Fifth Annual
October 18–20
Cedar Rapids, IA

Welcome

Hello! I am proud to welcome you to America's heartland and Cedar Rapids Iowa, the "City of Five Seasons"—if anyone knows what the fifth season is please let me know! Pearson is also proud to sponsor the 2016 Conference on Test Security and I thank you all for your attendance, support and participation.

"The Corridor", as referred to by us locals, is the nation's Mecca for Assessment development, technological enhancement, and processing of assessments with organizations like Pearson, ACT, The College Board and the University of Iowa located here—all known providers and supporters of learning though better measures. Pearson is happy to partner with all of these organizations and to sponsor this conference as we bring more focus and instructional support, enhanced by technology, to learning through better measures.

Our rapidly changing world, particularly with respect to technology, offers untold challenges in areas that are expanding rapidly. Such things as: "Personally Identifiable Information (PII)", the use of social media, fraud involving the use of assessments and legal issues regarding things like "off shore use or storage of testing data" are all issues we deal with daily. I am sure you will find these issues and others discussed during this conference and I suggest you engage with questions and actively participate as we all try to keep up with the logarithmic advance of change.

We have tried to expand the venue this year by provide more opportunities to include a broader base of participation both in sessions and in other activities. We hope this allows for more networking and more sharing—be it at sessions, in the halls or at social events. Let me or the Planning Committee know if you run into any issues but be sure to take advantage of the global expertise present this year. Thank you and let's all have a great conference.

Jon S. Twing, Ph.D.
Senior Vice President, School Assessments
Pearson

Conference History

The Conference on Test Security began in 2012 as the Conference on the Statistical Detection of Test Fraud and focused primarily on statistical methods. In 2013, an Executive Board was selected to provide guidance. In 2014, the Conference Executive Committee expanded the scope of the conference to include a broader range of test-security subjects, and the name changed to its current form to reflect the broader range of topics and experts presenting at the conference.

Wi-Fi

Guests will want to access the network “Marriott Conference” and enter conference code: **cidmc** (code is not case sensitive).

Name Badges

Name badges are not only a distinctive fashion statement; they are an important way for conference staff to identify conference attendees. Please wear your name badge in all sessions. **New this year:** Name badges will also contain a barcode that will be scanned at the door of each session and used to track attendance. Please have your name badge ready. Any attendee without a name badge will have to wait until others are seated to then manually check-in for the session with conference staff. This will allow you to receive session information specific to what you attend when the conference is over.

Conference URL

cete.ku.edu/2016-conference-test-security

Local Time

Iowa time zone is: Central Time Zone UTC -6:00

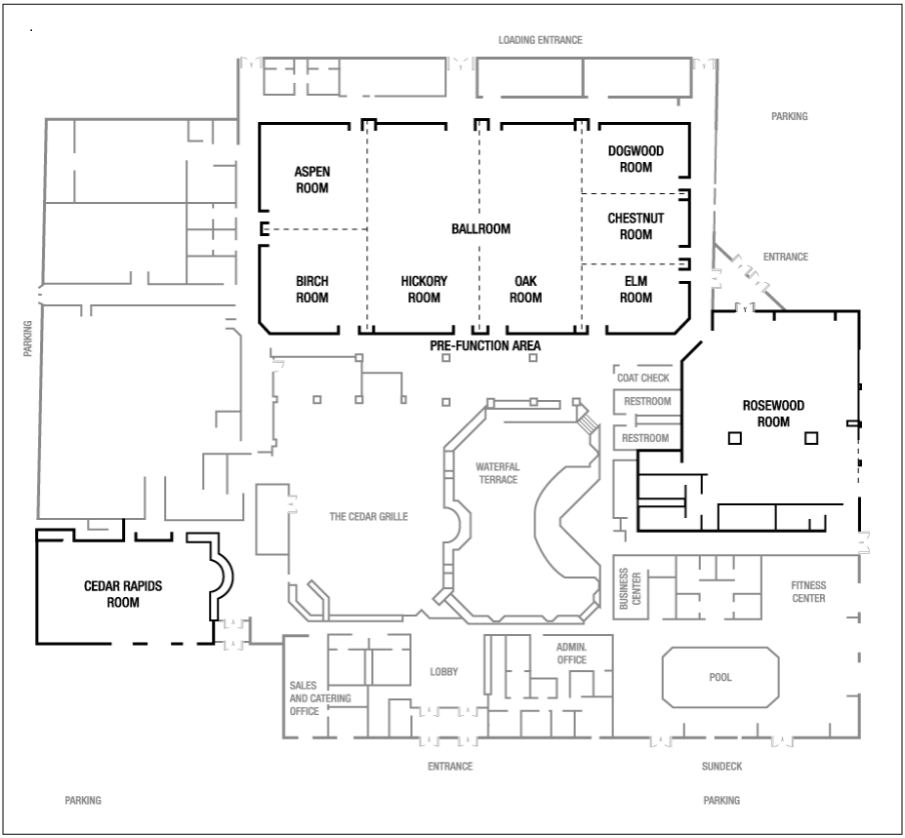
Thursday Night: Downtown Cedar Rapids

Sometimes the best kind of breakout session is the one you didn’t have to plan. There is a list of restaurants listed on the website for your convenience, or stop by the hotel concierge for more information on the shuttle service.

Schedule At-A-Glance

Tuesday, October 18		
Workshop Check-in/Registration – lunch on own	Hickory/Oak Foyer	12:00–1:00
Workshop 1	Chestnut & Elm Rooms	1:00–2:45
Break	Hickory/Oak Foyer	2:45–3:15
Workshop 2	Chestnut & Elm Rooms	3:15–5:00
General Conference Check-in – not needed if attended workshops	Hickory/Oak Foyer	5:00–7:00
Wednesday, October 19		
Breakfast	Hickory/Oak Ballroom	7:00–8:00
Welcome Keynote Speaker – David Callahan	Hickory/Oak Ballroom	8:00–9:30
Break	Hickory/Oak Foyer	9:30–9:45
Session 1	Chestnut, Dogwood, Elm, & Rosewood Rooms	9:45–10:45
Break	Hickory/Oak Foyer	10:45–11:00
Session 2	Chestnut, Dogwood, Elm, & Rosewood Rooms	11:00–12:00
Lunch	Hickory/Oak Ballroom	12:00–1:00
Plenary Session	Hickory Ballroom	1:00–2:30
Break	Hickory/Oak Foyer	2:30–2:45
Session 3	Chestnut, Dogwood, Elm, & Rosewood Rooms	2:45–3:45
Break	Hickory/Oak Foyer	3:45–4:00
Session 4	Chestnut, Dogwood, Elm, & Rosewood Rooms	4:00–5:00
Poster Session	Cedar Rapids Room	5:00–5:50
Social Networking Gathering	Waterfall Terrace	6:00–8:00
Thursday, October 20		
Breakfast	Hickory/Oak Ballroom	7:00–8:00
Session 5 (90 min)	Chestnut, Dogwood, Elm Rooms	8:00–9:30
Break	Hickory/Oak Foyer	9:30–9:45
Session 6	Chestnut, Dogwood, Elm Rooms	9:45–10:45
Break	Hickory/Oak Foyer	10:45–11:00
Session 7	Chestnut, Dogwood, Elm Rooms	11:00–12:00
Lunch	Hickory/Oak Ballroom	12:00–1:00
Session 8	Chestnut, Dogwood, Elm Rooms	1:00–2:00

Conference Layout



Keynote Speaker

David Callahan

Co-founder of Demos & Author of *The Cheating Culture*

In a cutthroat economic climate, everybody wants to get ahead, and decades of deregulation have made it easy to bend the rules. The author of *The Cheating Culture: Why More Americans Are Doing Wrong to Get Ahead*, David Callahan offers explanations for the proliferation of cheating. He argues that when the middle class sees wealthy cheaters get away with nothing more than a slap on the wrist, it inspires them to follow suit.

Callahan is co-founder of the independent think tank Demos, a public policy center based in New York City that combines research and advocacy to strengthen democracy and expand economic opportunity within the United States. It focuses on creating an open resource of knowledge and learning that operates beyond traditional parties, identities, and disciplines.

Callahan is the author of *Trading Up* as well as three books on US foreign policy and international affairs, including *Unwinnable Wars: American Power and Ethnic Conflict*. His latest book, *Fortunes of Change: The Rise of the Liberal Rich and the Remaking of America*, was published in 2010. His articles have appeared in *The American Prospect, Foreign Policy, The New York Times, USA Today*, and *The Washington Post*. He has also been a frequent commentator on television programs on CBS, CNN, Fox News, MSNBC, and PBS and has been a regular guest on radio talk shows across the United States, including such NPR programs as “The Connection,” “Morning Edition,” and “The Tavis Smiley Show.”

Prior to co-founding Demos, Callahan was a fellow at the Century Foundation,

engaging in wide-ranging public policy research and analysis.

David Callahan frequently speaks about issues of ethics and integrity to universities, corporations, and various associations, explaining how our cheating culture originated and offering solutions to fix it.

TUESDAY, OCTOBER 18

Workshop check-in/registration

Hickory/Oak Foyer - Welcome Table

12:00–1:00 pm Lunch on own

Workshop 1 — 1:00–2:45 pm

Option 1: Test Security for State & District Testing
Rachel Schoenig
Elm Room

Testing in K-12 entails unique exam security challenges. This workshop is designed specifically to address those challenges, providing a focus on fundamental test security practices to deter and detect potential test security breaches. Participants will leave with practical tools that can be used this academic year to improve test security at the State and District level.

Option 2: Statistical Analysis of Test Fraud
Dennis Maynes
Chestnut Room

This workshop is for individuals adept at statistical analysis. Participants will go deeper into particular statistical and analytical processes that occur when testing data are analyzed for evidence of test fraud. Bring your laptop and come prepared to assess data sets.

Break – Hickory/Oak Foyer

2:45–3:15 pm

Workshop 2 – 3:15–5:00 pm

Option 1: Conducting a Test Security Investigation
Mike Clifton
Elm Room

Even the strongest security protocols can be breached. This workshop addresses the steps to conduct a defensible test security investigation. Participants will learn effective methods for planning an investigation, gathering evidence, conducting interviews and preparing a report.

Option 2: How the Design of a Test Can Reduce Security Problems
David Foster
Chestnut Room

Test design can either contribute to or help reduce exam security concerns. This workshop will focus on ways a testing program can exert control over the security of its exams from the outset, enabling test developers to build tools to deter and detect security problems into the exams and exam delivery systems.

Conference check-in/registration – not needed if attended workshop sessions

Hickory/Oak Foyer
5:00–7:00 pm

WEDNESDAY, OCTOBER 19

Check-in/Registration & Breakfast – Hickory/Oak Foyer & Ballroom
7:00–8:00 am

Keynote Speaker — 8:00–9:30 am — Hickory/Oak Ballroom

Co-founder of Demos & Author of *The Cheating Culture*
» David Callahan

Break – Hickory/Oak Foyer
9:30–9:45

Session One — 9:45–10:45 am

Presentation 1 – Elm Room

What is Non-Independent Test Taking? White Papers You Won't Want to Miss!
» Janet Lehr, John Sowles, & Beverly Bone

Demonstration-Paper 1 – Chestnut Room

SIFT: Software for Investing Fraud in Testing
» Nathan Thompson & Terry Ausman

Presentation 2 – Dogwood Room

Dousing the Flames of a Media Firestorm
» Richelle Gruber, Ray Nicosia, & Steve Addicott

Presentation 3 – Rosewood Room

Access vs. Security: Considering for Accessibility
» Chelsea Seeley & Clare Rose

Break – Hickory/Oak Foyer
10:45–11:00 am

Session Two — 11:00–12:00 pm

Presentation 4 – Elm Room

Armadillos and other critters: Making your tests cheat-proof and theft-proof.
» David Foster

Presentation 5 – Chestnut Room

The Lifelong Journey to Valid Assessments
» Walt Drane, Alana Chamoun, Michael Clifton, Steve Addicott

Presentation 6 – Dogwood Room

Conducting a School Observation Program for State Assessments
» David Ragsdale

Presentation 7 – Rosewood Room

Real-Time Anomaly Detection for Online Testing
» Jim Sherlock

Lunch – Hickory/Oak Ballroom

12:00–1:00 pm

Plenary Session — 1:00–2:30 pm — Hickory/Oak Ballroom

Overview of the 'Handbook of Quantitative Methods for Detecting Cheating on Tests'
» Gregory Cizek, James Wollack, & Lorin Mueller

Session Three — 2:45–3:45 pm

Presentation 8 – Elm Room

Actionable Intelligence: Real world examples on how to move from reactive to proactive
» Bryan Friess, Brent Morris, & Laura Duffy

Presentation 9 – Chestnut Room

Statistical Detection: Where Do I Start?
» Nathan Thompson & Terry Ausman

Presentation 10 – Dogwood Room

Shine Spotlights on Test Sites: Effective Use of On-Site Monitoring to Assess & Improve Test Security
» Marc Weinstein, Thomas Gera, & Walt Drane

Presentation 11 – Rosewood Room

Quantitative Methods for Detecting Cheating on Tests: Methodologies & Applications
» William Skorupski, Jessalyn Smith, & Joseph Martineau

Break – Hickory/Oak Foyer

3:45–4:00 pm

Session Four — 4:00–5:00 pm

Panel 1 – Elm Room

The Every Student Succeeds Act & Implications for Test Security in States
» Dr. John Olson, Tammy Howard, Craig Walker, David Ragsdale, John Fremer

Presentation 12 – Chestnut Room

Web Monitoring Stories: A Tale of Three Testing Programs
» Janet Lehr, Karen Wood, & Alana Chamoun

Presentation 13 – Dogwood Room

Protecting Your Testing Program: A Practical Perspective
» Rachel Schoenig & Joe Brutsche

Presentation 14 – Rosewood Room

Real Cheating Data for Researchers: A Description of the Common Datasets from the 'Handbook of Quantitative Methods for Detecting Cheating on Tests'
» James Wollack & Gregory Cizek

Poster Session — 5:00–5:50 pm — Cedar Rapids Room

Assessing Interest in a Test Security Journal
» Mark Albanese & other from Executive committee

Exploring the Use of Multidimensional Scaling for Test Irregularity Cases
» Chi-Yu Huang, NooRee Huh, & Qing Xie

A Data Forensics Tool for the Statistical Detection of Test Fraud

» Sebastiaan de Klerk & Bernard Veldkamp

Caveon Core: Test Security Intelligence Platform

» Ben Fisher

Two-Stage Statistical Screening for Unusual Response Similarity in Large-Scale Assessments

» Mengyao Zhang

Social Networking Gathering – *Waterfall Terrace*

6:00–8:00 pm

Heavy hors d'oeuvre will be provided

THURSDAY, OCTOBER 20

Breakfast – *Hickory/Oak Ballroom*

7:00–8:00 am

Session Five — 8:00–9:30 am

Presentation 15 – Elm Room

Ensuring Trust in Assessment Results: A Global Perspective

» Rachel Schoenig, Ray Nicosia, & Ardeshir Geranpayeh

Presentation 16 – Chestnut Room

Backlash: How the Purposeful Disclosure of Test Items on the Internet Impacts Test Security

» John Fremer & Craig Mills

Break – *Hickory/Oak Foyer*

9:30–9:45 am

Session Six — 9:45–10:45 am

Presentation 17 – Elm Room

Building a Highly Available Online Testing Platform with Specific Focus on DDoS Protection

» Jim Sherlock

Presentation 18 – Chestnut Room

Statistical Methods of Detecting Test Fraud: Can We Get More Practitioners on Board?
» Nathan Thompson

Presentation 19 – Dogwood Room

New Test Security Requirements in USED Peer Review & States' Responses
» John Olson, John Fremer, Leila Williams, Jason Kolb, & Kathy Moore

Break – Hickory/Oak Foyer

10:45–11:00 am

Session Seven — 11:00–12:00 pm

Presentation 20 – Elm Room

Test Integrity in NITE Assessments: Prevention & Detection of Cheating, & Then What?
» Yekutiel Weiss

Presentation 21 – Chestnut Room

Using Test Results with Identical Answers to Obtain Credible Evidence of Test Security Breaches
» Dennis Maynes

Presentation 22 – Dogwood Room

A Test Security Framework
» Jamie Mulkey & Rachel Schoenig

Lunch – Hickory/Oak Ballroom

12:00–1:00 pm

Session Eight — 1:00–2:30 pm

Presentation 23 – Elm Room

Bayesian Detection of Cheating on Tests
» Dr. Wim J. van der Linden

Presentation 24 – Chestnut Room

Extension & Cross-validation of a Profile of Statistical Indices for Detection of Aberrant Test Responses

» Greg Hurtz & John Weiner

Response Time Based Non-Parametric Person-Fit Index for Aberrant Responding Behavior Detection in Large-scale Assessment

» Kaiwen Man & Hong Jiao

Presentation 25 – Dogwood Room

The Efficacy of Using Item Response Latency Statistics to Detect Pre-Knowledge

» Sarah Thomas

Presentation 1

What is Non-Independent Test Taking? White Papers You Won't Want to Miss!

Janet Lehr, John Sowles, & Beverly Bone

Trying to come up with a good way to tell your test takers that they don't need to cheat? Look no further, the ITCC just did. The ITCC's (IT Certification Council) Security Subcommittee took on the task of developing two whitepapers on Non-Independent Test Taking (NITT) for use with their IT certificants and those who employee IT professionals.

These two papers describe what NITT is, why it's cause for alarm, and activities that can be used to prevent it. The presenters will discuss the development process for these papers and how they are marketing and socializing them throughout IT community. These whitepapers have greater applicability beyond IT certification programs. There are key messages that all high stakes testing programs can use with their constituents.

Demonstration-Paper 1

SIFT: Software for Investing Fraud in Testing

Nathan Thompson & Terry Ausman

SIFT is a software program specifically designed to bring data forensics to more practitioners. Widespread application of data forensics, like other advanced psychometric topics, is somewhat limited when an organization's only options are to hire outside consultants or attempt to write code themselves. SIFT enables organizations with smaller budgets to apply some data forensics by automating the calculation of complex indices as well as simpler yet important statistics, in a user-friendly interface.

The most complex portion is a set of 10 collusion indices (more in development) from which the user can choose. SIFT also provides functionality for response time analysis, including the Response Time Effort index (Wise & Kong). More common analyses include classical item statistics, mean test times, score gains, and pass rates. All indices are also rolled-up into two nested levels of groups (for example, school and indices are also rolled-up into two nested levels of groups (for example, school and district or country and city) to facilitate identification of locations with issues.

All output is provided in spreadsheets for easy viewing, manipulation, and secondary analysis. This allows, for example, a small certification organization to obtain all of this output in only a few hours of work, and quickly investigate locations before a test is further compromised.

Presentation 2

Dousing the Flames of a Media Firestorm

Richelle Gruber, Ray Nicosia, & Steve Addicott

While its wildfire season in the West, it's also media wildfire season for many test programs! On any given day, an internet search for "cheating on tests," will find stories of testing programs of all sizes and types that have suffered a breach. This not only brings negative public attention to the program, but also leads stakeholders to wonder how this could happen, what steps the program had taken to prevent such an occurrence, and most importantly, whether the program's exam results are valid.

This session explores recent high-profile breaches, and provides examples of how high-stakes programs have dealt with them. We will walk attendees through media strategies, which include: creating a crisis communication plan before a breach occurs, being proactive with the media before a questionable situation arises, and how to handle things after developing effective messaging for all stakeholders.

ETS' head of test security, Ray Nicosia, a former journalist, has experience on both sides of the media challenge faced by test program leaders. He and Caveon's media expert Richelle Gruber will help attendees craft a plan in order to be prepared when you're approached by media regarding a test security issue.

Presentation 3

Access vs. Security: Considering for Accessibility

Chelsea Seeley & Clare Rose

Presentation on alternatives for providing access to users with disabilities to secure test content online. Review legal requirements and web standards for providing access to individuals with disabilities using external third-party assistive technology hardware and software. Review of challenges in providing access to both embedded operating system accessibility supports as well as integration of third party software. Summarize the considerations for accessibility while maintaining test security, and the limitations in meeting both user needs for access and security requirements. Review of current state solutions to enable access during secure testing, as well as ideas for improving future state secure testing solutions through the use of web-based user profiles.

Presentation 4

Armadillos and Other Critters: Making Your Tests Cheat-Proof & Theft-Proof

David Rose

It is amazing what we can learn from the animal kingdom. What kind of help can, animal varieties, especially those that have survived millions of years, give us to protect our tests? How can we emulate these small creatures to make tests that are more secure? We've grown up in a testing world that paid very little attention to security until the last ten years, which is probably why we are experiencing such problems today. The first and most important step toward a successful security effort is to help your tests protect themselves. Perhaps it is unrealistic to set as a goal, total prevention of theft and cheating, but it is worth considering. Or if prevention falls a little short, is it possible to deter individuals from committing test fraud? This session covers a host of rarely-used item and test design features that could save your life (actually save your tests? lives and save you a lot of trouble, money and headache). And you'll learn why, despite the obvious road kill evidence, animal species such as armadillos and skunks continue to thrive.

Presentation 5

The Lifelong Journey to Valid Assessments

Walt Drane, Alana Chamoun, Michael Clifton, Steve Addicott

Because threats to valid assessment results are continually evolving, measurement professionals need to maintain a focus and commitment to lifelong learning in the area of test security. An assessment that is not administered securely cannot be presumed valid. Test security threats occur and change at the various stages of test-taker development and learning from primary education through higher education to professional certification and beyond.

Over the last 20 years in the United States, primary education has been reformed by federal laws that require educators to raise student test scores. As a result, some educators have fraudulently manipulated scores. This session will address how state and local education agencies can contend with these threats.

All over the country, students feel great pressure to gain admittance to prestigious universities. This session will explore the impact of cheating scandals in university admissions testing and what can be done to reduce the likelihood of their occurrence.

Having earned a college degree, graduates seek careers, which require assessments, either for employment selection purposes and/or to obtain credentials. Because of the high stakes associated with certifications, test security threats are growing concerns. This session will overview validity issues in certification testing, and strategies for battling them.

Presentation 6

Conducting a School Observation Program for State Assessments

David Ragsdale

One of the recognized best practices for the security of state assessment programs is to conduct school observations during test administration. But there are many questions that need to be answered when working out the details. Which schools should be observed? Who should conduct the observations? How should observers be trained? What are they looking for? Should observations be announced or unannounced? What kind of feedback should schools receive? This presentation will discuss these questions and explore the different ways state assessment offices can plan and execute effective school observations.

Presentation 7

Real-Time Anomaly Detection for Online Testing

Jim Sherlock

Pearson has developed a real-time online testing feedback system for use in problem identification, triage, and anomaly analysis. This discussion will dive into the methodology and tooling with specific focus on real-world examples of how this kind of real-time analysis can help detect potential anomalies in the field.

Plenary Session — 1:00–2:30 pm

Overview of the *Handbook of Quantitative Methods for Detecting Cheating on Tests*

Gregory Cizek, James Wollack, & Lorin Mueller

This session will provide an introduction to the newly released *Handbook of Quantitative Methods for Detecting Cheating on Tests*, the most comprehensive volume dedicated to statistical approaches to detect cheating. The edited volume, which includes contributions from many of the best scholars and practitioners in the test security field, seeks to compile, describe, develop, expand, evaluate, and disseminate a body of best practices for identifying cheating on tests. The core of the book is dedicated to introducing new, innovative methodologies; however, the book also includes several review chapters to identify and evaluate current practices, as well as an entire section devoted to practical issues associated with using data forensics operationally.

In this presentation, the co-editors, Cizek and Wollack, will establish a foundation for studying and utilizing statistical approaches to detect cheating, present the framework used throughout the Handbook for thinking about cheating detection methodologies, and discuss the future of quantitative methods to detect cheating on tests. Handbook author Mueller will follow by discussing central themes that emerge throughout the chapters, and what they suggest about how we might improve the development and utilization of quantitative methods for detecting cheating on tests.

Presentation 8

Actionable Intelligence: Real World Examples on How to Move from Reactive to Proactive

Bryan Friess, Brent Morris, & Laura Duffy

As test security professionals, we are all exposed to a variety of different data points that provides potential indication of proxy testing, test center anomalies or candidate misconduct on a daily basis. How can you sift through the various false positives? More importantly, how can a program shift from being reactive to proactive? The answer: actionable intelligence.

Actionable intelligence is the process of determining the insights that need to be harvested from your data to take specific, risk based and decisive actions. These actions can be different for every program and specific situation. This session will discuss how one Exam Security program partners with their delivery partner to develop actionable intelligence that enables their joint teams to defensibly act against both test centers and candidates. We will discuss how the collective teams have identified, developed and implemented specific criteria and processes that can be monitored on an ongoing, proactive basis.

Presentation 9

Statistical Detection: Where Do I Start?

Nathan Thompson & Terry Ausman

How can statistical detection of test fraud be better directed? This presentation will begin by cogently outlining various types of analysis into a framework by aligning them with the hypothesis each intends to test, show that this framework should be used to direct efforts, and then provide some real experience by applying these to real data sets from K-12 education and professional certification.

In the first section, we will start by identifying the common hypotheses to be tested, including: examinee copying, brain dump makers, brain dump takers, proctor/teacher involvement, low motivation, and compromised locations. Next, we match up analyses, such as how collusion indices are designed to elucidate copying but can also help find brain dump takers. We also provide deeper explanations on the specific analyses.

In the second section, we apply this framework to the analysis of real data sets. This will show how the framework can be useful in directing data forensics work rather than aimlessly poking around. It will also demonstrate usage of the statistical analyses, facilitating learning of the approaches as well as driving discussions of practical issues faced by attendees. The final portion of the presentation will then be just such a discussion.

Presentation 10

Shine Spotlights on Test Sites: Effective Use of On-Site Monitoring to Assess & Improve Test Security

Marc Weinstein, Thomas Gera, & Walt Drane

Sponsors of high-stakes examinations must ensure that tests are administered under standardized conditions and that test administrators follow all applicable policies and procedures to ensure the security of the test and the validity of test results. On-site monitoring of test administrations is a powerful method to assess the fidelity of test administrations, deter, detect and respond to testing irregularities and test security violations, and improve the security of testing programs. Test sponsors may monitor test administrations through several methods, including announced visits and unannounced “secret shopper” visits, where the monitor poses as a test-taker.

During this session, presenters with extensive experience managing monitoring programs for an admissions testing organization, a state department of education and large metropolitan school districts will share monitoring strategies that have proven effective. Presenters will describe the components of effective test administration monitoring programs and explain how monitoring results can be used in conjunction with other data, including data forensics analysis, to provide powerful information about the performance of test administrators and/or test-takers when monitors are not present. Finally, presenters will explain how all of the information gleaned from monitoring can be used to improve test security and test administration policies and practices.

Presentation 11

Quantitative Methods for Detecting Cheating on Tests: Methodologies & Applications

William Skorupski, Jessalyn Smith, & Joseph Martineau

The focus of the *Handbook of Quantitative Methods for Detecting Cheating on Tests* is on new methodologies and their effective use operationally. In this presentation, we go more in depth into the book’s content with presentations from three chapter authors, two from methodological chapters and one from an applications chapter, detailing their contributions. The first presentation (Skorupski) will discuss the use of Bayesian hierarchical linear modeling strategies to detect aberrant growth at the group level, as might be used to identify test tampering/educator cheating or pre-knowledge/inappropriate coaching linked to specific test sites or test preparation centers. The second presentation (Smith) will discuss the use of lognormal and hierarchical response time models to assist in identifying compromised items and detecting examinees with pre-knowledge. The last presentation (Martineau) will address the present and future of accountability testing, with specific attention on security vulnerabilities and cheating prevention strategies.

Panel 1

The Every Student Succeeds Act & Implications for Test Security in States

Dr. John Olson, Tammy Howard, Craig Walker, David Ragsdale, John Fremer

Congress passed the Every Student Succeeds Act (ESSA) in early 2016, which will require a number of changes to state assessment and accountability systems. In addition, in October 2015, prior to the passing of this new ESEA law that replaces NCLB, the USED issued a letter to states highlighting a new "Testing Action Plan" that supports states in taking a closer look at their assessment systems to improve their overall efficacy and quality. Among the many issues described in both ESSA and the guidance from USED are critical ones like the validity and fairness of test scores and having high-quality processes in place to ensure this.

A related and very important issue that needs to be discussed as states leverage funds from ESSA to support "smarter high-quality assessments" is the need for enhanced test security. The integrity of an assessment hinges on the accuracy and fairness of the test scores, and if cheating is happening in districts and/or schools, then the state's results, as well as their use to make important decisions, are in question. ESSA funds can be used to improve the quality of assessments, which includes the policies, procedures, and practices that states use to develop, administer, and report test scores. An approach that focuses more on implementation of best practices for the prevention, detection, and remediation of testing irregularities or improprieties is needed by many, if not all, states.

In this session, a panel of experts will discuss the regulations as specified in ESSA and the support/guidance documents from USED and others that guide states in the implementation of better, and more secure, assessments. Representatives with many years in the testing industry and expertise on test security for states and three states will discuss the new law from their perspectives and recommend ways to make enhancements to state assessment programs, with a particular focus on their security. Audience participation will be encouraged to interact with the panel members on the nuances of ESSA as they relate to improving test security. In addition, key documents will be shared with attendees that highlight the latest guidance on ESSA and useful resources for states.

Presentation 12

Web Monitoring Stories: A Tale of Three Testing Programs

Janet Lehr, Karen Wood, & Alana Chamoun

Web and media monitoring's main purpose is to remove test content from unauthorized web sites, chat rooms, and social media sites. However, it can also help to understand the landscape of a testing program's online presence. It can provide candidate awareness and education, as well as allow for relationship development to halt test content from being posted in the first place.

Join us for three different tales of testing program web monitoring. Each program is from a different industry sector: IT certification, medical education, and professional certification. Each program will discuss:

- Goals and strategies for removing test and copyrighted content
- How to get site owners to monitor for inappropriate content and conversation
- Exam design strategies that mislead illicit web sites
- How web and media monitoring morphs as the program gains a better understanding of the online environment

Presentation 13

Protecting Your Testing Program: A Practical Perspective

Rachel Schoenig & Joe Brutsche

Trust in test results and protection of an examinee's private data are fundamental requirements of any assessment program. This session will address practical and cost-effective tips to address test security and protection of an examinee's private data. Attendees will leave with useful and accessible tools designed to better identify key test security and examinee data risks and mitigation strategies from a practical business level.

Presentation 14

Real Cheating Data for Researchers: A Description of the Common Datasets from the *Handbook of Quantitative Methods for Detecting Cheating on Tests*

James Wollack & Gregory Cizek

The best way to understand the properties of new cheating methodologies is through a simulation study in which researchers carefully control the cheating mechanism/process, the prevalence of cheating, and the magnitude of compromise. However, the generalizability of simulation results hinges on the reasonableness of the simulation assumptions and the similarity of simulated conditions to those experienced in practice. One of the unique aspects of the *Handbook* is that all methodological chapters applied their methods to one of two common datasets, one from an educational context and one from a professional credentialing context. The application of different methods to a common dataset not only aids in promoting a coherent focus across chapters, but also allows for comparison and critique across methods. In addition, by incorporating information about the various methodologies from simulation studies, it is possible to learn a considerable amount about the datasets themselves, including the likely compromise status of people and items. As the characteristics of these datasets become better understood, the datasets become even more valuable tools for validating future methodologies, within the context of a real-life, operational testing program. Participants in this session will learn about the common datasets used throughout the *Handbook* and will learn how to obtain access to the datasets for their own test security research purposes.

Presentation 15

Ensuring Trust in Assessment Results: A Global Perspective

Rachel Schoenig & Ray Nicosia

Trust in assessment results is a critical aspect of the learning process. As the value and stakes of an assessment increases, however, so too does the willingness of examinees and others to engage in dishonest assessment practices; practices that erode trust in individual assessment results and, more broadly, in our industry. Join seasoned professionals with experience securing workforce skills credentials, certifications and admissions exams across the globe as they share case studies and lessons learned in their practice. Lessons that can help ensure trust in assessment results and support lifelong learning.

Presentation 16

Backlash: How the Purposeful Disclosure of Test Items on the Internet Impacts Test Security

John Fremer & Craig Mills

Recently, there have been a number of examples of purposeful disclosure of test items. Live test items have been Snapchated, Facebooked, and Instagrammed. Test items have appeared in blog posts, where their distribution to parts unknown is eminent. This activity is related to a "backlash" against standardized testing and it can only have detrimental effects on test security. Such disclosure of test items has implications for all types of assessments that are used to help make important decisions.

This panel discussion will explore social media backlash and its impact on test security. The panel will discuss what can be done as by measurement professionals to reduce the number and severity of such exposures and to cope with item exposure when it does occur.

Objectives, as a result of this session, the participant will be able to: 1) Recognize the impact of purposeful test content disclosure 2) Identify examples of purposeful disclosure and how testing programs have coped with the problem 3) Describe solutions for a testing program to employ to discourage and prevent purposeful item disclosure. John Fremer, Moderator/Submitter

Presentation 17

Building a Highly Available Online Testing Platform with Specific Focus on DDoS Protection

Jim Sherlock

Building a highly scalable and resilient assessment platform is a challenge. This discussion will focus on some best practices for building highly available services for the delivery of high stakes assessments. Topics will include a review of current threats, vendor partnerships, and some engineering best practices.

Presentation 18

Statistical Methods of Detecting Test Fraud: Can We Get More Practitioners on Board?

Nathan Thompson

Statistical methods of detecting test fraud have been around since the 1970s, but are still not in general use by most practitioners, instead being limited to a few specialists. The purpose of this presentation is threefold. First, we will discuss the types of methods that have been suggested in the literature, categorizing them in a way to make them more digestible to non-specialists. Second, we will present comparisons of the methods using real data, allowing attendees to better see differences in performance and conceptualization of the indices. Finally, we will discuss approaches to making these methods more generally accessible to practitioners, primarily with the use of user-friendly software or even forensic reports that are directly integrated with online testing platforms. Methods will include classical approaches such as Frary, Tideman, and Watts (1979) and Bellezza and Bellezza (1989) as well as more recent methods such as Wollack's (1997) omega.

Presentation 19

New Test Security Requirements in USED Peer Review & States' Responses

John Olson, John Fremer, Leila Williams, Jason Kolb, & Kathy Moore

In 2016, the U.S. Education Department (USED) included new requirements on test security for peer review of states. All states must now show evidence that their assessment systems are secure. Specifically, Critical Element 2.5, Test Security, asks States to prove they have "implemented and documented an appropriate set of policies and procedures to prevent test irregularities and ensure the integrity of test results." To fulfill this requirement, States must submit documentation of their policies and procedures in four categories of test security: prevention, detection, remediation, and investigation.

One of the best ways for states to show this evidence is to develop a Test Security Handbook that addresses the requirements of Critical Element 2.5. In this session, we will describe the components of Handbooks developed for all types of high-stakes testing programs, including those created for states. Caveon's model for State Test Security Handbooks is to compile/describe all evidence a state has, including training materials, test administration manuals, irregularity reports, memos on state policies/guidance to districts, investigation procedures, and many other things. By putting all this information together into a single document, the Handbook serves as a comprehensive collection of all state policies, practices, and procedures related to test security.

Representatives from three states will describe the evidence submitted to the USED for peer review and the value of Test Security Handbooks. Another presenter, who is a peer reviewer, will discuss what evidence was found particularly supportive of high-quality/secure state assessment systems. Session attendees will receive a summary of the USED peer review security requirements and a model outline and Table of Contents for a State Test Security Handbook.

Presentation 20

Test Integrity in NITE Assessments: Prevention & Detection of Cheating, & Then What?

Yekutiel Weiss

There is a considerable body of literature concerning prevention and detection of suspected test cheating, but there is little information about the validity of the measures being taken after detection. ETS and ACT publish their policy for dealing with suspected irregularities; nevertheless, no reports on hit rate and reliability were found. NITE (Israel's National Institute for Testing and Evaluation) formulated a policy of dealing with suspected cheating and has operated accordingly for the last 30 years.

When suspected test misconduct is detected, NITE suspends the score in question and offers the examinee to take a retest in a controlled environment. A decision vis-à-vis the original score is made based on the retest score. In the years 2010-2014, 1,692 scores were suspended. 40% of the examinees did not show up for the retest and their scores were cancelled, 31% of the scores were released after the retest and 29% were cancelled.

In order to assess the decisions made, statistical analyses were performed on the scores of the retested examinees and on a simulation sample. The results suggest that the group of examinees whose scores were cancelled is considerably different from the group of examinees whose scores were released. This finding supports the validity of the process.

Presentation 21

Using Test Results with Identical Answers to Obtain Credible Evidence of Test Security Breaches

Dennis Maynes

Answer-copying and similarity statistics have been studied by researchers for several years with the intent of detecting potential collusion among test takers. These statistics have been used both to confirm allegations of test-taker cheating and to detect potential situations of test-taker cheating. A special case of situations involving collusion is that of test results with identical answers ("identical tests"). This special case is interesting because identical tests are usually only observed at very high score levels where existing answer-copying and similarity statistics have low power (i.e., not able to detect the security problem). Additionally, these situations are symptomatic of egregious test security breaches in which the exam content has been disclosed with a nearly perfect answer key to a large number of test takers. The presented research demonstrates the probability analysis of identical tests and an approach for obtaining credible evidence that test security has been breached from the analysis of identical tests. Practical implementation of the approach is shown using two cases studies from actual data. The theoretical efficacy of the approach is demonstrated through simulations.

Presentation 22

A Test Security Framework

Rachel Schoenig & Jamie Mulkey

Reliable and portable credentials will serve a critical role in closing the skills gap. Efforts underway across the US are providing a structure for evaluating portable workforce skills credentials. To ensure reliability, there is a need for information about the level of security incorporated into a credential. Because security isn't a "one size fits all" proposition, a test security framework can help drive greater transparency and, ultimately, consumer trust in portable credentials and in the credentialing industry. Join us for a discussion of why a test security framework is needed, the benefits to examinees and credential users, how the framework is being developed, and how you can get involved!

Presentation 23

Bayesian Detection of Cheating on Tests

Dr. Wim J. van der Linden

A Bayesian approach to the detection of cheating on tests has several advantages relative to classical statistical hypothesis testing. First, it is based on the correct probability distribution of the number of items on which the test taker has cheated given his observed number of aberrant responses. Second, whereas classical hypothesis testing only allows us to control its Type I error, the Bayesian approach does allow us to directly account for the incidence of cheating in the population of test takers. Third, the approach resolves the problem of whether or not to condition on the responses by the source in the detection of answer copying, which has plagued the literature since Frary et al. (1977). Fourth, it automatically accounts for the presence of estimation error in any of the parameters of the psychometric model (e.g., ability parameters). A natural Bayesian way of presenting evidence of cheating is through reporting of its posterior odds given the responses observed for the test taker. In this presentation we will show the odds for four different types of cheating: item pre-knowledge, item harvesting, answer copying, and fraudulent erasures on answer sheets. For each of these types of cheating, the odds can be calculated using a simple, extremely fast algorithm known as the Lord-Wingersky algorithm in test theory. The only difference exists in the parameters that need to be fed into the algorithm.

Presentation 24

Extension & Cross-Validation of a Profile of Statistical Indices for Detection of Aberrant Test Responses

Greg Hurtz & John Weiner

Many statistical indices have been developed over nearly 100 years of research, with many being variations or modifications of early indices but some having more distinct foundations that are sensitive to different data patterns. For example, many indices are measures of similarity in raw item responses consistent with answer-copying, while others are more general person-fit type measures that detect aberrances in item scores. No single index can effectively detect and diagnose all patterns of aberrant responding, so a profile of indices that are differentially sensitive to different patterns is expected to be more effective as a multi-faceted evaluation process than any single index alone. The presenters will describe recent research developing a profile of indices for detecting two patterns of cheating (pre-knowledge and answer copying) where discriminant function analysis was used to explore unique linear combinations of indices that maximally differentiate groups of test-takers, for whom their actual responses were manipulated in accordance with the two cheating patterns. They will then present new research extending the results to detecting additional aberrant patterns (e.g., randomness, carelessness), and present cross-validation results for the profile across multiple tests and test forms. Strategies for applying the discriminant function results in practice will be discussed.

Response Time Based Non-Parametric Person-Fit Index for Aberrant Responding Behavior Detection in Large-scale Assessment

Kaiwen Man & Hong Jiao

Response time has been widely used lately in measuring and evaluating students' latent characteristics, including but not limited to intelligence, personality, knowledge, skill, and ability. In particular, response time can be used to identify aberrant behaviors in educational and psychological tests. Statistical procedures based on response time can help us achieve this goal by creating a more flexible framework (i.e., discordant pattern analysis) compared to the traditional item response-based modeling within the item response theory framework. The purpose of this research is to propose a non- parametric index based on response time only in computer-based testing to identify aberrant responding behavior that might indicate cheating, such as pre-knowledge or non-cheating behaviors like warm-up effects.

Presentation 25

The Efficacy of Using Item Response Latency Statistics to Detect Pre-Knowledge

Sarah Thomas

The detection of examinees who accessed proprietary test content before taking an exam, called examinees with pre-knowledge, is a top priority in data forensics. Some authors (e.g. van der Linden, 2006) have proposed using item response latencies (i.e. the time taken to answer test questions) to detect cheating. However, this research has been hampered by a lack of data containing credible information about which examinees are suspected cheaters. In the current study, we used item response latency statistics to identify examinees who are suspected to have pre-knowledge, based on their responses to Trojan Horse (TH) items. TH items are easy, unscored items that are published with an erroneous answer key. Using the erroneous answer key to answer the TH items demonstrates a lack of content knowledge and probable pre-knowledge. Groups of TH scores were predicted using examinees working speed, variability and skewness in response times, the correlations between examinees, item response times and the expected item response times, and combinations of these predictors in a logistic regression analysis. The results will be discussed in terms of the overall performance of the model, the most effective predictors, and the implications of the findings for data forensics investigations.

THANK YOU

FOR ATTENDING THE 2016 CONFERENCE ON TEST SECURITY

Hosted by

Pearson