

Key areas include: Search Engine Dynamics; Social Media Surveillance; Encryption, Tracking, and Privacy Challenges; Emerging Technologies and Adaptive Tactics; and Proactive Monitoring Strategies. Gain practical insights from industry experts with 40+ years of experience, and learn through real-world case studies how to remain agile amid rapid technological change.

FRIDAY, OCT 18

CLOSING KEYNOTE

10:45 AM – 12 PM | BALLROOMS A & B

COTS KEYNOTE DEBATES

Camille Thompson, College Board | Kim Brunnert, Elsevier | Michael Clifton, Cornerstone Strategies | Ray Nicosia, ETS

Join us for the Closing COTS Debates! This always informative and definitely entertaining keynote session will feature industry luminaries as they debate topics of interest to testing professionals. This year, we will explore whether high-stakes testing programs should accept mobile drivers licenses, a question which has been cropping up more and more as additional governments participate in mobile identity solutions. With the maturing capability to secure and monitor testing on a test taker's own device, we will also explore whether college and commercial test centers should allow test takers to bring their own devices to the test center. Is it time to accept mobile ID's and bust open test centers so test takers can BYOD? Join us for a spirited debate as we discuss these issues!

A SPECIAL THANK YOU TO OUR CO-HOSTS:

Pearson VUE

caveon

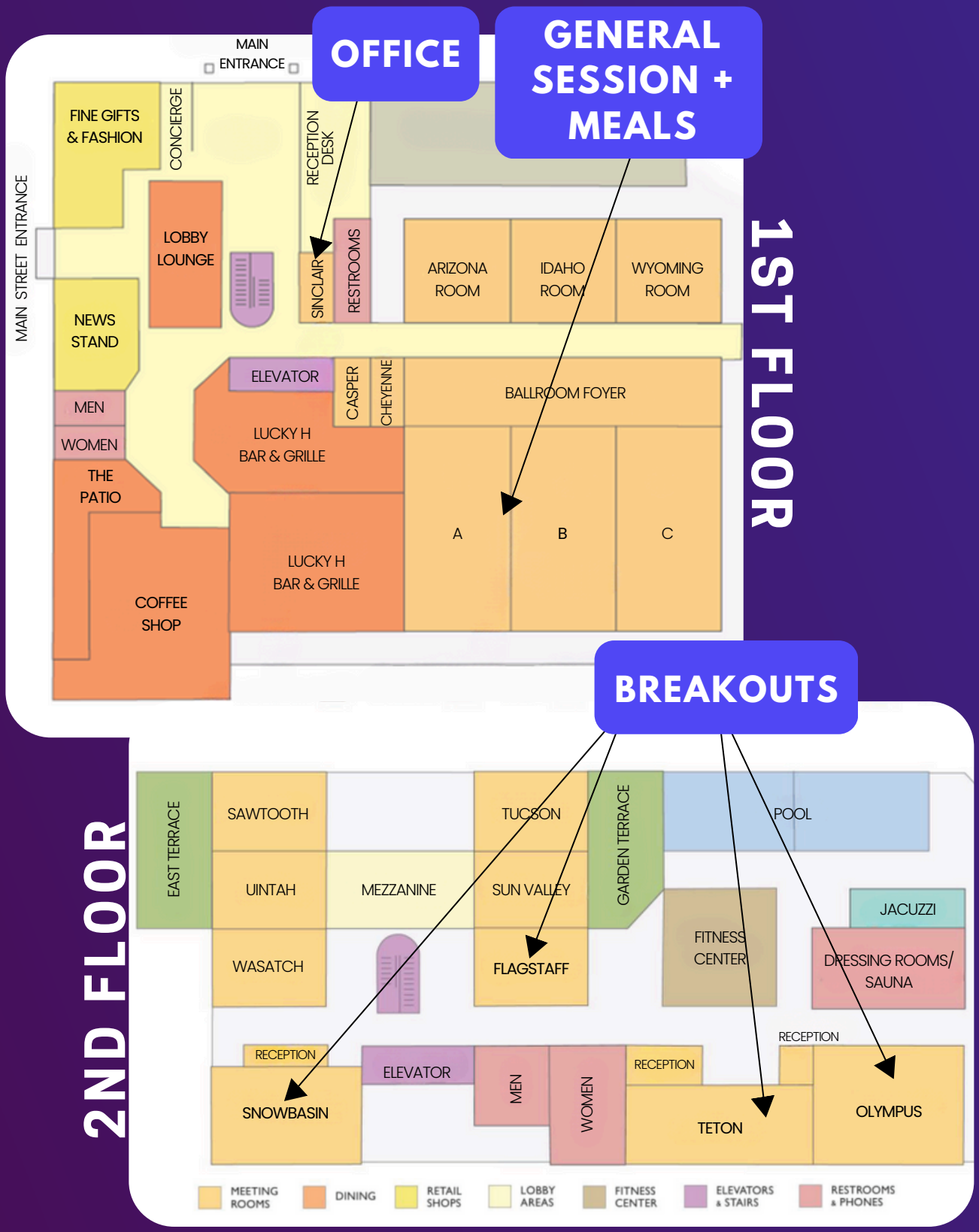
MEASURE LEARNING

Honorlock

duolingo english test

CONFERENCE MAP

THE LITTLE AMERICA HOTEL



GENERAL INFO

TIPS FOR A PRODUCTIVE AND MEMORABLE COTS 2024!

PRESENTER INFORMATION

If you are scheduled to present at COTS, please go to bit.ly/cotspresenter to view important updates and information.

If you encounter any technical issues during your presentation, please call Brinnlie Knavel for immediate assistance: (801) 520-8758.

FOOD ALLERGIES & DIETARY RESTRICTIONS

All food allergies and dietary restrictions identified during the registration process have been communicated to the catering staff. If at any point you should have questions, please speak with one of the hotel staff or visit the registration desk.

LOCAL ATTRACTIONS

Salt Lake City offers a unique blend of natural beauty, cultural attractions, and rich history, making it the perfect destination for those seeking both adventure and inspiration.

Explore local attractions here: https://bit.ly/slcactivities

NEED HELP?

If you have any questions, please visit the registration desk or call Brinnlie Knavel: 801.520.8758

SESSION INFO

YOU CAN LOOK FORWARD TO THESE PRESENTATION FORMATS AT COTS 2024

(15-20 mins) Standard Presentation
In Standard Sessions, two to three groups will present on similar topics during one 60-minute time slot. This means that each presentation will be 15-20 minutes long.

(60 mins) Poster Presentation
60-minute session to include multiple posters on various topics. Poster presentations are informal and involve one-on-one interactions with many attendees.

(10-20 mins) Coordinated Symposium
Three to five separate presentations, all focused on a common theme. One of the presentations may consist of a discussion, analysis, and/or contextualization of another session or sessions. All symposia will occupy a 60-minute time slot.

(60 mins) Panel Presentation
Two to five individuals discussing different aspects of a common theme with limited audience participation. All panel presentations will occupy a 60-minute time slot.

(60 mins) Demonstration
60-minute session in which presenters demonstrate a technique or method related to a core aspect of test security.

(60 mins) Facilitated Roundtable
60-minute session that promotes networking and invites audience engagement around an important test security topic. Session will begin with a short, informal introduction to frame a conversation, followed by a free-flowing dialogue among audience members.

NEED HELP?

If you have any questions, please visit the registration desk or call Brinnlie Knavel: 801.520.8758

SESSIONS BY ROOM

THURSDAY

OLYMPUS

How to Thrive in Your Rookie Season
Schoenig, Thompson, Nicosia, Alam

Friend or Foe? Generative AI in Assessment
Gonthier, Yunger, Weinstein, Dight

Test Security Myths Debunked
Svensson, Bostwick, Meade, Chamoun

How AI is Impacting Exam Security Programs
Brunnert, Hussey, McCauley, Schoenig

Hot to Stay One (Or More!) Steps Ahead of Online Cheating Rings
Schoenig, Samit, Snyder, Hastings

SNOWBASIN

Pre-Keynote Kickoff Workshop: The What & How of Technology-Enabled Cheating
Schoenig, Nicholson, Clark, Hastings

Test Security in Longitudinal Assessment
Sauder, Lee, Yee Prendez, Wollack

Standard Presentations:
"New Approaches in Exam Security"

Standard Presentations: *"Considering Proctoring and Exam Security"*

Standard Presentations: *"Approaches to Detecting Examinee Preknowledge"*

Standard Presentations: *"Evolving Methods in Statistical Detection (Part 1)"*

TETON

Generative AI in Secure Item Development for State Assessment
Drane, Quesen, Araneda, Olson

Through-Year and Interim Benchmark Assessments by States
Olson, Black, Shockley, Drane, Jordan

Dealing with AI Cheating Threats and Solutions
Foster, Foster

Using Test Security Data to Improve the Monitoring of Schools
Olson, Hicks, Ragsdale, Fenby, Drane

Remote Test Administration for State Assessments
Olson, Tohinaka, Carter, Butcher, Drane

FLAGSTAFF

Conducting Candidate Interviews in Exam Security Investigations
Samit, Friess, Morris

Using ChatGPT for the automatic creation of SmartItems™
Araneda

Enhancing Test Security: Real-Time Decision Making & Automatic Interventions in High-Stakes Tests
Gonthier, Tucker, Nicosia

Similarity Statistics on CATs?!
Toton, Scott, Huynh, Butterbaugh

How Delivery Modality Influences Candidate Behavior and Performance
Svensson, Bostwick, Muir

SESSIONS BY ROOM

FRIDAY

OLYMPUS

How to Identify and Prioritize Risks
Ritz, Clark

Balancing Security and Accessibility in Remote Assessments: Implications for Testing Policies
Belzak, Munson, Becker, Norris

TETON

The Essentials of Web Monitoring and Engagement
Clifton, Friess, Hussey

Controlling Difficult Security Threats with Test Design
Foster

SNOWBASIN

Standard Presentations: *"Evolving Methods in Statistical Detection (Part 2)"*

Standard Presentations: *"Post-Breach Response: Risk Metrics, Digital Incident Response, and Sanction Trends"*

FLAGSTAFF

Is it Cake? Or Is it AI?
Foster, Marder

Unraveling the Impact of Current Search and Sharing Technologies on Your Web Monitoring Efforts
Straw, Baldwin

PRESENTER RESOURCES

REMINDER:

Presenters are responsible for managing their own time during the session. Please be mindful of the schedule and ensure your presentation, *including any Q&A*, stays within your allotted time.

Tech/AV Support:

- Miles Johnson (801) 448-3970
- Mario Trujillo (801) 259-3592

Phone Support:

- Call Brinnlie Knavel (801) 520-8758 for immediate assistance

Wi-Fi Login:

Network: LA Red
Password: riverrun

THANK YOU TO OUR SPONSORS

CO-HOSTS:

caveon®

Pearson VUE

Honorlock®

MEASURE LEARNING

duolingo english test

FRIENDS:

*psi

PROMETRIC

NCTA NATIONAL COLLEGE TESTING ASSOCIATION

ncme NATIONAL COUNCIL ON MEASUREMENT IN EDUCATION

kryterion by DRAKE INTERNATIONAL