ATP 2015

Innovations in Testing

Palm Springs, CA

March 1-4

Westin Mission Hills Golf Resort & Spa

Palm Springs, CA

INSIGHT happens here

#atpconf

Schedule at a Glance / Table of Contents

SUNDAY - March 1st

10:00AM – 6:00 PM	Speaker Lounge Open Mission Hills Boardroom
7:15 AM – 12:00 PM	Golf Tournament
2:00 PM – 7:30 PM	Registration Open
5:00 PM – 6:00 PM	First-time Attendee Reception and Remarks Oasis Den
6:00 PM – 7:30 PM	Innovations in Testing Opening Reception Masters Plaza

MONDAY - March 2nd

7:30 AM – 7:00 PM	Registration Open
7:30 AM – 6:00 PM	Speaker Lounge Open Mission Hills Boardroom
7:30 AM – 8:30 AM	Breakfast with Exhibitors Ambassador Ballroom
7:30 AM – 8:15 AM	European Regional Division Breakfast Meeting – All are Welcome! East Patio
7:30 AM – 8:15 AM	Workforce Practice Area Division Breakfast Meeting – All are Welcome! Polo Room
7:30 AM – 8:15 AM	Session Coordinator Breakfast Meeting Celebrity Ballroom C
8:30 AM – 10:00AM	Opening Keynote Celebrity Ballroom D-E
10:15 AM – 11:15 AM	Breakout Sessions
11:30 AM – 12:30 PM	Breakout Sessions
12:30 PM – 1:45 PM	Lunch with Exhibitors Ambassador Ballroom
12:30 PM – 1:30 PM	Education Practice Area Division Luncheon – All are Welcome! Oasis Den
12:30 PM – 1:30 PM	Asia Regional Division Luncheon – All are Welcome! Hospitality Suite 1019
1:45 PM – 2:45 PM	Breakout Sessions
3:00 PM – 4:00 PM	Dessert with Exhibitors Ambassador Ballroom
4:00 PM – 5:00 PM	Breakout Sessions
5:00 PM – 7:00 PM	Reception with Exhibitors and Innovation Showcases Ambassador Ballroom

TUESDAY - March 3rd

7:30 AM – 5:15 PM	Registration Open
7:30 AM – 6:00 PM	Speaker Lounge Open Mission Hills Boardroom
7:30 AM – 8:30 AM	Breakfast with Exhibitors Ambassador Ballroom

TUESDAY - March 3rd (cont'd)

7:30 AM – 8:15 AM	Security Committee Breakfast Meeting – All are Welcome! Oasis Den
7:30 AM – 8:15 AM	Clinical Practice Area Division Breakfast – All are Welcome! Celebrity Ballroom A
8:30 AM – 10:00 AM	Workshops
10:00AM – 10:30 AM	Break with Exhibitors Ambassador Ballroom
10:30AM – 12:00 PM	Workshops & Product Demos
12:00 PM – 1:00 PM	Certification/Licensure Practice Area Division Luncheon – All are Welcome! Oasis Den
12:00 PM – 1:15 PM	Lunch with Exhibitors Ambassador Ballroom
1:15 PM – 2:15 PM	Breakout Sessions
2:30 PM – 3:15 PM	Ignite Sessions & Exhibits Open
3:30 PM – 4:15 PM	Dessert with Exhibitors & Roundtable Discussions Ambassador Ballroom
4:15 PM – 5:15 PM	Breakout Sessions
5:30 PM – 6:30 PM	Sponsor Thank You and Debrief Reception Oasis Den
6:00 PM – 7:00 PM	Industrial/Organizational Practice Area Division Reception – All are Welcome! Hospitality Suite 1419

WEDNESDAY - March 4th

7:30 AM – 11:00 AM	Speaker Lounge Open Mission Hills Boardroom
8:00 AM – 12:30 PM	Registration Open
8:00 AM – 10:00AM	Breakfast with Closing Keynote Celebrity Ballroom D-E
10:15 AM – 11:15 AM	Breakout Sessions
11:30 AM – 12:30 PM	Breakout Sessions

TABLE OF CONTENTS

Welcome . . .

Dear Colleague,

On behalf of the Association of Test Publishers (ATP) Board of Directors and the Conference Planning Committee, welcome to ATP's **16th annual Innovations in Testing Conference!**

As a global industry leader, ATP's reputation continues to spread. This year we welcome delegates from **over 15 countries** – from North America to Europe, the Middle East and Asia. Your Innovations Conference is reaching new heights because delegates eagerly participate as speakers and volunteers. **Records have been broken.** Your willingness to share your knowledge and experience with one another makes our time together a true learning opportunity.

Since we last met, the Conference Planning Committee has been committed to identifying and providing **new features** to the conference. This year you can experience 16 Workshops, 93 Breakouts, 7 Ignite Sessions, 8 Product Demonstrations (New!), 12 Innovative Showcases, and 7 Fishbowls. We are very proud to announce that we have chosen to **Webcast** the Opening and Closing Keynote, 16 Breakout Sessions and Ignite Presentations. And we are excited to premiere the **Innovations Hackathon** (Shark Tank) contest where 6 teams compete through developing their best certification program.

The conference is complete with many **networking opportunities**, starting with the First Time Attendees Reception and the Welcome Reception on Sunday evening. As you enjoy all that the conference has to offer, please join us in the **exhibit hall** and meet your friends, colleagues, partners and sponsors.

Make yourself comfortable with our **enhanced mobile app**; scan all sessions and create your custom schedule. Be sure to join us on Monday at 8:30 a.m. for our **Opening Keynote**. Dan Shapero from LinkedIn will share his insights and predictions to help attendees anticipate emerging trends. Don't miss our **Closing Keynote** on Wednesday morning with Constance Steinkuehler. Constance will review the current landscape of games for learning and assessment, highlighting recent innovations in assessment design and delivery.

ATP truly is the **"intelligent voice for testing"**. Due to our sponsors' and loyal members' continued financial support and dedication, the Innovations Conference is once again on track to be a huge success. Thank you for participating in this conference. To make sure that we stay on track, your **feedback** is encouraged and appreciated. Your suggestions and insights drive us to brighter innovations.

Welcome to Palm Springs!

Best regards,

Douglas F. Becker, PhD
HMH Riverside
ATP Board Chair, 2015

Erwin van Schaffelaar
Cito
Innovations in Testing Conference Chair, 2015

Got Test Security?

The ATP Security Committee provides a forum to encourage assessment organizations to collaborate in addressing test security concerns.

We strive to:

- **Identify methods to improve test security;**
- **Establish and disseminate security best practices; and**
- **Protect the integrity of assessments and assessment programs.**

This is a list of security related sessions that you don't want to miss!

MONDAY, MARCH 2ND, 2015

10:15 AM - 11:15 AM
- We Are Not the NSA: How to Factor Privacy Into Your Security Program

11:30 AM - 12:30 PM
- PII Data is an Important Business Asset. How Well Are You Protecting It?
- Locking it Down: The Key to Test Readiness and Security

1:45 PM - 2:45 PM
- The Incredible Journey…to Trustworthy Assessment Results: A Panel Discussion of Best Practices for Using Statistics to Invalidate Scores
- Including Special Populations in Test Security Policies and Procedures

4:00 PM - 5:00 PM
- Security at the Forefront

TUESDAY, MARCH 3RD, 2015

7:30 AM - 8:15 AM
- Security Committee Breakfast Meeting – All are Welcome!

8:30 AM - 10:00 AM
- Exam Security Incident Response Workshop
- Trust But Verify: Exam Integrity Audits Made Easy Through The Adoption of Technology

10:30 AM - 12:00 PM
- Test Security on a Dime: Tips and Tools for a Cost-Effective Test Security Program
- We're Not in Kansas Anymore: Supporting International Expansion with Best Known Security Controls

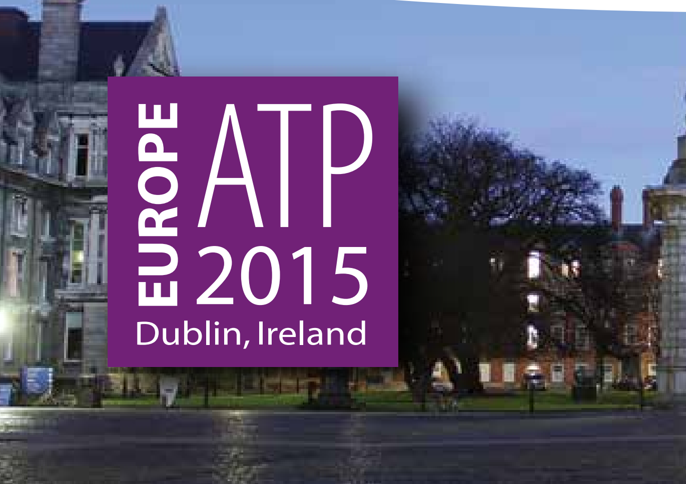
1:15 PM - 2:15 PM
- Maximizing Exam Security through & Susceptibility of Various Item Types
- Test Security Tools and Measuring Their Effectiveness

4:15 PM - 5:15 PM
- New Security Trends in High Stakes International Examinations: Communicating with Regulators

WEDNESDAY, MARCH 4TH, 2015

10:15 AM - 11:15 AM
- Security and Transparency – Do You Know What You Purchased? Do They Know What They Sold You?
- Using Secure Item and Test Designs to Protect Your Tests

11:30 AM - 12:30 PM
- Working a Case: Best Practices in Conducting Exam Integrity Investigations



SAVE THE DATE!

E-ATP 2015 Annual Conference
23-25 September 2015
Dublin, Ireland

Call for Papers Now Open!

www.eatpconference.org

Association of Test Publishers (ATP)

Promoting and preserving the general welfare of testing and its value to society in all its forms and uses since 1992.

ATP is a global, nonprofit organization representing providers of tests and assessment tools and/or services related to assessment for clinical, occupational, certification, licensure, educational, workforce skills credentialing or other similar professional measurement uses. ATP members are the leading publishers and assessment service providers in today's testing industry who are committed to encouraging international strategies which advance the position of the industry, its technology and the science that supports it.

ASSOCIATION OF TEST PUBLISHERS
MAKING THE WORLD SMALLER FOR TESTING

The vision of the Association of Test Publishers (ATP) is to become the "International Voice of Testing." Although the ATP was founded as a trade association in North America in 1992, testing and assessment have been developing into a global community at a rapid pace. Consequently, the ATP expanded by adding a Europe Division (2005), an Asia Division (2011), and a founders' group has met to organize an India Division (2012). Accordingly, the ATP is now poised to become recognized as the global trade organization for test publishers, thus making the world smaller for testing. Such expansion is appropriate, inasmuch as the language of testing and psychometrics is international and many issues faced by publishers are the same all over the world. Members of the ATP are truly part of a global community of testing, benefiting from the coordinated efforts of a strong organization that offers education, networking, advocacy, advertising, and industry advancement/protection activities on behalf of publishers worldwide.

ATP Regional Divisions

The Association of Test Publishers' membership is composed of the leading publishers and assessment service providers in today's testing industry. Presently there are four Regional Divisions that are active as subgroups of ATP's global membership: Asia, India, Europe and North America. Regional Divisions provide added networking opportunities for members and provide a forum for the unique regional needs of assessment professionals in different geographic areas.

Asia Regional Division

Asia-ATP (A-ATP) was formed to meet the local needs and interests of the Asia test publishing and test delivery community and to ensure that the Asian test publishing community participates in ATP activities that have global importance and relevance. A-ATP seeks to join together with ATP's other Regional Divisions to form a strong foundation with a common global position and policy on testing issues such as intellectual property protection, privacy, best practices for all forms of testing, cultural adaptation of tests, interoperability and innovative testing methods and procedures.

Europe Regional Division

Europe-ATP (E-ATP) is the Regional Division that brings together European test publishers and related organisations for the purposes of networking, sharing innovative ideas and creating both educational and marketing opportunities related to testing and assessment in Europe.

India Regional Division

India-ATP (I-ATP) is the regional Division that provides a forum focused on positively influencing consumers and government agencies in India to establish policies and guidelines that are advantageous to all of the stakeholders in the testing process. Much like ATP's other Divisions and Regions, I-ATP focuses on advocacy and education for the test publishing/ assessment services industry in India.

North America Regional Division

The North American Regional Division is ATP's oldest founding Division. Established in 1992 as a North American trade association, ATP brought together the leading test publishers and assessment service providers working in the areas of Industrial/Organizational, Certification/ Licensure, Clinical and Educational testing. In 2000 ATP launched the Association's flag ship conference, Innovations in Testing. ATP continues to expand its global mission to promote and preserve the general welfare of testing and its value to society, in all its forms and uses.

Do you want to learn more at ATP? Visit us in the Ambassador Foyer!

#atpconf

We’re Mobile!

Download the Conference Guidebook App

Use your smart phone or your tablet to access the program, maps and exhibitor information throughout the conference.

1. Simply scan the QR code here:

2. Search for “Guidebook” in your App store, or go to www.guidebook.com/getit to download the Guidebook app.

3. Once downloaded, search “Innovations in Testing 2015.”

Once you have downloaded the app...

My Schedule: Personalize your conference experience by adding sessions and events to the ‘My Schedule’ feature.

Sponsors: Browse or search the exhibitor list for general information, booth numbers, and link to an exhibitor’s website to learn more.

Maps: Interactive hotel map helps you find that intriguing discussion or fun networking event quicker and easier.

To-Do: Create your own To-Do List to ensure you accomplish your personal and professional goals at Innovations.

Special Thanks to the 2015 Conference Committees

Executive Committee

Board Chair
Douglas F. Becker

Board Past Chair
Eugene Burke

Conference Chair
Erwin van Schaffelaar

Conference Vice Chair
Eric D'Astolfo

Association of Test Publishers, CEO
William G. Harris

Association of Test Publishers, COO
Lauren Scheib

Program Chair
Ashok Sarathy

Program Vice Chair
Rachel Schoenig

Marketing Chair
Kelly Goeke

Marketing Vice Chair
Emily Kurze

Certification/Licensure Chair
Manny Straehle

Clinical Chair
Henrik Skovdahl Hansen

Education Chair
Gerald Melican

Industrial/Organizational Chair
Gary Behrens

Workforce Skills Credentialing
Dan Rinn

European Chair
Matthew Poyiadgi

Asia Chair
Alex Tong

Security Chair
Rory McCorkle

Marketing Committee

Marketing Chair
Kelly Goeke

Marketing Vice Chair
Emily Kurze

Members
Liz Burns
Vita Greco
Jill Liles
Jamie Rosa

Program Committee

Program Chair
Ashok Sarathy

Program Vice Chair
Rachel Schoenig

Conference Chair
Erwin van Schaffelaar

Conference Vice Chair
Eric D'Astolfo

Certification/Licensure Vice Chair
Beth Kalinowski

Clinical Chair
Henrik Skovdahl Hansen

Education Vice Chair
Jesse Markow

Industrial/Organizational Vice Chair
Cicek Svensson

Security Vice Chair
Chuck Friedman

European Vice Chair
Cicek Svensson

Members/Reviewers

Linda Andries Houghton Mifflin Harcourt

Kathy Angelucci National Board Of Medical Examiners (NBME)

Terry Barham SAS

Douglas F. Becker HMH Riverside

Kirk Becker Pearson

Julie Boyce Pearson VUE

Jim Brinton Certification Management Services (CMS)

Mary Browne National Board of Osteopathic Medical Examiners (NBOME)

Belinda Brunner Pearson VUE

Matthew Burke The College Board

Liz Burns NetApp, Inc.

Jill Burroughs Alpine Testing Solutions

Nathalie Collautt Project Management Institute

Maggie Collins Human Resources Research Organization (HumRRO)

Matt Comber Conference of State Bank Supervisors

Susan Cooley Alpine Testing Solutions

Drew Dallas National Commission on Certification of Physician Assistants (NCCPA)

Eric D'Astolfo Pearson VUE

Ingrid de Vries Medical Council of Canada

Philip Dickison The National Council State Boards of Nursing

Holly Dickson Microsoft Corporation

John Dight Kryterion

Merri Dillinger Adecco Group North America

Cathy Donath The Donath Group

Ken Doucet Professional Exam Service

Jeremy Dugosh American Board of Internal Medicine

Nikki Eatchel Scantron

Lisa Ehrlich Measured Progress

David Elfassy Netlogon

Henrietta Eve American Institute of Certified Public Accountants (AICPA)

Daniel Eyob Graduate Management Admission Council (GMAC)

Sean Fahey VidCruiter

Deborah Ford CPSRH

Jeff Foster Hogan Assessment Systems

Blaine Gaddis Hogan Assessment Systems

Holly Garner Ascend Learning

Jennifer Geraets ACT, Inc.

Dawn Gibas ASA/SSSA

Stella Gibbs Pacific Metrics Corporation

Kristin Gibson IBM

Kirsten Gobeski Institute for Personality and Ability Testing

Kelly Goeke Pearson VUE

Jerry Gorham Ascend Learning

Maurilo Gorito Riverbed Technology

George Gray NetApp, Inc.

Vita Greco Professional Examination Service

Sandra Greenburg Professional Examination Service

Cabell Greenwood Internet Testing Systems

O'Neal Hampton Pearson VUE

Jennifer Harvel CPS HR Consulting

Katherine Heppner Pearson VUE

Jim Hogan ACT, Inc.

Beth Holst Holst & Associates, LLC

Deb Houghton Pearson VUE

Julie Huber Teradata Corporation

Crystal Hughes The Booth Company

Peg Jobst Graduate Management Admission Council (GMAC)

Erika Johnson Caveon

Vishal Kapoor Pearson

Siddiq Kassam International Credentialing Associates (ICA)

Chad Kinart Board of Certification, Inc. (Athletic Training)

Reid Klion pan - Performance Assessment Network, Inc.

Brenda Knowles Pearson VUE

Elizabeth Langston ARDMS

Harry Layman The College Board

Mary Lehnard Alpine Testing Solutions

Daniel Lewis McGraw-Hill Education/CTB

Vincent Lima Professional Testing, Inc.

William Lorié Questar Assessment, Inc.

Kari Lucio Flaskerud Sales Force

Peter Mackey CFA Institute

Marita MacMahon ACER

Dennis Mahoney The Wharton School, University of Pennsylvania

Carl Matheson Australian Medical Council

Joy Matthews-Lopez National Association of Boards of Pharmacy

Shannon Matzke Questar Assessment, Inc.

Jennifer McGlashon Kryterion

Jeff McGregor Strasz Assessment Systems

David Mott Tests for Higher Standards

Canda Mueller Questar Assessment, Inc.

Lorin Mueller The Federation of State Boards of Physical Therapy

Jacqui Mulligan CPA Ontario

Liberty Munson Microsoft

Kristin O'Brien National Board of Osteopathic Medical Examiners (NBOME)

Stefani Okamoto Microsoft

Joe Orban Questar Assessment, Inc.

Corina Owens Professional Testing

Steven Peluso Steven Peluso Law Office

Michelle Porter International Code Council

Clarke Porter Pearson VUE

Mary Jane Prendergast Loyalist Exam Services

Ruth Ramstad Pearson VUE

Sandi Resnick Micrsoft

Shungwon Ro IBM

Lawrence Rudner LMP Associates

Sara Rupp CompTIA

Lindsay Sanders NWEA

Ashok Sarathy Graduate Management Admission Council (GMAC)

Judd Sather Data Recognition Corporation

Stacy Saul Association for Financial Professionals

Jarin Schmidt Pearson VUE

Jason Schwartz Pearson VUE

Matt Scicchitano Teradata

Sonya Sedivy University of Wisconsin

Jennifer Semko Baker & McKenzie LLP

Linjun Shen American Board of Psychiatry and Neurology

Bruce Shotts National Board of Chiropractic Examiners

Rachel Siegel CFA Institute

Russell Smith Alpine Testing Solutions

Hao Song National Board of Osteopathic Medical Examiners (NBOME)

Michelle Sparacino Graduate Management Admission Council (GMAC)

Myisha Stokes Alpine Testing Solutions

Michael Tack BDO

Stephanie Terry ACT

Beinvenda Thorne ARDMS

Jill van den Heuvel Alpine Testing Solutions

Joan Van Tol Law School Admission Council (LSAC)

Timothy Vansickle Questar Assessment, Inc.

Veronica Vele ACER

Desley Ward CPA Australia

Linda Waters Prometric

Patrick Watts CFA Institute

Lynn Webb Consultant

John Weiner PSI Services LLC

Marc Weinstein Dilworth Paxson LLP

Bob Whorton National Council of Examiners for Engineering and Surveying (NCEES)

William Wilkins Kryterion

Brodie Wise Internet Testing Systems

Cynthia Woodley Professional Testing, Inc.

Heather Young Pearson VUE

Iva Zemanovi A Systems Incorporated

Platinum Sponsors

Alpine Testing Solutions

ata

cito now you know

CMS Certification Management Services

CollegeBoard

ETS®

GMAC® GRADUATE MANAGEMENT ADMISSION COUNCIL

iTS internet TESTING SYSTEMS

KRYTERION™ GLOBAL TESTING SOLUTIONS

PEARSON

Professional Examination Service™ Credentialing Insight

Professional Testing

PROMETRIC®

psi

Yardstick Testing and Training Experts

Gold Sponsors

ASSESSMENT SYSTEMS
For Good Measure

ACT®

AMP™

ANSI
American National Standards Institute

BrightLink

Castle
Design.
Develop.
Deliver

caveon™
Test Security

comira
Test us.
www.comiratestina.com

Commsmultilingual

CERTNT

Copyright Clearance Center

Cyveillance
a QinetiQ Company

enlightKS

examity®
BETTER TEST INTEGRITY.

EXCELSOFT

Heuristic
SOLUTIONS

GENERAL DYNAMICS
Information Technology

LearningBuilder™

Gold Sponsors

INNOVATIVE EXAMS
SECURE | ONLINE | PROCTORING

ICE Institute for Credentialing Excellence™

International Credentialing Associates

Learnosity

NPC® Integrated Print & Digital Solutions

oat home of tao

PAN Performance · Assessment · Network

ProctorU

Software SECURE

strasz ASSESSMENT SYSTEMS

Questionmark

uCertify

WIRIS™ math & science

ZOOMORPHIX SYSTEMS

Silver Sponsors

Center for Applied AxioMetrics
tracking competence carefully and concisely

BUROS
CENTER FOR TESTING

Dilworth Paxson LLP

GRANDMASTERS

Houghton Mifflin Harcourt

Loyalist
EXAM SERVICES
FLEXIBLE • SECURE • TRUSTED

LSAC®

measured progress™

Mountain Measurement, Inc.

netlogon CONTENT DEVELOPMENT

NOCTI®

PACIFIC METRICS CORPORATION

Paradigm TESTING

Conference Program Overview

#atpconf

Innovations in Testing 2015 is full of exciting events and opportunities to expand your knowledge of the testing industry and interact with colleagues. Following is a brief overview of the program guide, events at the conference and ways to take full advantage of your time here.

First Time at Innovations in Testing?

You won't want to miss the First Time Attendee Reception on Sunday from 5:00 – 6:00 PM in the Oasis Den, immediately before the opening reception. You will hear a brief overview of ATP and the conference, and get the chance to meet other first timers as well as conference veterans.

An Abundance of Networking Opportunities

Don't be shy! This is your chance to make lasting connections and learn from your colleagues. Interacting with other assessment professionals is often cited as the #1 reason people attend Innovations in Testing. We have ample ways for you to get connected:

- Receptions allow you to mingle with other assessment professionals in a casual setting.
- The exhibit hall is open on Monday and Tuesday. The hall is filled with assessment-related companies and experts in testing ready to talk with you.
- Breakfast and lunch are great times to introduce yourself to people you don't know and connect with them.
- Go online! We are promoting social networking more than ever as a way to learn from each other and stay connected throughout the year. During and after the conference, check out our discussion group on LinkedIn, Twitter and YouTube. See the ad in this program book for specific links.

Informative Educational Sessions

Once again, we had a record number of session submissions this year, resulting in a fantastic selection. In addition to the main general sessions, you will see that we offer four types of education sessions – Workshops, Breakout Sessions, Fishbowls, Innovation Showcases, Ignite Sessions and (NEW) Product Demonstrations.

- What is the difference between Workshops and Breakouts? Workshops are longer and designed to go a little deeper or broader on the subject matter. They are also more interactive, engaging the audience for participation and learning. Breakouts are content rich but shorter and with a more narrow focus.
- Who leads them? Experts in the testing industry. The program has been designed from the ground up to provide expert speakers across a wide range of testing related topics.
- How were they selected? All sessions were selected through a rigorous blind review process, designed to ensure relevance, breadth of topics, expertise of the speakers and clear outcomes.

General Sessions

The conference includes two all-conference general sessions. On both Monday and Wednesday, there are exciting keynote speakers. Check the Monday morning and Wednesday morning program pages within this book for full details. Don't miss them!

Ignite Sessions encapsulate the Ignite motto "enlighten us, but make it quick." Each talk is exactly five minutes long. Ignite presenters share their personal and professional passions, using 20 slides that auto-advance every 15 seconds. These will take place on Tuesday from 2:30 PM – 3:15 PM in the main General Session Room (Celebrity D-E). See page 84 for details.

Fishbowl Discussions are moderated open discussions on various topics that will allow you to learn in a format best for you – whether you like to contribute to the conversation, or listen to the dialogue and learn from the viewpoints of your colleagues. Each Fishbowl discussion will take place during breakout timeslots on Monday and Tuesday.

Product Demonstrations are computer-based demonstrations or presentations showcasing the latest technologies, products, services and/or solutions within the testing industry. Multiple demonstrations will take place during the Tuesday, 10:30 AM Workshop timeslot.

NEW!

Innovation Showcases are your chance for one-on-one interaction with a wide variety of presenters while you enjoy a cocktail. Each Showcase is a chance for presenters to highlight a particular study, product, technology, or approach in an individual manner, sometimes using interactive media. We truly have a vast array of fascinating topics, specifically designated as Innovation Showcases to allow attendees to dig deep with the presenters in their areas of expertise. See page 57 for details.

Conference Program Overview

#atpconf

Conference Topics

We continue to offer world-class content that you have come to expect at the Innovations in Testing conference. The 2015 program offers attendees several choices of session types and subject matters, giving you more choices and more exposure to the content that matters most to you. Whatever your interest and learning style, the Innovations in Testing conference is the place to learn, share and grow. The conference topics are as follows:

- **Program Design**
- **Legislation, Policy and Accessibility**
- **Testing, Measurement, and Psychometrics**
- **Security**
- **Data Management**
- **Business Strategy and Operations**
- **Test Administration and Delivery Models**

FEATURED SPEAKER SERIES

Various sessions are highlighted within this program book as part of a Featured Speaker Series. These speakers have been hand-picked and invited to present on industry topics because they are experts in their field.

FUNDAMENTALS OF TESTING

Sessions denoted within a Fundamentals of Testing box are appropriate to attendees who are new in the testing industry.

Get to Know Our Sponsors – Industry Experts

This event wouldn't happen without our sponsors. They bring many industry experts to the conference who have the most up-to-date information on innovative ways to develop, deliver, manage and support your assessment needs. Please get to know them by visiting the exhibit hall.

ATP Division Areas

To help you navigate the program and identify sessions that align with your key areas of interest, you will see each session specify its associated division. ATP content related divisions include:

- Certification/Licensure
- Clinical
- Education
- Industrial/Organizational
- Workforce Skills Credentialing NEW!

Webcast Sessions

Sessions denoted with this icon throughout this program book will be webcast live. These sessions will also be recorded and available for conference attendees to view after the conference. All attendees will receive an email with details explaining how to access these recorded sessions.

Please Complete Surveys!

This conference is as good as it is because we listen to you! Make sure your voice is heard. You will be asked to fill out a survey for each session and for the conference as a whole. Our speakers and conference committee members are volunteers and greatly appreciate the feedback to help create even better sessions and conferences in the future...not to mention a nice pat on the back when appropriate!

More Information:

- **Visit the ATP booth in the exhibit hall.**
- **ATP Website: www.testpublishers.org**
- **Innovations in Testing Website: www.innovationsintesting.org**

These sites provide you with several ways to get connected, learn more about ATP, the conference, and what's happening in our industry.

Opening Day

SUNDAY, MARCH 1, 2015 • 8:00 AM – 7:30 PM

8:30 AM – 12:00 PM
Golf Tournament
Sponsored by General Dynamics Information Technology

Golfers should arrive by 7:15 AM for breakfast and instructions.

Location: Meet at the Driving Range

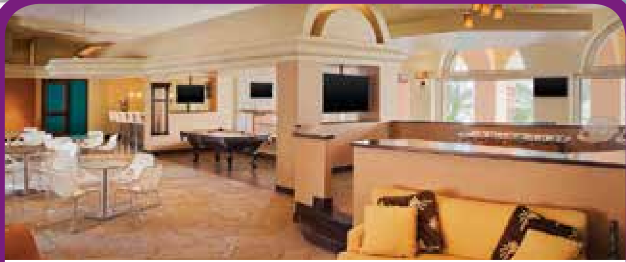
The 2015 conference will kick-off with a Golf Tournament sponsored by General Dynamics Information Technologies. Join us on the Pete Dye Golf Course for exceptional golf, a stunning desert scape and a beautiful mountain skyline. Designed for golfers at all stages, new and pros. This shotgun tournament is a perfect way to start the conference with networking, along with friendly competition.

Registration was required prior to the conference.

10:00 AM – 6:00 PM
Speaker Lounge Open

Room: Mission Hills Boardroom

Are you presenting at the Innovations in Testing conference? Stop in to practice your presentation, collaborate with co-presenters, or just rest for a while. This room is open to all conference presenters.



5:00 PM – 6:00 PM
First Time Attendee Reception

Location: Oasis Den

If you're new to the Innovations conference, this is a great chance to learn what to expect and how to navigate our many offerings. Make the most of your time here by joining us for this informative session and networking opportunity.

6:00 PM – 7:30 PM
Opening Reception

Location: Masters Plaza

Please join your colleagues in the testing industry and celebrate the opening of ATP's 16th Annual Conference. Don't miss this opportunity to meet up with old friends and make some new ones!

Registration / Breakfast / Keynote

MONDAY, MARCH 2, 2015 • 7:30 AM – 6:00 PM

7:30 AM – 6:00 PM

Speaker Lounge Open
Room: Mission Hills Boardroom

Are you presenting at the Innovations in Testing Conference? Stop in to practice your presentation, collaborate with co-presenters, or just rest for a while. This room is open to all conference presenters.

7:30 AM – 8:15 AM

European Regional Division Breakfast Meeting – All are Welcome!
Location: East Patio
(Next to the Ambassador Ballroom)

Join the European Division of ATP as we discuss the current plans and initiatives in place for 2015 and beyond. Bring your thoughts and ideas as new ideas are encouraged.

Workforce Practice Area Division Breakfast Meeting – All are Welcome!
Room: Polo Room

Join the newest division of ATP Workforce Practice for breakfast and see what they have to offer.

Session Coordinator Breakfast Meeting
Room: Celebrity Ballroom C

Join us for breakfast and learn your responsibilities as a session coordinator.

7:30 AM – 8:30 AM

Breakfast with Exhibitors
Room: Ambassador Ballroom

8:30 AM – 10:00 AM

Opening Keynote & Awards

Room: Celebrity Ballroom D-E

Dan Shapero
Head of Careers Products, LinkedIn

The Future Value of Certifications: Insights from LinkedIn's Data Trove

In a world in which the status of individuals, organizations and economies becomes increasingly represented digitally through massive amounts of online data, what does the future hold for the value of skills validation, credentials, and formal education? In this talk, Dan Shapero will share his insights and predictions to help attendees anticipate emerging trends.

Attendees will learn how big data can be used to show the impact and value of specific certifications. Key take ways include:

- How career path data already available online can be used to track and observe dynamics in education and the credentialing industry across economies and geographies.
- The prevalence of educational options such as micro badges and certifications relative to a traditional university degree.
- Examples of how credentialing organizations and schools can leverage the power of data analytics to gain a new understanding of the value of their services.

Dan Shapero has held various leadership roles in sales, operations, and product management at LinkedIn since he joined in 2008. Dan is now Head of Careers Products at LinkedIn, where he oversees how LinkedIn helps professionals navigate their career and get their dream job by leveraging LinkedIn. Prior to LinkedIn, Dan was a Manager at Bain & Company, a management consultancy. Dan holds an MBA from Harvard Business School and a BS in Applied Mathematics from Johns Hopkins University.

REMARKS BY:

Douglas F. Becker *ATP Board Chair*
Erwin van Schaffelaar *Innovations in Testing Conference Chair*

Achievement Awards

MONDAY, MARCH 2, 2015 • 8:30 AM – 10:00 AM

2015 ATP Career Achievement Award and the 2015 Professional Contributions and Service to Testing Award

Don't miss this opportunity to join us to recognize and celebrate two leaders within the testing industry.

2015 CAREER ACHIEVEMENT AWARD

Presented to Betty Bergstrom, PhD

Dr. Bergstrom has been part of the testing community for more than twenty-five years. She began her career working with testing companies and certification organizations. Since 2006, she has worked for Pearson VUE where she served as Vice President of Testing Services, managing global teams including Measurement and Research, Content Development and Test Publishing.

Dr. Bergstrom holds a PhD in Measurement, Evaluation and Statistical Analysis from the University of Chicago. She has research expertise in the areas of computer-based testing, computerized adaptive testing, innovative item types, item response theory, and large-scale item banking. She is the author of numerous articles related to computer-based testing. Dr. Bergstrom is a member of the American Educational Research Association (AERA) and National Council on Measurement in Education (NCME). She served as an ATP Board member from 2007-2010.

2015 PROFESSIONAL CONTRIBUTIONS AND SERVICE TO TESTING AWARD

Presented to Alan J. Thiemann

Alan has practiced in Washington, DC since graduating from the University of Texas School of Law in 1975. Although he has been in private practice since 1987, he has spent time as a government lawyer and as in-house counsel to several national trade associations. He currently is the founder and owner of the Law Office of Alan J. Thiemann, located in Washington, DC.

Alan became Legislative Counsel to the Association of Personnel Test Publishers in 1988, which became the Association of Test Publishers (ATP) in 1992. He has represented the organization and its members on corporate/non-profit matters, state and federal legislation/regulations, intellectual property issues, and litigation involving psychological and educational testing. On ATP's behalf, Alan has handled cases involving educational measurement used in employment/personnel, education and training, certification/licensure, and clinical psychology, and he has served as advisor to the Test Security Committee and the Interoperability Study Group. During his tenure at ATP, Alan has served as counsel to the US Technical Advisory Group developing ISO 10667 (Delivery of Assessments in Occupational Settings). Since 2006, Alan has served as counsel/consultant to a joint effort between ATP and the Council of Chief State School Officers to develop and maintain a set of operational best practices for statewide K-12 large-scale assessment program (published in 2010 and 2013); this group is currently working on a related white paper dealing with how states/testing entities using cloud-based technology should deal with FERPA issues. Alan also serves as ATP's international counsel, assisting the organization in its globalization activities, including projects undertaken by the regional groups.

Alan is a frequent lecturer and has written extensively on business issues for various trade organizations, ranging from patent and trademarks, to Digital Copyright and Licensing, to labor/human resources, to the Americans with Disabilities Act.

Fishbowl Session

MONDAY, MARCH 2, 2015 • 10:15 AM – 11:15 AM

Leader2Leader 2.0: What are your Challenges in Running a Credentialing Business?

Room: Celebrity Ballroom A
Practice Area Division(s):
Topic: Business Strategy and Operations
Session Type: Fishbowl

Building on the success of a Leader2Leader session in 2014, this session provides a forum for certification leaders to explore the greatest challenges directly affecting them and their organizations. This fishbowl discussion will focus on the challenges of maintaining a quality credentialing program, while managing resource constraints. The questions posed by the moderator will focus on several topics:

- Growing program revenue, volumes, and engagement
- Managing expenses: staffing, budgetary, and other resource constraints
- Leveraging innovation and technology to stay ahead of the pace of change

The panelists for this session have managed certification programs, dealing with such concerns on a daily basis while expanding their programs in numbers and global reach. The moderator and participants will draw on their experience to discuss examples of these issues, such as taking an agile approach to keep up content changes, maintaining market share in a competitive environment, and taking effective development approaches while following proper processes.

This discussion will provide credentialing organizations with tactics to address such difficulties through the expertise of the panelists and learning from the other leaders sharing their expertise.

PRESENTERS:

Rory McCorkle *International Credentialing Associates (ICA)*

Dania Eter *Cisco Systems*

Jennifer Naughton *Association for Talent Development (ATD formerly ASTD)*

Breakout Session

MONDAY, MARCH 2, 2015 • 10:15 AM – 11:15 AM

Can Assessment Tools Survive the Rise of Video Recruiting?

Room: Celebrity Ballroom B
Practice Area Division(s):
Topic: Business Strategy and Operations
Session Type: Breakout

Assessment tools, such as personality scales, provide relatively low predictive validity (Hogan, ATP 2014) compared to assessment centers or structured interviews. However, they require minimal time and effort on the part of hiring managers who dislike the disruptions in productive work necessitated by recruiting and therefore assessment tools remain popular. More recently, video recruiting is allowing hiring managers to gain the benefits of a structured interview without the time and cost previously required. Can assessment tools survive this disruption?

Points covered will be:

- How asynchronous video recruiting is replacing phone interviews and the challenges and opportunities for the psychometric community.
- How video recruiting is bringing the structured interview process into organizations much earlier in the recruiting cycle.
- How adding rating scales and group raters to candidates asynchronous video answers is changing the recruiting landscape
- How very few video recruiting companies are adding quantitative analysis and the opportunities that exist in this market for the psychometric community
- How skill testing is now being incorporated in the recruiting process
- How skill testing can be administered over the cloud
- How organizations are using testing and assessment with video to enhance recruiting

PRESENTER:

Sean Fahey *VidCruiter*

FUNDAMENTALS OF TESTING

Accessibility 101

Room: Celebrity Ballroom C
Practice Area Division(s):
Topic: Legislation, Policy, and Accessibility
Session Type: Breakout

Accessibility is about fostering a testing experience for candidates so that the test measures what it is intended to measure, rather than measuring the effects of a person's disability. Many of the accessibility features of computer-based tests actually enhance accessibility for non-disabled test-takers as well, which is the essence of Universal Design. There are a number of practical considerations that must be kept in mind when developing tests and test-content so that accessibility is enhanced and test-takers with disabilities have access to the material. On the other hand, providing accommodations on tests may alter the validity of the test scores.

The legal obligations of the ADA must be balanced with the fundamental objectives of the test—a balance that can be difficult to achieve. This session will introduce you to the world of accessibility as it pertains to the psychometric and practical considerations of test development and test delivery. This session will focus on key areas to consider when innovating and developing certification exams with accessibility in mind, including a description and demonstration of commonly-used accessibility software tools that may enhance access for test-takers with disabilities, and the practical considerations of developing test content so that candidates can utilize this software. This session will include case examples of test publishers and testing organizations that have tackled these topics when designing their next-generation assessments. In addition, this session will include a discussion of psychometric concerns that may be raised about the validity of tests that are delivered with accommodations. Relevant research in this area will be discussed as well. Finally, the audience will be invited to discuss the prospect of balancing legal requirements with ensuring that the fundamental nature of the test is not altered.

PRESENTERS:

John Hosterman *GED Testing Service*

James Sandstrum *Assistive Technology Partners, University of Colorado*

Jill Moe *GED Testing Service*

Aligning the Content of an Adaptive Test with the Most Important Knowledges and Skills: How to Accomplish it and How to Measure it

Room: Celebrity Ballroom F
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Breakout

In every assessment situation, the test we use to measure a construct is only a sampling from a complex content domain. A question from test takers, measurement professionals, and other stakeholders is whether this sampling results in an assessment that is useful for the program's purposes. With an adaptive test, this question is more interesting, because each test taker sees individualized content, possibly changing the content alignment. This session discusses the basic process used to move from a job description or set of content standards to the content that is delivered on an adaptive test. It then proposes innovative metrics that might be used to evaluate content alignment.

Zara will detail how the issue manifests itself in a certification/licensure context. In particular, processes for going from a job analysis to the identification of the content to be tested are crucial to this process. These processes involve more art then science, as practical constraints influence decisions made concerning content. For certification and licensure, the approaches include the identification and translation of what an entry-level job holder needs to know and be able to do into a functional test.

Kingsbury will describe the issue as it manifests itself in an educational context. Examples using the common core content standards will detail processes for unpacking content standards to create an adaptive test that might have better alignment than a fixed-form test.

Wise will describe approaches to quantify the alignment of an adaptive test taken by a single individual or a group of individuals. He will also discuss how the adaptive test's ability to center on the individual might influence alignment. In addition, he will describe a family of statistics that might be used to evaluate alignment.

A discussion among presenters and attendees will investigate how different approaches to adaptive testing that might allow the investigation of the alignment of tests in complex situations that include different item styles, content areas, and the targeting of the test to the test taker. In particular, the discussion will identify processes that might be put into place to help adress the tradeoffs between test precision and content coverage within particular testing programs.

PRESENTERS:

Gage Kingsbury *Psychometric Consulting*

Anthony Zara *Pearson VUE*

Steve Wise *Northwest Evaluation Association*

We Are Not the NSA: How to Factor Privacy Into Your Security Program

Room: Celebrity Ballroom G
Practice Area Division(s):
Topic: Security
Session Type: Breakout

News reports about companies losing personally identifying information and identify theft are becoming daily occurrences in the media. As a result, the protection of one's personal information is a priority for individuals, including their exam data and testing experience. Certification programs need to think "privacy by design" when implementing their security plans. In today's exam world, unique scenarios such as online proctoring, protection of exam results, testing data use, and biometrics create an intersection between privacy and security that is raising new challenges for certification programs to consider and manage.

For example, what does an exam security team need to know when it comes to protecting their candidates' data? How do privacy laws and regulations affect a company's security plans and procedures? What special considerations must be in place for the testing of minors, especially on a global scale? And what does a company do when privacy and security conflict? In this session, we will discuss the areas of privacy, data protection, and security in relation to a testing security program, touching on three key topics:

- Implementing a privacy framework into a security program, and what upfront considerations that will strengthen security measures later;
- Identifying security practices that have privacy components, such as the use of biometrics, video surveillance, online/remote proctoring, testing of minors, and global testing;
- Maximizing investigation and enforcement measures while maintaining candidates' privacy expectations.

Good privacy practice starts with privacy by design principles; this session will help highlight some of the principles that clients and test sponsors need to know in order to frame their security programs while keeping privacy principles in mind.

PRESENTERS:

Kerri Davis *Microsoft*
Joe Brutsche *Pearson VUE*
Bryan Friess *Pearson VUE*

How Many Distractors Do You Need?

Room: Celebrity Ballroom H
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Breakout

CFA Institute is globally recognized as a leader in the training and assessment of investment professionals, and recognizes candidate accomplishment through the CFA charter designation. This designation is obtained through successful completion of three successive exams that have been developed and administered for over 50 years. World-wide candidate numbers approach 170,000 each year across the three exam levels. While the exams employ three different item formats across levels (i.e., selected-response, item sets, and constructed response formats), our focus is on recent format changes that we implemented with respect to selected response portions of the exams. We shifted all item development from four- to three-response items.

While the majority of the psychometric literature advocates for three-response alternatives, many organizations are reluctant to make the change to only two distractors. We propose to share our motivations and experiences in making this transition to provide a case study for other testing organizations to review and debate. We believe test developers overestimate their ability to write high quality fair distractors, and may also overestimate their ability to identify unfair distractors.

We will present a longitudinal perspective of exam psychometrics for three years leading into the change and five years since employing three-response alternatives. We will focus on observed changes in exam difficulty, discrimination, variability, non-functioning distractors, reliability, and construct validity as evaluated through latent variable modeling with focus on dimensionality and multi-group confirmatory factor analyses across groups that differ by geographic region. Results are generally consistent with an expectation of 'no harm done' in these quantitative measures. The benefits of our shift to three-choice items include better distractors, efficiencies in development processes, and improved resource utilization. The topic will be of interest to all exam developers and psychometricians.

PRESENTERS:

Peter Mackey *CFA Institute*
Timothy Konold *University of Virginia*

A How-To Guide to Successfully Managing a Portfolio of Credentials

Room: Mirage Room
Practice Area Division(s):
Topic: Business Strategy and Operations
Session Type: Breakout

More than one certifier has discovered the hard way that having one successful credential (even a spectacularly successful credential) is no guarantee that the next one (or the next) will match this success. Developing and managing a portfolio of credentials is more complex and challenging than tending to a single certification program. It requires a different focus, more in-depth analysis and more sophisticated business strategies. Certifiers that lack these abilities will likely find themselves over-extended, saddled with under-performing programs and legacy credentials and struggling to accomplish business goals.

Drawing on the experience and lessons learned of professionals who manage credential portfolios, we will answer the following essential questions:

- What are the critical considerations when determining whether to add a new credential - whether it's your first "new addition" or one of many?
- What impacts (both positive and negative) might the new credential have on a flagship credential and other credentials in the portfolio?
- What is the right "balance" of products for a credential portfolio?
- How can you maximize and optimize the total value of your credential portfolio?
- What branding and marketing issues should be taken into consideration when managing a credential portfolio?
- When is enough, enough – can you ever have too many credentials in your portfolio?
- How do you know when it's time to re-engineer or discontinue a credential?

This interactive session will provide veteran managers of credential portfolios with an opportunity to further expand their knowledge base and to share tips with colleagues. For those who are thinking about adding a new credential to their solo certification program, the session will provide a wealth of information to inform future decision making.

PRESENTERS:

Lenora Knapp *Knapp & Associates International*
Rahul Keshap *CFA Institute*

FUNDAMENTALS OF TESTING

Content Development Cohorts: The Role They Play in the Changing Face of Innovative, Interactive Item Development

Room: Oasis Ballroom 1
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Breakout

External content developers (CDs) can be used to broaden a resource base for high-quality item development. This requires connecting with skilled individuals whose content development expertise is utilized for the development of assessment items, particularly technology enhanced items (TEIs).

This session will cover best practices related to locating skilled individuals or companies (through associations, referrals, etc.), including how to qualify candidates by ensuring that prospects understand minimum requirements; requesting samples or remotely administering performance tasks; and vetting candidates to ensure that fundamental content knowledge requirements are met.

Various aspects we will cover include how to identify those CDs that possess skills best aligned to project goals; guidelines CDs need in order to supply viable items, particularly TEIs, while giving them the broadest leeway possible in areas of creativity and ingenuity; how allowing broad flexibility in creativity by experts provides engaging, thoughtful items that get to the core of the test taker's knowledge and understanding; the crucial role of feedback in minimizing rejected items when working with new or seasoned CDs; protection of intellectual property rights involving proprietary documentation and systems, as well as developed items; how communication with CDs not under assignment keeps those relationships active; how best to deal with issues such as dropping acceptance rates, plagiarism, or conflict of interest; the need for open communication in maintaining a successful cohort base where there is little, if any, face-to-face interaction.

PRESENTERS:

Lindsay Sanders *Northwest Evaluation Association*
Joanne Zimmer *Northwest Evaluation Association*
Michael Laus *Northwest Evaluation Association*

Online Proctored Testing: Navigating Uncharted Waters in High Stakes Assessment

Room: Oasis Ballroom 4
Practice Area Division(s):
Topic: Test Administration and Delivery Models
Session Type: Breakout

Methods for test delivery and management are rapidly evolving as technology continues to advance and become more cost-effective. Depending upon the purpose and stakes of the examination, a range of models is now available for deploying tests and addressing various threats to security. This session will examine an emerging technology-enabled approach, online proctored testing, and its applicability to certain high stakes testing scenarios.

The presenters will examine approaches, considerations and challenges in planning and adopting online proctored testing for different types of testing programs. In addition, empirical studies will be reviewed in which psychometric properties, data forensics properties, security risks, and candidate reactions are examined for a language interpreter oral certification test and end-of-course university exams that were administered under online and onsite proctored testing conditions. A questions-and-answers session will be conducted with audience members following the presentations.

PRESENTERS:

John Weiner *PSI Services LLC*

Ruben Garcia *Innovative Exams*

FEATURED SPEAKER SERIES

Crowdsourcing Leadership: Evaluating CEO Performance via Big Data and Public Digital Records

Room: Oasis Ballroom 5-6
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Breakout

Would you hire Steven Ballmer, or Silvio Berlusconi, without first giving them a psychometric test? This talk will examine how public digital records of CEO's reputation and big data analytics may improve our ability to predict leadership performance vis-a-vis established methods, such as valid psychometric tests. In line, it will discuss how recent innovations in talent identification (e.g., social media analytics, gamification, and big data) may impact our ability to assess leadership potential. For example, to what degree can we effectively evaluate CEO's emotional intelligence from their digital reputation? Can public employee engagement data be used to predict future CEO performance? And can CEO's online footprint predict future levels of employee engagement? Theoretical and practical implications, as well as ethical issues, will be discussed.

PRESENTER:

Tomas Chamorro-Premuzic *University College London*

Breakout Session

MONDAY, MARCH 2, 2015 • 10:15 AM – 11:15 AM

Implementing a Universal Optimal Test Assembly (OTA) System for Linear, Multi-Stage and Computer Adaptive Testing

Room: Oasis Ballroom 7

Practice Area Division(s):

Topic: Test Administration and Delivery Models

Session Type: Breakout

In this breakout session, we cover the aims and implementation of a revolutionary universal Optimal Test Assembly (OTA) system at McGraw-Hill Education CTB, showcase the system's ability to build flexible, accurate, and valid tests in its use of the shadow test technique, and demonstrate how OTA can be used to power a variety of test administration models in K-12, higher education, and certification testing.

The OTA system is built upon the theoretical research of linear models for optimal test design (van der Linden, 2005), which provides the unifying approach required for flexible, transparent, and optimal item selection regardless of delivery format. The OTA system includes two components: an Automated Test Assembly (ATA) application and a partner adaptive test engine (ShadowCATTM). The ATA application converts user input test requirements, test blueprints, and a test objective into a mixed integer programming (MIP) model. These models are submitted to the system to obtain an optimal set of items from an item bank, in the case of fixed forms. Alternatively, the models can be integrated into the ShadowCAT adaptive test engine for use in constructing shadow tests instantaneously during linear on-the-fly test (LOFT), multistage on-the-fly test (MSOFT) and item-level computer adaptive test (CAT) administrations. The ShadowCAT adaptive test engine differs from traditional CAT engines, in that at every step, ShadowCAT engine looks ahead and constructs a complete test, and then selects the optimal item from that fully compliant test.

Benefits of using the OTA system include guaranteed optimization (best collection of items from the item pool), guaranteed compliance with test blueprints and psychometric requirements, and the highest precision possible by maximizing test information for each individual test taker. The OTA system is also a highly scalable system built with cloud computing technology.

The session starts with an introduction to optimal test design, form assembly and adaptive testing. The user interface and functionalities of the OTA system will be demonstrated. We will then walk through a form assembly example step-by-step with input directly from the audience, and demonstrate a live ShadowCAT simulation, to give the audience first-hand experience of the OTA system. Finally, empirical research results from the OTA system will be summarized.

PRESENTERS:

Jie Li *McGraw-Hill Education CTB*

Michelle Barrett *McGraw-Hill Education CTB*

Seung Choi *McGraw-Hill Education CTB*

Craig Mills *McGraw-Hill Education CTB*

Breakout Session

MONDAY, MARCH 2, 2015 • 10:15 AM – 11:15 AM

Avoiding the "Techie"/"SME" Disconnect: Collaboration Principles for Performance-Based Test Development

Room: Polo Room
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Breakout

Performance-based testing is of increasing interest in certification, and the merits of testing via practical demonstration of skills are well-documented. Traditional multiple-choice test items are utilized by many programs because development is affordable and there are few technological issues to address. However, performance-based items in general and simulation items specifically have technical challenges that may restrict the item intent. In this respect, performance-based items are at greater risk than multiple-choice items for variance between item intent as defined in the job-task analysis and exam blueprint, and the actual item task as limited by technical constraints.

The focus on iteration and collaboration in Agile-like software development processes can be brought to bear with great results on the content development process for assessments that involve a complex technical implementation component. This approach ensures that content designers and the technical teams are frequently validating, together, that the technical implementation of an assessment or item are matching with the intent from the content point of view, and adjusting accordingly between iterations is an extremely effective way to ensure that the typical 'lost in translation' risks are mitigated and the final test will indeed be fit for purpose.

This presentation will explore how variations on the Agile methodology can be adapted successfully to the content and technical development process for complex exam development. You will gain insight from both the test sponsor and the test delivery vendor perspectives regarding actual lessons learned in developing and implementing performance-based items. We will list specific issues we encountered, how we addressed these issues, and the impact of our corrective measures on time to market and cost.

PRESENTERS:

Philip Koneman *Autodesk*
Sinéad Hogan McSweeney *Certiport*

Breakout Session

MONDAY, MARCH 2, 2015 • 10:15 AM – 11:15 AM

The 2014 Testing Standards: Implications for Assessment Organizations

Room: Rancho Room
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Breakout

The Standards are designed to promote sound and ethnical use of tests, as well as provide a basis for evaluating the quality of testing practices. The Standards provide criteria for evaluating tests, testing practices, and the effects of test use. This breakout session will focus on changes and refinements within the 2014 Standards for Educational and Psychological Testing (AERA, APA, and NCME, 2014) which have potential implications for organizations which develop, operate and use assessments in educational, employment and credentialing settings.

For example, changes from the 1999 to 2014 Testing Standards could have significant impact on the following processes:

- Broader view of the intended audience for the Standards – broadened to include students, policy-makers, and test-users, as well as measurement professionals (Plake and Wise, In press).
- Test development - reference to universal design and evidence-center design practices.
- International contexts – impact of globalization and more frequent use of assessments with different cultures across traditional national boundaries.
- Fairness – a major reorganization and elevation of this concept as foundation to the validation argument, as well as the focus away from special interest groups (e.g., ELL) to the needs of all examinees.
- Overarching standards – identified for all chapters except applications and meant to convey the intent of each chapter. These standards are viewed as applicable and essential to all tests (Plake and Wise, In press).
- Formative and interim assessments and how Standards may or may not apply to these types of uses.
- Technology – how innovative item types, the use of multiple platforms or a variety of devices might impact comparability across devices, modes (paper and digital) as well as security and other related issues.
- Evidence – what types of documentation would be required to support claims and intended uses?

In addition, the introduction notes that all standards should be met before operational test use, unless otherwise specified in the Standards. While this statement is not new, what are the implications for innovative assessments and overall test development?

Panelists were members of either the AERA-APA-NCME Management or Joint Committees which oversaw and completed the 2014 revision, respectively.

PRESENTERS:

Wayne Camara *ACT, Inc.*
Nancy Tippins *CEB/Valtera*

Thinking in New Boxes: Thought Provoking Conversations for Certifiers

Room: Celebrity Ballroom A
Practice Area Division(s):
Topic: Business Strategy and Operations
Session Type: Fishbowl

Increasingly, certifiers find themselves operating in an environment of accelerating change and competition. Even the most successful certifiers cannot afford to rest on their laurels. Innovation and creativity are key to sustaining a competitive advantage and consistently anticipating and meeting the needs of the market.

But who has time to be creative when we're already spending eight hours a day (usually more) just keeping up with routine operations? And when we do find the time to just sit and think, we soon discover that attempting to think outside the box, "on demand", simply is not effective or at best only yields ideas for incremental change.

So what's the solution? Bring your team members and spend an hour with us for an experience that will jump-start imaginative thinking about your certification business. In this interactive session, we'll pose provocative questions designed to catalyze creative thinking about the future of certification and the credentials you offer.

We'll also share tips and tools that will help you to think outside the "business as usual" box when you return to the office and will start you on the path to bringing strategic innovation to your certification business.

PRESENTERS:
Lenora Knapp *Knapp & Associates International*
Liberty Munson *Microsoft*

Taking Exam Development Virtual

Room: Rancho Room
Practice Area Division(s):
Topic: Business Strategy and Operations
Session Type: Breakout

When budget cuts happen, tough decisions need to be made. What changes can be made within a certification program to reduce cost BUT maintain the integrity of the exam development process? This presentation will outline the steps to migrate a standard exam development process from a 100% in-person development model to a 100% virtual development model based on real world examples including a case study from IBM. Not ready to go completely virtual? As part of the case study we will also cover the benefits of finding a hybrid approach between the two methodologies.

PRESENTERS:
Kristin Gibson *IBM*
Susan Farago *Farago Enterprises, LLC*

FUNDAMENTALS OF TESTING

Conquering the Great Divide Between Exam and Curriculum Development

Room: Celebrity Ballroom C
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Breakout

Individuals in the assessment industry often struggle with the age old question: How much information should be shared between the test development side of the business and the curriculum development side? Should exams be developed from supporting curriculum, or should the only overlap in information be the job task analysis (from which each group takes its development queue)?

This session will facilitate the discussion of best practices for exam/curriculum intersection. Four perspectives will be shared, discussing both the theoretical psychometric constructs and the business drivers that facilitate decisions around this issue. There will also be discussion around terminology and the titles of documents used in defining exams: job task analysis, exam blueprint, test definition, test expansion, etc. This session will feature a lively discussion of existing processes and innovative practices in conquering the great divide between testing and curriculum development. Share the issues that you have faced in protecting the validity of your exams and the solutions that you found. Where do you draw the dividing line between test development and curriculum development?

PRESENTERS:
Donna Senko *Ericsson*
Jerry Gorham *Ascend Learning*
Dania Eter *Cisco Systems*
Kris Vezina *Google*

Breakout Sessions

MONDAY, MARCH 2, 2015 • 11:30 AM – 12:30 PM

FUNDAMENTALS OF TESTING

The Power of Community: Leveraging Open-Source Solutions for Item Banking and Test Delivery

Room: Celebrity Ballroom F
Practice Area Division(s):
Topic: Test Administration and Delivery Models
Session Type: Breakout

Open-source software solutions, such as Linux, Drupal, Moodle, TAO, Jasper and others, allow for use of their programs without any licensing fees. In addition, open-source software source code is openly available for viewing, customization, modification and extensibility. The altruistic concept of open-source software is that the vast community of developers who use the product will contribute all – or at least most – of their enhancements made to the program back to the community to leverage.

The presenters will discuss the open-source platforms available for leverage by the education, employment and certification markets, will offer attendees deep insight into what open-source means, its value as a tool, and different applications of open-source software. Examples of how open-source software is currently being used for assessment and certification will be presented. Attendees will also have an opportunity to pose questions regarding their specific program's requirements and learn how open-source platforms may offer solutions.

PRESENTERS:

Kate O'Connor *Breakthrough Technologies, LLC*
Doug Wilson *Breakthrough Technologies, LLC*
Mark Gierl *University of Alberta*

FEATURED SPEAKER SERIES

Want to Report Subscores? Things to Think About

Room: Celebrity Ballroom G
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Breakout

Quite often, in reporting results of a test, there is a desire to report secondary information regarding performance on specific content areas or sections of the test to supplement the total score or classification. This additional information is usually reported in the form of subscores. These subscores are meant to benefit examinees by providing them more detailed information about their performance than is found in the total score. However, oftentimes the reporting of subscores provides little, if any, additional information than is found in the total score. Or worse, sometimes the information in subscores provides misleading information to the examinee. This presentation will discuss issues related to the validity, reliability and added value of subscores, including the calculation of "augmented subscores" which has been proposed in the literature as a way to increase the reliability of subscores. Examples will be used throughout to illustrate the various points.

PRESENTER:

Marc Gessaroli *National Board of Medical Examiners (NBME)*

PII Data Is an Important Business Asset. How Well Are You Protecting It?

Room: Celebrity Ballroom H
Practice Area Division(s):
Topic: Security
Session Type: Breakout

Your organization may be under pressure to locate, identify and classify personally identifiable information (PII) data to meet regulatory and industry requirements. Or you may need to undertake an audit exercise in response to a recent security breach or to identity vulnerabilities. PII data is information that, when lost, can lead to significant contractual or legal liabilities; serious damage to your organization's image and reputation; or legal, financial, or business losses. PII includes a broad spectrum of data, such as the following:

- Social security number
- State-issued identification number
- Mother's maiden name
- Driver's license number or similar operating license information
- Passport number
- Any other government-issued identification number
- Credit history
- Criminal history

This session will review how such critical information should be protected like any other valuable asset. As the world has become more reliant on technology, high-tech breaches have also adapted, becoming more and more sophisticated and organized. They can exploit human error and weak security controls to steal test taker personal information. Once this occurs, the integrity of a test sponsor's program is impugned, and its brand and reputation is damaged.

Awareness of PII theft and personal privacy has never been higher, and regulatory bodies and industry groups require you to be responsible caretakers of other people's information. A group discussion will consider the effects of recent legislation such as FISMA, Safe Harbor, the pending EU Data Protection Act as well as U.S. state-based privacy laws for evaluating how well you are protecting your test takers' PII data as a strategic business imperative. Presentation points will broadly cover the use of social security numbers (SSNs) and personal identifiers and a best-practices approach to have in place the right organizational policies, data/network security, records retention/disposal and security practices.

PRESENTERS:

Jason Young *Prometric*

Susan Layton *Federation of State Boards of Physical Therapy (FSBPT)*

Determining Whether Innovative Item Types are Right for Your Program

Room: Mirage Room
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Breakout

Have you heard this one about your program? "Passing the test only tells you so much about whether or not you can do the job." At the heart of any certification program is an assessment instrument or set of instruments that measure whether a candidate is suited for a job role or specialty. Innovative item types (or alternative item formats) are an increasingly popular means of developing examination items because they require exam candidates to engage in mental activities that more closely emulate job tasks and/or they capture concepts that are difficult to assess via traditional multiple-choice formats.

The downside? They are a lot of work. Diving into the innovative item "waters" involves needs and feasibility analyses, investigation and adoption of new technology, and substantial revision of the test development process – not to mention change management and clear messaging of the rationale for new item types to exam candidates and other stakeholders.

So are they worth it? Innovative item types are still on the "bleeding edge" of how certification exams are developed and administered. Many programs have adopted the use, while many more struggle to justify their implementation. This session provides insight from psychometricians and program directors as to how innovative item types may be a valuable component of their assessment instruments. Panelists will discuss the considerations and motivations behind the adoption and growth in the use of innovative item types, as well as the measurement considerations involved with developing, scoring, and evaluating the effectiveness of innovative item types.

PRESENTERS:

Amin Saiar *PSI Services LLC*

Timothy Muckle *National Board of Certification & Recertification for Nurse Anesthetists (NBCRNA)*

Ellen Julian *American Registry for Diagnostic Medical Sonography (ARDMS)*

Lee Skinner *American Nurses Credentialing Center*

Locking it Down: The Key to Test Readiness and Security

Room: Oasis Ballroom 4
Practice Area Division(s):
Topic: Test Administration and Delivery Models
Session Type: Breakout

With the move towards online, browser-based assessment for high stakes assessment as well as for formative assessment, exam integrity and test security have become a constant talking point within the online testing industry.

Browser-based assessment has obvious advantages in that test takers have immediate access to the assessments; no requirement to spend time downloading assessment software. It also means that candidates can use their own devices and do not necessarily need to go to designated testing centres. It does however, bring its own set of disadvantages.

How do we know that the test takers aren't finding the answers online? Or that they aren't recording the item questions to be distributed to their peers at a later date? What measures can we put in place to ensure the probity of online, browser-based assessment? Meeting these challenges requires creativity, technology, and, of course, funding.

In this session we will explore some of the challenges and successes that we and our clients have faced in implementing security measures to ensure the integrity of online assessment.
Solutions include:

- Secure browsers: Test takers are prevented from accessing other files, websites and folders on their devices. In order to protect valuable intellectual property and to prevent cheating they are also typically unable to print, copy, cut or paste data from their screens.
- Detailed event tracking: Monitors every online action the test taker makes for the duration of the assessment and alerts the system administrator if the test taker is outside of accepted behavior patterns.
- Test Windows: Setting the time during which the assessment may be taken.
- Student verification process and pacing: test administrators issue unique codes to individual test takers and have the ability to monitor the test taker's progress in real time. They can also start, save, quit, pause and allocate extra time all at the click of a button.

As well as looking at some of the current solutions for test security we will also discuss the future of test security and explore how security risks may evolve in the coming years and what new technologies may provide the greatest risks to the security of online assessment.

PRESENTERS:

Gavin Cooney *Learnosity*

Linda Andries *Houghton Mifflin Harcourt*

Learning Maps as a Basis for Educational Assessment

Room: Oasis Ballroom 5-6
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Breakout

Learning maps have been around a long time and in research projects have been used as the basis for the development of assessments. The Dynamic Learning Maps Alternate Assessment is an operational testing program that has been developed by a consortium of 20 states. In creating this program much has been learned about processes and procedures for developing tests in a learning map environment. The current operational version of the map consists of 146 foundational nodes that indicate early precursor skills necessary for academic learning, 1,823 nodes and 4,707 connections for English Language Arts, and 2,382 nodes and 5,096 connections for mathematics.

In this session we will describe and analyze the map and organizing structures that we created to support test development (conceptual areas and essential element concept maps), approaches to learning map visualization, item analysis and selection approaches, standard setting methods, and scoring and reporting practices a learning map based assessment.

PRESENTERS:

Neal Kingston *Achievement and Assessment Institute*

Angela Broaddus *Achievement and Assessment Institute*

Transitioning to Multiple Response Items: A Case-Study

Room: Oasis Ballroom 7
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Breakout

Research suggests that traditional multiple choice items are limited in their ability to assess higher cognitive levels and multiple aspects of a process, procedure, or technique. As such, many organizations look beyond traditional multiple choice items to address these assessment needs. Historically, K-type items have been one of the only item types available to address these needs. While these items do address assessment needs, there are limitations to them. Research has shown that K-type items are more difficult and less discriminating than other multiple choice item types. Additionally, K-type items can be difficult for exam committee members or subject matter experts to create.

Like K-type items, multiple response item types assess higher cognitive levels and multiple aspects of a single topic. Multiple response items look like multiple choice items with the only difference being that an examinee can select one or more response options. Additionally, the difficulty and discrimination of multiple response items may be more in line with traditional multiple choice items. Therefore, the use of multiple response items may allow for organizations to avoid some of the limitations of K-type items.

In 2012, the Council of Landscape Architectural Registration Boards transitioned from K-type items to multiple response item types. Data has been collected on over 100 items that were originally presented to candidates as K-type items and have now been presented as multiple response items. Item difficulty and discrimination were analyzed for each item across item formats. The ability to analyze the same items across two formats is unique to this study, as previous research on these two item types has been done on each item type separately.

In this session, we will compare the difficulty and discrimination of K-type items and multiple response items, discuss multiple response items that were converted from K-type versus those that were developed from scratch, and provide recommendations on the development and implementation of multiple response items.

PRESENTERS:

Corina Owens *Professional Testing, Inc.*

Adrienne Cadle *Professional Testing, Inc.*

Jim Penrod *Council of Landscape Architectural Registration Boards*

Lessons Learned: Designing and Building Technology Enhanced Items Using a Writing and Editing Tool

Room: Polo Room
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Breakout

This session, which will be presented by the NWEA Content Acquisition team, will demonstrate the functionality of the technology enhanced item (TEI) writing and editing interface that NWEA uses. Participants will take away an understanding of the advantages of enabling subject matter experts (SMEs) to create TEIs and the advantages of how item specifications can be designed to allow SMEs the flexibility to choose the item type necessary for assessing student knowledge. The presentation and hands-on workshop will be relevant to any conference participants who are interested in the development of TEI items. Although some of the presentation may address particular aspects of assessing K-12 educational standards, participants will not need to be in the K-12 educational assessment industry to benefit from this session.

NWEA uses a content management system that enables the NWEA Content Acquisition team to assign the creation of QTI-compliant TEIs directly to SMEs. SMEs can use standard WYSIWYG editing tools to create fully functional TEIs in real time. SMEs are given the ability to evaluate how a TEI's functionality will allow students to interact with the item content and demonstrate the depth of their understanding of the skills being assessed. SMEs can select from a tool set of available QTI item types—such as Gap Match, Graphic Gap Match, Text Entry, and Choice Interaction—to determine which TEI type will best fulfill the item specification, assess the learning objective, and match the desired level of cognitive complexity. Providing SMEs with the tools to build TEIs themselves streamlines the item development process by eliminating traditional workflow loops in which programmers "translate" the vision of an item writer by reading descriptions of items on item-writing templates. SMEs can now be given the ability to build items instead of simply describing them, allowing for a flexibility that will yield more effective and creative approaches to assessing student knowledge.

PRESENTERS:

Lindsay Sanders *Northwest Evaluation Association*
Michael Laus *Northwest Evaluation Association*
Joanne Zimmer *Northwest Evaluation Association*
Justin Marks *Northwest Evaluation Association*

Demonstrating the Value of Assessment in Talent Management

Room: Celebrity Ballroom B
Practice Area Division(s):
Topic: Business Strategy and Operations
Session Type: Breakout

A longstanding challenge for testing professionals is to communicate the value of assessments to non-technical stakeholders. While technical experts understand that tests that exhibit good psychometric properties (e.g., validity, reliability and fairness) will lead to positive results for the organizations that use them, business leaders think in different terms and metrics (e.g., risk management, organizational effectiveness, and company-specific metrics). Industrial/ Organizational psychologists have attempted to demonstrate the value of assessments using complex utility models (e.g., Brogden), however, the resulting estimates of utility are not widely understood or accepted by business leaders; hence the challenge.

In this interactive session, four seasoned assessment experts from four test publishing and assessment consultancy firms will review and discuss key concepts underlying the value of a range of assessment types for the acquisition, development and management of talent in organizations. Classic and alternative metrics will be examined, with an eye towards bridging the gap between psychometric and business metrics, and leveraging "big data" to align talent assessment with strategic objectives and outcomes. Real life examples and case studies will serve as a backdrop for exploring alternative approaches and metrics for expressing value to organizational leaders and stakeholders.

PRESENTERS:

Eugene Burke *CEB*
John Weiner *PSI Services LLC*
Ryan Ross *Hogan Assessments Systems*
John Jones *Vangent, A General Dynamics Company*

Platinum Sponsor Sessions

MONDAY, MARCH 2, 2015 • 11:30 AM – 12:30 PM

Alpine Testing Solutions

PLATINUM SPONSOR SESSION

Don't Be Late to the Game – Security Should Not Be An Afterthought

Room: Oasis Ballroom 1

Session Type: Sponsor Session

Test development, when done correctly, can be an arduous process. If time is not taken to build in security, all of that hard work could vanish with one breach. Not only could thousands of dollars of test content be lost, but you could also damage your program's reputation and your company's brand. The best credentialing programs weave security into all aspects of their development process, because if you wait until your test is published, you are already late to the game.

Creating a holistic plan of how you will approach test development and delivery will increase the security and integrity of your credential. By incorporating a risk assessment into the program design at the beginning of your development, you can identify potential threats to your program and develop mitigation strategies upfront that can be incorporated throughout the exam's life cycle.

Every successful security plan contains prevention, detection, and enforcement strategies but the most important part of any security plan is that it be proactive. Creating a comprehensive plan of how you will address security threats throughout the lifecycle of your program, before you have a breach, is the best way to protect your investment.

Presenters in this session will provide practical information and resources to assist attendees in how to develop a holistic security plan that takes into consideration every phase of a credential's life cycle.

PRESENTERS:

Jill Buroughs *Alpine Testing Solutions*

KJ (Ken) Salchow *F5*

Sara Rupp *CompTIA*

Internet Testing Systems

PLATINUM SPONSOR SESSION

What's New in Accessibility for Computer-Based Testing

Room: Oasis Ballroom 2

Session Type: Sponsor Session

During this session, we will demonstrate the ability to respond to Technology Enhanced Items using the keyboard. Over the past year we have made extensive progress with accessibility for computer-based exams and we now have solutions for item types like drag and drop or hot spot. Accessibility is now a core function of our test delivery system, not an add-on. We'll also update you on progress we've made in other areas of our technology.

PRESENTERS:

Pat Ward *Internet Testing Systems*

Cabell Greenwood *Internet Testing System*

Cito

PLATINUM SPONSOR SESSION

High Stakes CBT Examinations Breaking Through

Room: Oasis Ballroom 3

Session Type: Sponsor Session

Questify is in its design driven by international success and in this presentation we emphasise the use of international open standards (QTI, LTI). We will bring you up to speed on the latest developments in high stakes CBT examinations. In collaboration with the Dutch Test & Exam Board (CvTE), innovations are being achieved in secondary national examinations. "Facet" - the assessment solution developed in name of CvTE – plays as breaking through technology a dominant role. In addition to this, we will demonstrate the use of our authoring and delivery platform, Questify, for the Driving Standards Association.

PRESENTERS:

Nynke de Boer *CvTE*

Mark Molenaar *Cito*

Thank you to our 2015 Innovations in Testing Golf Sponsor

GENERAL DYNAMICS
Information Technology

12:30 PM - 1:30 PM

Education Practice Area Division Luncheon – All are Welcome!

Room: Oasis Den

The Education Division meeting will take place at lunch to allow time for associating with each other and reviewing the past year. We will also discuss what the Division might take on followed by a brief meeting and collection of ideas.

Asia Regional Division Luncheon – All are Welcome!

Room: Hospitality Suite 1019

Join the Asia Regional Division for lunch while learning more about current initiatives and volunteer opportunities.

12:30 PM – 1:45 PM

Lunch with Exhibitors

Room: Ambassador Ballroom

FEATURED SPEAKER SERIES

1:45 PM – 2:45 PM

Value-Added and the Use of Student Test Scores to Evaluate Teachers

Room: Celebrity Ballroom A

Practice Area Division(s):

Topic: Legislation, Policy, and Accessibility

Session Type: Fishbowl

The use of indicators based on student achievement to evaluate teacher effectiveness is often mandated by political decision-makers. And, although the calculations can be performed, how do they function? This fishbowl session will address the concept of value-added and the questions, concerns and observations of members of the audience with two experts in assessment, Drs. Henry Braun and Mark Reckase, as leaders of the session. Dr. Alina von Davier will moderate. Short introductory remarks will be followed by a lively interchange.

PRESENTERS:

Henry Braun *Boston College*
Mark Reckaase *Michigan State University*
Alina von Davier *Educational Testing Service (ETS)*

Breakout Session

MONDAY, MARCH 2, 2015 • 1:45 PM – 2:45 PM

Common Challenges in Developing & Updating Test Content

Room: Celebrity Ballroom B
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Breakout

Regardless of practice area, test publishers face common challenges in developing items and other test content. Also, common practices can vary across different fields such as those represented by ATP's different practice area divisions. As such, the goals of this session are threefold. First, panelists will use actual case studies, scenarios, and other real-world examples to discuss current best practices, alternative methods, and innovative approaches to developing and evaluating test content within their field. Second, we will compare and contrast methods across divisions. Finally, we will demonstrate how members of ATP's different practice area divisions can work with and learn from one another to mutually benefit all parties.

Questions we will address include:

- How does your company develop test content?
- How do your development procedures vary by construct type (e.g., aptitudes, competencies)?
- How do you evaluate the effectiveness (e.g., reliability, validity, etc.) of test content?
- How do you adapt your development processes for normative vs. standards-based scoring?
- On average, what is the shelf life of your items and tests?
- When and how do you update assessments?
- What are the biggest obstacles you face in developing or updating test content?
- Do you proactively or reactively work towards developing new test content?
- How do you guard against test misuse?

Audience members will leave this session with (a) a better understanding of the challenges typically faced when creating, reviewing, and updating test content, (b) new ideas they can take back to their own organizations, and (c) a greater appreciation of the value of cross-division interactions as part of the ATP community.

PRESENTERS:

Blaine Gaddis *Hogan Assessment Systems*
Jeff Foster *Hogan Assessment Systems*
Manny Straehle *International Credentialing Associates*
Joseph Orban *Questar Assessment, Inc.*
David Herzberg *Western Psychological Services*

Including Special Populations in Test Security Policies and Procedures

Room: Celebrity Ballroom C
Practice Area Division(s):
Topic: Legislation, Policy, and Accessibility
Session Type: Breakout

This session brings together two important issues that are often considered separately from each other in state assessments: (1) the need to provide accommodations to students who require them so that they are able to show their knowledge and skills and (2) the challenge of preserving test security throughout the testing process as the needs of the students are met. Stronger test security policies and procedures are needed in states and school districts to help ensure the integrity and validity of test results. Some students with disabilities and English language learners (ELLs) need to use accommodations to meaningfully access an assessment, and a few participate in alternate assessments. There is the risk that some policies may adversely affect accessibility for some of these students. For example, some accommodations (e.g., interpreter, reader, and scribe) may require someone other than the student to see the test. Similarly, some students may need to use assistive technology that has the potential to compromise test security.

This session will include information about how these issues are addressed by several sets of guidelines and standards such as:

- The National Center on Educational Outcomes' (NCEO) Test Security Policy Considerations
- The Council of Chief State School Officers' (CCSSO) Technical Issues in Large Scale Assessment (TILSA) Test Security Guidebook for States
- Caveon Test Security Standards
- ATP/CCSSO Operational Best Practices for Statewide Large-Scale Assessment Programs
- APA/AERA/NCME Standards for Educational and Psychological Testing
- NCME Testing and Data Integrity documents

This session will also present the results of an analysis of states' test security and confidentiality policies and how they include accommodations, alternate assessments, and issues related to students with disabilities.

The focus of this session will be on how these guidelines and standards can be used to inform the development of effective test security policies at the K–12 level that address issues related to the assessment of special populations. The intent of this session will be to share important information from a variety of sources and to start a lively discussion about how to provide comprehensive test security measures while maintaining accessibility. Recommendations on how to enhance security procedures will be discussed, and links to all the key resources on test security will be provided to attendees.

PRESENTERS:

Sheryl Lazarus *National Center on Educational Outcomes (NCEO)*
Martha Thurlow *National Center on Educational Outcomes (NCEO)*
John Olson *Olson Educational Measurement & Assessment Services*

FUNDAMENTALS OF TESTING

The Subject Matter Expert Lifecycle: Pitfalls and Assessment Implications

Room: Celebrity Ballroom F
Practice Area Division(s):
Topic: Program Design
Session Type: Breakout

A persistent challenge in developing quality assessment programs is not only the continuity of Subject Matter Expert (SME) involvement throughout development or revision cycles, but also the quality and vigilance of SMEs. This session will reflect on a model for utilizing SMEs to develop statewide standards and assessments for career-technical education programs in Ohio—The Subject Matter Expert Lifecycle. Attendees will be presented with pitfalls that can occur during the SME lifecycle and associated dangers to assessment quality. Best practices around SME definition, development and training will be discussed; along with balancing best practice within the constraints of less-than ideal implementation.

PRESENTERS:

Brooke Parker *The Ohio State University*
Michael Loyd *The Ohio State University*

Extending the Applications of Measurement Decision Theory

Room: Celebrity Ballroom G
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Breakout

While its strength is in rank ordering individuals, Item Response Theory has been a useful tool in classification testing. But, because of its unidimensionality and continuous trait assumptions, IRT may not be the optimal or most efficient model for classification testing. This session will explore the utility of Measurement Decision Theory - an alternative, innovative measurement model designed specifically for classification testing - using actual data that spans pre-employment selection, certification, and personality examinations.

MDT is based on the likelihood of group classification (e.g., qualified v. not qualified) given the pattern of candidate responses. Lawrence Rudner will present a brief, non-mathematical overview of the model. Advantages in terms of the need for only small samples for calibration, applicability to multidimensional data, management of false positives and false negatives, adaptive testing, and shorter tests will be emphasized.

Because the model is only concerned with the determining group membership rather than pinpoint ranking, fewer test questions are typically needed. Steven Barkley will present on the comparative accuracy of shortening a certification exam using IRT and MDT. Using MDT, a fixed form test could be reduced to 25% of the original length with only a slight loss in classification accuracy.

Forced-choice testing, in which test takers must choose among alternatives and leads to multiple classifications such as don't hire, hire, or proceed with caution are easily accommodated by MDT. Julie Carswell will evaluate the accuracy and efficiency of a variable length MDT computer adaptive algorithm for forced-choice testing with multiple classifications based on candidate test data derived from a pre-employment selection assessment.

The decision to use MDT needs to be based on pragmatic as well as psychometric considerations. Patrick Wadlington will discuss the decision to use MDT on a highly visible and well-known personality assessment that has been completed by more than 100,000 examinees over four years and the ramifications of that decision.

Reid Klion will serve as discussant. He will identify the key points of each of the presentations and highlight recommendations and considerations for practice. Additionally, he will present an overview of free software that can be used to explore MDT.

PRESENTERS:

Lawrence Rudner *LMP Associates*
Steven Barkley *Paradigm Testing*
Julie Carswell *Sigma Assessment Systems, Inc.*
Reid Klion *pan, Performance Assessment Network*

Driving Automated Scoring: Avoiding the Bumps in the Road

Room: Celebrity Ballroom H
Practice Area Division(s):
Topic: Program Design
Session Type: Breakout

Automated scoring engines, like Pacific Metrics' CRASETM, offer great promise in reducing scoring costs and time while also ensuring scoring consistency and accuracy. These engines have the potential to increase the use of constructed-response, writing prompts, and technology-enhanced items at a time when reducing costs is paramount. Automated scoring, combined with the ability of innovative items to tap skills not readily assessed with multiple-choice items, stands to greatly enhance the types of items that can be scored cheaply, quickly, consistently, and validly.

However, implementing automated scoring into an assessment program is new territory for many organizations, and can be disastrous if done incorrectly. The purpose of this session is to highlight key considerations when planning to use automated scoring. The Presenter will discuss the topics below with detailed examples of both problems and solutions.

Is AS worth the expense? Automated scoring is not always a cost-effective scoring option. The up-front costs of engine configuration, system integration, and decisions about response handling, can outweigh per-response scoring costs. What items can AS score? Automated scoring performance varies by item and response types. AS has been shown to perform comparably to humans on essay scoring and less comparably on constructed response items. Do you have electronic hand-scored training data? Automated scoring typically requires hand-scored training data for each item that is administered. The nature and quality of the data are critical to the success at online implementation. Do you have an integration plan? Even with detailed specifications, the integration of two systems can be a thorny issue. Do you have a monitoring plan? Automated scoring is not a plug and play solution, and a monitoring strategy for both scoring and technology infrastructure is critical to success. Do you have a fallback plan? If the automated scoring system fails or the scoring is not adhering to specifications, a contingency plan must be in place. How will you communicate the use of automated scoring to stakeholders? For some stakeholders, automated scoring has a trust problem. Decisions about how to incorporate automated scoring may be influenced by the value your stakeholders place on automated scoring.

PRESENTERS:

Susan Lottridge *Pacific Metrics*
Nikki Eatchel *Scantron*

Using Automatic Item Generation to Develop Practice Non-Verbal Reasoning Items for a High Stakes Medical Admission Test

Room: Mirage Room
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Breakout

The Undergraduate Medicine and Health Sciences Admission Test (UMAT) is developed by the Australian Council for Educational Research (ACER). It is used by Australian and New Zealand universities to assist with the selection of students into undergraduate medicine and health science courses. Approximately 15,000 candidates sit UMAT each year.

Because of the high stakes nature of UMAT, a significant number of test preparation providers offer courses to coach potential UMAT candidates. These businesses charge high prices for attendance and in some cases the information they provide is incorrect and potentially misleading.

ACER has put in place a number of cost-effective strategies, to provide all the resources candidates need to prepare for UMAT. Nevertheless, a continual candidate request is to provide more practice materials; it is this that drives many candidates to UMAT preparation providers. To support further test candidates and avoid depleting the ACER item bank, ACER approached Mark Gierl and his team at the University of Alberta, to develop an increased number of non-verbal reasoning items utilizing automatic item generation software developed at the University.

This presentation looks at how effective the automatic generation of non-verbal, UMAT style, items is. It also poses the question of the viability of the methodology for the development of actual test items.

PRESENTERS:

Marita MacMahon Ball *ACER, Inc.*
Mark Gierl *University of Alberta*

The Incredible Journey… to Trustworthy Assessment Results: A Panel Discussion of Best Practices for Using Statistics to Invalidate Scores

Room: Oasis Ballroom 4
Practice Area Division(s):
Topic: Security
Session Type: Breakout

As testing professionals, we all pursue a common goal—valid, trustworthy test results. However, setting off to attain this goal can feel a bit like starting a long, arduous journey, one filled with myriad challenges and dangers. Some of the most treacherous aspects of this adventure can involve creating policies and procedures around the use of Data Forensics to determine which test takers may have gained unfair advantage and should have their scores invalidated.

Fortunately, this breakout session will serve as a map for avoiding many perils and risks on this incredible journey.

Our panel of industry experts has decades of experience using statistics to invalidate scores in programs that test anywhere from a few hundred candidates a year to hundreds of thousands. Their combined expertise on this topic provides a comprehensive set of guidelines that will benefit assessment professionals managing any high stakes test program. Panel members will address several of the crucial questions any program manager must grapple with when instituting the powerful deterrent of statistical score invalidation. These questions involve legal, operational, organizational, community, and media dynamics. Panelists will also accept questions from the audience, facilitating an interactive, engaging conversation between professionals who share the same validity concerns.

PRESENTERS:

Steve Addicott *Caveon Test Security*
Jennifer Semko *Baker and McKenzie LLP*
Liz Burns *NetApp*
Jack Terry *National Board of Examiners in Optometry*
Rachel Watkins Schoenig *ACT, Inc.*
Dennis Maynes *Caveon Test Security*

How to Develop Video-Based Content for Computer-Based Examinations: Experiences and Lessons

Room: Oasis Ballroom 5-6
Practice Area Division(s):
Topic: Program Design
Session Type: Breakout

Once upon a time, using videos to supplement tests was every test developers' dream. Today, computer delivery mechanisms provide a full range of opportunities for video-based content. While all test developers are fascinated by the potential of video-enhanced testing, it is understood that the development of video content is quite a different undertaking from the development of regular text-based test items. This session will share the experiences and lessons of American Board of Psychiatry and Neurology's (ABPN) video content programs that develop both videos and video-associated test items. Topics covered will include:

- The multi-faceted nature of the video content programs
- Assessing the needs for the video content
- Assessing vendors' delivering capabilities for video content
- Demonstration of ABPN video content
- The general process of the video content development
- Lessons and challenges of developing scripts for actor-based videos
- Lessons and challenges of developing real-patient-based videos
- Lessons and challenges of publishing exams with videos
- Lessons and challenges of developing items for videos
- Examinee feedback on video content

The session will be a dynamic and interactive event between presenters and attendants. At the beginning of the session, the presenters will ask the attendants to share any specific interests in video content development. The presentation will focus on the interests of the attendants. In addition, the presenters will engage the attendants by presenting real challenging cases and encouraging the attendants to brainstorm with the presenters for solutions.

PRESENTER:

Linjun Shen *American Board of Psychiatry and Neurology*

Selecting Civil Servants Across 28 European Countries: Psychometric Challenges and Solutions

Room: Oasis Ballroom 7
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Breakout

The European Personnel Selection Office (EPSO) delivers a staff selection service on behalf of the Institutions of the European Union. For each selection process candidates from the 28 member states are assessed in order to select the best for possible recruitment as EU officials within the Institutions.

Around 40,000 graduates apply every year for less than 300 posts of generalists and undergo a two stages selection process: a pre-selection phase with CBT and an assessment centre. The pre-selection phase consists of tests assessing verbal, numerical and abstract reasoning as well as situational judgement. Verbal and numerical reasoning is delivered in all 24 official languages of the European Union. The major psychometric challenges for CBT are translating and calibrating the items and assembling them into equivalent test forms.

In order to best meet these specific challenges of multilingual large-scale testing EPSO has developed innovative and sophisticated in-house solutions involving comprehensive item analyses based on the Rasch model, equating across the 24 languages and a test form assembling algorithm. By using these solutions EPSO can ensure fair and equal treatment of all candidates.

PRESENTER:

Gilles Guillard *European Personnel Selection Office (EPSO)*

Candidate Self-Perceptions of Exam-Related Knowledge and Their Role in the Interpretation and Use of Subscore

Room: Polo Room
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Breakout

In recent years, there has been much discussion in the testing industry surrounding the topic of subscore reporting (i.e., domain-level score reporting). On one hand, we know candidates want detailed feedback regarding their exam performance. Furthermore, professional testing standards encourage subscore reporting if it is psychometrically defensible. On the other hand, subscores are often unreliable, and there is legitimate concern that these subscores will be misinterpreted or misused by candidates or other recipients of score reports.

Relatively little attention, however, has been devoted to investigating how candidates actually interpret and use subscore information. One issue worthy of consideration is candidate self-awareness of their own areas of strength and weakness with regard to the examination content. Candidates' self-awareness of their domain-level knowledge or abilities may be relevant for their interpretation of subscore information for a variety of reasons. For example, it is possible that a candidate's acceptance of the subscore results as accurate and his or her reliance on those scores for guiding future study efforts depends, in part, on the extent to which the subscores confirm or disconfirm that candidate's self-perceptions. We sought to investigate the accuracy of candidates' self-perceived knowledge or abilities as well as the relationship between candidate self-perceptions and candidate interpretations or uses of subscore information.

We collected candidate self-assessments of their general knowledge and of their exam performance in the content domains of a certification examination they had recently taken. We compared these self-ratings to their actual test scores for several fitness professional certification programs with reliable subscores. Consistent with the literature on self-awareness and self-assessment, one of the main findings of this research is that candidates have very little insight into their own areas of strength and weakness. In this presentation, we will discuss these findings and their implications for how subscores are interpreted and used by both passing and failing candidates. We will also discuss factors that may contribute to candidates' lack of self-awareness. Finally, we will facilitate an interactive discussion about the implications these findings could have on subscore reporting best practices.

PRESENTERS:

Andrew Dwyer *Castle Worldwide*
Sarah Carroll *Castle Worldwide*
Todd Galati *Cisco Systems*

Implementing Automated Scoring in K–12 Assessments: Development and Scoring Considerations

Room: Rancho Room
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Breakout

As K-12 assessments are moving online, new item types are being developed that can be automatically scored through rule-based machine rubrics or scoring engines that utilize forms of artificial intelligence to replicate human scoring. Use of these scoring methods has implications for item authoring, development, and review. Online item types currently in use in K-12 assessments are most easily defined by student response types such as selecting one or more response options in a list or table format, entering mathematical numbers or symbols from a keyboard or number palette, manipulating the position of item elements on the screen (drag and drop), creating a graph or diagram by placing points on a grid, or typing a text response on a keyboard. For each of these response types, scoring rules or rubrics need to be created for automated scoring. Item writers must determine how complex responses correspond to item score points. This requires a thoughtful degree of alignment between item authoring and the required blueprint and design that goes beyond the traditional item development plan.

Automated scoring of constructed responses using artificial intelligence has been the subject of research over the past two decades. Artificial intelligence engines have proven to be quite good at scoring general writing ability, but have been less effective scoring shorter responses and evaluating the quality of the content. Nevertheless, given the large numbers of responses to be scored in K-12 assessments, automated constructed response scoring can offer significant benefits in terms of scoring turnaround time and cost efficiencies. Scoring engines may be deployed as an adjunct to a human reader with human adjudication of discrepant scores or multiple scoring engines may be deployed with human ratings provided in the case of discrepancies.

Presenters will discuss their experiences in authoring and scoring the 20,000 items that were field-tested by the Smarter Balanced Assessment Consortium including specifications and validation of rubrics for machine-scored items and feedback from artificial intelligence scoring to inform item development. Examples of items, machine and human scoring rubrics, and student responses will be shared as illustrations. Implications for K-12 assessment and for assessment in other contexts such as certification and licensure will be discussed.

PRESENTERS:

Sally Valenzuela *McGraw-Hill Education*
Craig Mills *McGraw-Hill Education*
Kevin Sweeney *The College Board*
Vincent Kieftenbeld *CTB/McGraw Hill*

Pearson VUE

PLATINUM SPONSOR SESSION

If You Come to a Fork in the Road, Which Road Should You Take? Test Owners no Longer Need to be Limited to One Road When it Comes to Test Delivery

Room: Oasis Ballroom 1
Session Type: Sponsor Session

Test owners have more test administration options than ever when deciding how and where to deliver their exams. When faced with deciding between traditional test center delivery, client-managed test centers, Internet-based testing, online proctoring, observational testing, mobile delivery or a combined approach, how do you determine the best test administration solution for your program? It can be confusing.

Pearson VUE can help. This session will define and explore each of these test delivery models. We will discuss the advantages and tradeoffs each model provides your candidates and program as a whole. Testing programs that are successfully using new delivery options will be highlighted to show how the different administration options may fit unique testing needs. We will also show you how you can combine the different models across your testing program and leverage an integrated multi-modal delivery approach.

Leave this session with a more complete understanding of each test administration option and which options are best for your program.

PRESENTER:

Peter Pascale *Pearson VUE*

Platinum Sponsor Sessions

MONDAY, MARCH 2, 2015 • 1:45 PM – 2:45 PM

Professional Examination Service

PLATINUM SPONSOR SESSION

The Future Role of Noncognitive Assessments in Credentialing

Room: Oasis Ballroom 2
Session Type: Sponsor Session

Hiring organizations and academic institutions increasingly recognize that workplace and educational success depends on a multitude of interpersonal and intrapersonal competencies, in addition to the traditionally expected set of cognitive skills. These so-called non-cognitive or "soft" skills include everything from work ethic, teamwork and emotional competence to time management, integrity, resilience and intercultural competence.

Dr. Richard D. Roberts, Head of ProExam's Center for Innovative Assessments, will share some of his research plans tied to new, innovative measures of non-cognitive skills and the value they can provide to licensure and certification organizations.

PRESENTER:

Richard Roberts *Professional Examination Service*

ATA

PLATINUM SPONSOR SESSION:

Leveraging Mobile Technologies to Enable Secure and Timely Deployment of Test Delivery – with Global Case Studies

Room: Oasis Ballroom 3
Session Type: Sponsor Session

With the popularity of mobile devices and high internet accessibility, test takers and awarding bodies are urging test services providers to leverage mobile technologies to provide flexible test center deployment and test administrations. Without the trade-off in security and test integrity, ATA has successfully developed and implemented a global mobile testing system (MTS) that could instantly turn an ordinary classroom or an assembly hall into a secured test room with locked-down mobile devices. MTS has been successfully implemented domestically throughout China in thousands of classrooms and also used in hundreds of English learning classrooms around the world since 2013. Such MTS is expected to gain its popularity with its efficient deployment and secured implementations for exams needing scattered test locations and test windows.

PRESENTER:

Alex Tong *ATA, Inc*

ASSOCIATION OF TEST PUBLISHERS
MAKING THE WORLD SMALLER FOR TESTING

The vision of the Association of Test Publishers (ATP) is to become the "International Voice of Testing." Although the ATP was founded as a trade association in North America in 1992, testing and assessment have been developing into a global community at a rapid pace. Consequently, the ATP expanded by adding a European Division (2005), an Asian Division (2011), and a founders' group has met to organize an India Division (2012). Accordingly, the ATP is now poised to become recognized as the global trade organization for test publishers, thus making the world smaller for testing. Such expansion is appropriate, inasmuch as the language of testing and psychometrics is international and many issues faced by publishers are the same all over the world. Members of the ATP are truly part of a global community of testing, benefiting from the coordinated efforts of a strong organization that offers education, networking, advocacy, advertising, and industry advancement/protection activities on behalf of publishers worldwide.

Dessert / Fishbowl Session / Breakout Sesion

MONDAY, MARCH 2, 2015 • 3:00 PM – 5:00 PM

3:00 PM – 4:00 PM

Dessert with Exhibitors

Room: Ambassador Ballroom

Enjoy dessert with the industry leading vendors in the exhibit hall.

4:00 PM – 5:00 PM

Innovating Report Design in Testing

Room: Celebrity Ballroom A

Practice Area Division(s):

Topic: Data Management

Session Type: Fishbowl

This session will engage participants in a fishbowl discussion and interactive working session on designing reports in the testing industry. Using generated reports, the facilitators will lead participants through a discussion on the design considerations for reports used in the testing industry as well as noting the key areas that should be improved as we move forward. Topics discussed will include usability, analytics, interactivity, and data visualization as it pertains to reports in the testing industry.

The facilitators will begin by engaging the first set of participants in a discussion about the issues that have arisen for them when designing reports for their own programs. The fishbowl discussion will proceed from there into showing constructed reports to illustrate common problems and innovative solutions to report design.

PRESENTERS:

Daniel Wilson *Mountain Measurement, Inc*

Brian Bontempo *Mountain Measurement, Inc*

Including Video in an Assessment: Use of a Storyboard to Support Item Writer

Room: Celebrity Ballroom B

Practice Area Division(s):

Topic: Testing, Measurement, and Psychometrics

Session Type: Breakout

When an exam is administered on computer, the addition of certain elements, like the use of video clips, becomes possible. And video, whether used in the item stem or the response options, has the potential to increase the fidelity of the exam and to improve measurement of key elements from the test construct. As part of ABIM's ongoing goal of making the exams as close to real practice as possible, several video-related efforts have been pursued.

Video items can serve a variety of purposes, but medical certification videos can be divided into two primary types of scenarios: real encounters and simulated encounters. Real encounters involve real patients and real physicians in a practice or clinical setting, and are most often used when a patient's physical appearance or condition cannot be accurately portrayed by an actor. Simulated encounters involve standardized actors, filling the roles of patient or physician in a simulated clinical setting, and are most often used when a patient's physical appearance or the patient/physician encounter can be effectively replicated. Simulated encounters might also be used to re-enact scenarios, including the simulation of appropriate (or inappropriate) patient safety techniques, like hand-washing.

Nevertheless, a number of challenges remain, including those related to item writing and video scripting. Even seasoned item writers may struggle with the markedly new task of writing video scripts. To address the demanding nature of item-writing for video, ABIM has been developing a video storyboarding process to support item writers, providing prompts for information that might otherwise be forgotten and building important considerations, like maximizing video utility, into the process. The storyboarding process is designed to streamline the video item-writing process and increase efficiency. It also ensures the consistency and quality of both videos and related items and guides item writers to consider video only if it will add value to the examination, rather than using video simply because the technology is available.

This session will include small group activities structured around the tasks of storyboarding and writing a video item. Participants will obtain an understanding of how to develop a standard storyboarding process to apply to their practice area to increase efficiency and support their own item writers.

PRESENTERS:

Bryn Herrschaft *American Board of Internal Medicine*

Cynthia Parshall *CBT Measurement*

Pushing the Boundaries – Going Global with Your Certification

Room: Celebrity Ballroom C
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Breakout

This session will provide information about the challenges faced by a certification body in providing their exam in Arabic and how these challenges were overcome. At E-ATP 2013, Martin Mackain-Bremner from the Association of Accounting Technicians in the UK presented on ‘Pushing the Boundaries’, including some of the challenges around ‘Going Global’ with their certification. The focus was on localisation in the contexts of technical, cultural and linguistic considerations.

Soon after the 2013 conference, AAT embarked on a pilot localisation project that oversee the translation of one qualification (Bookkeeping) into Arabic, with partners Comms Multilingual and BTL. This presentation will provide a description of and status update for that project, known as Transit, and will describe some of the principal issues and challenges that the AAT development team faced.

The aim was to test the processes out on a complex language like Arabic with a view to being able to offer the certification in many more languages later. Many important lessons were learned and we developed some crucial tools and processes to deliver this and future similar projects.

Many organisations are considering offering their certifications in other languages and to other cultures. This will then help attendees to be prepared for the challenges they will face and to provide them with practical guidance as to how to overcome such challenges, so that they can offer their certifications, tests and assessments successfully in other languages.

The example will be from a certification body, but it is equally relevant to anyone involved in testing, who is looking to make their exam, assessment or test available in other languages and to other cultures.

PRESENTERS:

Sue Orchard *Comms Multilingual Ltd*

Martin Mackain-Bremner *Association of Accounting Technicians (AAT)*

Cloud Based Assessment – A Match Made in Heaven

Room: Celebrity Ballroom F
Practice Area Division(s):
Topic: Test Administration and Delivery Models
Session Type: Breakout

The increasing prevalence of cloud-based assessment can be taken as a validation of this assessment delivery option. However, despite its increasing popularity it is still a relatively new concept and as such is viewed as uncharted waters in many respects.

In this session, we hope to better prepare those considering implementing a cloud-based assessment strategy by sharing some of our experiences gained in delivering cloud-based assessment to millions of students across the USA. As well as discussing the general technical, business and strategic advantages of leveraging the power of the cloud, we will also discuss specific challenges we have faced when delivering cloud-based assessments and the solutions implemented to overcome these obstacles.

We will focus on:

- the improved testing experience for test takers i.e anytime, anywhere access;
- the administrative and reporting benefits that real-time, test progress tracking can offer;
- ease of integration as compared with other assessment delivery methods;
- perceived barriers to using cloud-based assessment, such as test security and academic validity;
- technical challenges such as archaic technical infrastructure, schools with limited bandwidth and firewall and proxy restrictions;
- the cost efficiencies offered by the ability to dynamically scale to cope with fluctuating demand;
- the business benefits & challenges of having a constantly evolving product;
- the change from a traditional pricing and distribution model to a subscription based model.

In all of the above cases, we will discuss cloud-based assessment in the specific context of the Common Core state standards and how it may help states, testing companies and educational publishers to adapt.

By the end of the session you will have a clear picture of the current challenges and opportunities faced by cloud-based assessment as well as some insight into what’s coming down the line in both the short and medium term.

PRESENTERS:

Gavin Cooney *Learnosity*
Mark Lynch *Learnosity*
Mitch Benson *Instructure*

Breakout Session

MONDAY, MARCH 2, 2015 • 4:00 PM – 5:00 PM

Opportunities in International Credentialing: Organizations, Relationships and Collaboration

Room: Rancho Room
Practice Area Division(s):
Topic: Business Strategy and Operations
Session Type: Breakout

This session will feature the experiences of two organizations that have expanded their oversight process to include the utilization of a US-developed examination for the purpose of credentialing candidates internationally. Each organization will share its experiences in meeting the challenges faced when implementing global credentialing.

The Canadian Council of Registered Nurse Regulators (CCRNR) will utilize the NCLEX-RN (developed by the US-based National Council of State Boards of Nursing) as its licensure examination for entry-level nursing practice beginning in 2015.

The Institute of Internal Auditors (IIA), an international professional association with 180,000 members in 190 countries is the internal audit profession’s global voice, recognized authority, chief advocate, and principal educator. The IIA’s experience and leadership in managing a comprehensive portfolio of certifications while working with such a vast array of international affiliates will be invaluable to attendees assessing ways to engage with global audiences and establish global credentials.

For the CCRNR, this session will outline the regulatory issues involved, the decision-making process, challenges encountered, and solutions developed. This is a historic joining of regulatory processes and the lessons learned from this partnership should be interesting to all attendees. For the IIA, this session will outline all the issues of credentialing globally through the leadership of a single organization. Again, the complexities of managing global certification programs through a local affiliate structure will be discussed and lessons learned will be shared with attendees.

PRESENTERS:

Phil Dickison *National Council of State Boards of Nursing*
Lisa Hirtzinger *The Institute of Internal Auditors*
Reno Deschaine *The Institute of Internal Auditors*

FUNDAMENTALS OF TESTING

A Systems Approach to the Design of an Adaptive Test

Room: Celebrity Ballroom H
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Breakout

Fixed-test blueprints have been used for years to describe the specific nature of a test and to provide directions for building a test for a particular purpose. They include information such as the number of items, content strands, etc. The use of this type of test blueprint was natural with fixed-form tests and was adopted for use when adaptive tests were first developed. The fixed-test administration guide was also adopted by developers of adaptive tests.

Unfortunately, the blueprint and guide developed in the context of a fixed-form test fall short when applied to an adaptive testing setting. This makes it difficult for programs designing adaptive tests to make the decisions that they need to make. This is primarily because the system of an adaptive test is so different from that of a fixed test. As examples (more are included in the full paper):

- A fixed-test blueprint specifies the number of items to be administered in each content strand to each test taker. With the advent of item forms with differing amounts of item information, this may lead to an adaptive test that has imbalanced information. Adaptive testing programs can work around this by specifying levels of information per strand, but this moves far from the original blueprint.

- A fixed-test guide doesn't include information that an adaptive testing program might be able to use across test events. For instance, item exposure isn't considered in guides, so adaptive test developers have to add external controls (such as Sympson-Hetter) to the system.

This difficulty in the use of fixed-form blueprints and guides is exacerbated when a test may be administered multiple times to a single individual (common in education, certification, and licensure). A system-based model for creating blueprint tools and administration tools that attend to the needs of an adaptive test is needed. This presentation describes such a system, including specifications of an adaptive testing system for a particular test event, a series of tests for an individual test taker, and a set of test events across a number of test takers. The study then applies lessons from thirty year of adaptive testing to describe how programs can use this approach to improve adaptive testing program design, development, and performance.

PRESENTER:

Gage Kingsbury *Psychometric Consulting*

If at First You Don't Succeed: An Analysis and Interpretation of Candidate Retake Response Patterns

Room: Oasis Ballroom 4

Practice Area Division(s):

Topic: Testing, Measurement, and Psychometrics

Session Type: Breakout

The testing industry supports and encourages examinees to be offered at least one exam retake opportunity. Credentialing programs often define and enforce specific exam retake policies that include requirements such as waiting periods, retake fees, or education/retraining. The intent of these policies is to limit candidates' exposure to the exam content for security purposes and encourage candidates to make a serious effort at the earlier attempts. Although programs typically provide failing candidates with guidance on preparing for their second attempt, there are a number of strategies that candidates may use to achieve a higher score.

The research in this area supports the idea that examinees perform better on a second attempt at an exam, but there is little empirical research providing insight as to how examinees' response patterns change or remain the same between attempts. This information can help program stakeholders better understand how candidates prepare for a second attempt at an exam and if having prior exposure to some content provides an advantage. From a program maintenance perspective, analysis of retake performance and response patterns can help identify compromised items or help detect cheating.

In this session, representatives from two credentialing programs will provide an overview of their exam retake policies, including the rationale for such policies. Presenters will explain how response patterns were analyzed for repeat test takers for both programs, including in-depth investigations of these response patterns by item response time, item difficulty, item type, and exam length.

The main focus of this presentation will be a review of the key findings that emerged from this investigation, along with the resulting impact for these programs. From a psychometric perspective, specific findings will be detailed that raise validity and security concerns. From a policy perspective, key findings will be interpreted with regards to program impact and future policy discussions. Guidance will be provided to program leaders as to why such investigations should be conducted, how to analyze response patterns, and suggestions for psychometric and policy interpretations.

PRESENTERS:

Susan Davis-Becker *Alpine Testing Solutions*

Amanda Wolkowitz *Alpine Testing Solutions*

Jared Zurn *National Council of Architectural Registration Boards*

Jack Terry *National Board of Examiners in Optometry*

Defending Cut Scores Across Various Contexts: Common Questions, Answers, and Insights

Room: Mirage Room

Practice Area Division(s):

Topic: Testing, Measurement, and Psychometrics

Session Type: Breakout

Setting cut scores (i.e., passing score/standard) is more art than science, requiring not only the "right" method but good facilitation. Because the cut score has clear implications for test takers (e.g., failing eliminates them from opportunities while passing increases those opportunities), testing programs must take care not only when selecting a standard setting method but in how they use the input from that process to establish the cut score for their exams and assessments, and they must be prepared to defend the process and the associated decisions. The perspective from which the cut score is scrutinized depends heavily on the exam's or assessment's purpose. In situations where legal challenges are more likely, such as when it's tied directly to high-stakes consequences, selecting the "right" standard setting method and applying the output of that process to set the cut score are critical. Testing programs have more flexibility when decisions are not as closely tied to high-stakes outcomes, but may have other important reasons to defend their selection of standard setting method and the resulting cut score, such as public perceptions or test user confidence.

This panel discussion brings experts from the fields of certification, licensure, employment, and educational testing to discuss: (a) typical stakeholder concerns related to cut-scores in various contexts, (b) how to select the "right" standard setting method to support the exam's purpose, (c) how to do so within programmatic and organizational constraints, and (d) how to defend the method selected and resulting cut score in various contexts—both in a court of law and in the court of public opinion. During this discussion, the panelists will share their experiences with various cut score setting methods, their effectiveness in setting appropriate cut scores, and the successes and failures that they've had in defending the associated pass/fail decisions to their stakeholders, constituents, and other interested parties. They will discuss common questions that are asked about the cut score, the challenges that stakeholders raise about the process, and who is likely to question the cut score and the process used to establish it. They will share the pros and cons of using several common cut score setting methods (e.g., Angoff, Bookmark, Borderline, etc.) in terms of defending the resulting cut score.

PRESENTERS:

Liberty Munson *Microsoft*

Lorin Mueller *The Federation of State Boards of Physical Therapy*

Steve Ferrara *Pearson School*

Kayo Sady *DCI Consulting Group*

Improving the Odds for Success of High-Potential Programs

Room: Oasis Ballroom 5-6
Practice Area Division(s):
Topic: Business Strategy and Operations
Session Type: Breakout

As economies emerge from one of the worst economic crises in decades, surfacing the talent that organisations need to realize the challenges of the post-crisis world has become an even more critical imperative. Yet, current talent practices such as high-potential programs are failing organizations with 55% of high-potential candidates dropping out of their programs within 5 years, 46% of new leaders failing to meet their business objectives and 1 in 4 high-potential candidates looking for opportunities with other organizations. This session shows the strategic value of assessment data in addressing a critical need for organizations worldwide including those in the US and Canada both public and private – identifying those employees with the aspiration, ability and engagement to be the next generation of managers and leaders. It addresses issues such as the confusion between high-performance today and high-potential for more senior roles in the future, and how assessment data answers three critical questions for high-potential programmes – will they get to a senior position, will be they be effective when they get there and will they still be with the organization when they get there?

This session is structured around short segments that invite the sharing of audience experience and challenge the audience to answer a simple but tough question from the CFO – how do we know that this development investment today will pay off in the future? For the more technically minded, the session will present new data from the one of the most extensive research programs into to what defines a high-potential employee drawing on client survey research stretching back a decade and an assessment database of 6.6 million people worldwide. For the more business oriented, the session also shows how an assessment solution can be framed and articulated for an executive audience to address a market need.

PRESENTERS:

Eugene Burke *CEB*
Emily Hill *CEB*
Sara Asady *CEB*

International Perspectives: Clinical Computer-Based Testing in the UK

Room: Oasis Ballroom 7
Practice Area Division(s):
Topic: Test Administration and Delivery Models
Session Type: Breakout

The Royal College of Paediatrics and Child Health is responsible for training and examining all paediatricians in the United Kingdom. Over the last 24 months, it has completely moved from conventional paper-based testing, to an innovative computer-based testing approach, making it a pioneer within the Royal College medical group.

This presentation focuses on the selection criteria for its computer-based testing solution (satisfying the clinician's need for an easy-to-use question authoring environment, supporting automatic test assembly, and analytics for the Angoff review panels), along with the business change needed to support the adoption of computer-based testing, and how this change has helped to improve efficiency and productivity throughout the organization.

This presentation is for clinical institutions which use, or plan to use, advanced digital testing technologies for tests which contain detailed clinical images or video, or traditionally require an examiner to be physically present to observe a candidate. Attendees will gain an insight into the Royal College medical group and how computer-based testing is being addressed within the UK, and learn about the innovative technological techniques RCPCH is implementing.

Examples of techniques that attendees will learn about include:

- Embedding clinical images within the question content that can be "zoomed in" on, allowing clearer and more detailed image-based questions that are easy to use for candidates.
- Breaking down Extended Matching Questions into multiple parts, so that metadata can be applied to each part of the question, and then analysed independently during the Angoff review session.
- Using high quality videos within grouped questions, and delivering these into test centres with weak infrastructure.

Following the successful roll-out of RCPCH's computer-based testing program, its next plan is to utilise a tablet-based observational testing tool, so that all test data can be consolidated into one place. The presentation will conclude with a demonstration of the observational testing tool, and invite discussion and comments on how their US counterparts see comparable test methods using similar technology, along with the shared challenges and experiences these test methods bring within the clinical profession.

PRESENTERS:

Jim Crawford *BTL*
Romana Moss *Royal College of Pediatrics and Child Health (RCPCH)*

Learning from Disruption – How Industries are Rapidly Evolving and the Trends They are Setting

Room: Polo Room
Practice Area Division(s):
Topic: Business Strategy and Operations
Session Type: Breakout

What does it mean to be disruptive and what can we learn from other industries that are experiencing one wave of disruption after another?

Square™ took on the financial services industries and brought credit card processing to the streets. Mint™ made managing your money cool and easy and Uber™ is zipping execs and students alike across cities.

These organizations, and others like them, took on established industries and applied innovative and disruptive processes to break through the old school thinking of how things should be done. Not all were successful, but many were able to create a new 'normal' that has infiltrated other industries including the testing industry.

Start-up organizations have raised consumer expectations in many areas including the speed and availability of customer service, a user experience that is a perfect blend of functionality and museum quality design, and of course the ability to access a service 24 hours a day, 7 days a week.

In this session we'll explore the different marketing, operations and business processes that have enabled organizations to grow and apply them to the testing industry. We'll then flip the coin and discuss how the testing industry can be a part of this disruptive movement. We'll cover the essentials of testing and demonstrate how an innovative certification program can help an organization change the paradigm, expand existing and enter new markets, raise customer satisfaction, engage partners and expand market share rapidly.

PRESENTERS:

William Dorman *Kryterion, Inc.*
David Meissner *Kryterion, Inc.*

FEATURED SPEAKER SERIES

Security at the Forefront

Room: Celebrity Ballroom G
Practice Area Division(s):
Topic: Security
Session Type: Breakout

Security continues to be one of the hottest topics for the testing industry, as demonstrated through the many presentations here at Innovations. This session is a unique opportunity to have your questions about security answered by experts in the field. A panel of experts representing the legal, communications, proctoring, remote delivery, crisis management, and forensic perspectives in security will answer your questions, moderated by the ATP Security Committee. Questions will be asked both during the session, as well as responding to questions submitted ahead of the conference. Come and get insight about the security issues specific to you and your organization!

PRESENTERS:

Jennifer Semko *Baker & McKenzie*
Jennifer McGlashon *Kryterion, Inc.*
Roddy Meade *Prometric*
Neal Kingston *University of Kansas*
A. Benjamin Mannes *American Board of Internal Medicine*
Bob Whorton *NCEES*

Platinum Sponsor Sessions

MONDAY, MARCH 2, 2015 • 4:00 PM – 5:00 PM

The College Board

PLATINUM SPONSOR SESSION

All Roads Lead to Rome: Using Multiple Selection Tools

Room: Oasis Ballroom 1

Session Type: Sponsor Session

Often in selection scenarios, candidates have scores on different instruments. For example, institutions may accept more than one admissions test (e.g., GRE or GMAT). Or, a score on one test may exempt a candidate from taking another test. Sometimes, we want to treat all scores from a test battery as equivalent (as with British A-Level exams). This talk explores ways to determine equivalent scores across multiple instruments (e.g., equating, concordance, prediction models, expert judgment), and how practitioners and consumers can interpret and use this information

PRESENTERS:

Deanna Morgan *The College Board*

Rosemary Reshetar *The College Board*

Michael Walker *The College Board*

ASSOCIATION OF TEST PUBLISHERS

MAKING THE WORLD SMALLER FOR TESTING

Prometric

PLATINUM SPONSOR SESSION

The Impact of the Updated Standards for Psychological & Educational Testing – What You Need to Know and Do

Room: Oasis Ballroom 2

Session Type: Sponsor Session

Over the past three years, members of the board of The Standards for Educational and Psychological Testing have updated the criteria developed to "promote sound testing practices and to provide a basis for evaluating the quality of those practices." This session will discuss some of the key changes to the standards to ensure your testing program meets the updated standards.

Prometric will describe major challenges faced and investments made while staying current in a new era of testing. By this new set of standards, how will you hold your vendor responsible to be more forthcoming and transparent throughout the test development and administration process?

- Delivering valid and secure tests and having a robust item bank to support this initiative.
- Concerns about item harvesting and security necessitate thinking about using alternate test delivery formats that present equivalent test forms to all candidates but limit the exposure of items.
- Linear-on-the-Fly Testing (LOFT) offers organizations the option of implementing a process that makes more efficient use of an item bank while controlling item exposure.
- Automation of form assembly and score reporting diagnostics to ensure accuracy and effectiveness of producing the best possible results for examinees while providing the candidate insight into their score while protecting the content
- Investments in testing accommodations and test development tools that automate the design, form creation and publication processes to standardize the testing landscape.

PRESENTERS:

Kenneth Ridgely *Prometric*

Steven Winicki *Prometric*

The Innovation Showcases are your chance for one-on-one interaction with a variety of presenters while you enjoy a cocktail. Each showcase is a chance for presenters to highlight a particular study, product, technology, or approach in an individual manner, sometimes using interactive media.

Innovation Showcase 1

SIFT: New Software for Investigating Fraud in Testing

Room: Ambassador Foyer
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Innovation Showcase

SIFT is a new software program that has been developed to help testing organizations investigate the possibility of fraud. It provides a user-friendly interface that allows you to select the type of analysis and collusion indices desired, and outputs them into Microsoft Excel files for easy viewing and additional work (e.g. sorting). The primary purpose of the program is to bring the psychometrics of data forensics to more practitioners, increasing the overall level of test security.

The most advanced portion of the program is the collusion indices, which include: G2; K; K*; PAIR1; PAIR2; Bellezza and Bellezza (observed and random versions); Harpp and Hogan; and Harpp, Hogan, and Jennings. Additional indices might be added by March 2015. SIFT also provides functionality for response time analysis, including the Response Time Effort index (Wise & Kong). Of course, a classical item analysis is also performed, as those statistics remain useful, and are sometimes integrated into indices.

Finally, SIFT also crosses these analyses by two levels of nested locations that users are allowed to specify, such as State and District, or Country and City. For example, you might look at the mean Response Time Effort and number of collusion-flagged examinees in each city where your tests are administered, allowing you to flag possibly compromised or problematic test centers.

The goal of the presentation is to explain the types of analysis that are available and how they can benefit testing organizations. A program such as this is not yet readily available for practitioners, so this could be quite helpful in the field of test security.

PRESENTER:

Nathan Thompson *Assessment Systems*

Innovation Showcase 2

Using Automated Essay Scoring to Mark Innovative Item Types Administered in Multiple Languages

Room: Ambassador Foyer
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Innovation Showcase

Written-response tasks are included in assessments because they allow educators to measure an examinee's ability to think, reason, solve complex problems, communicate, and collaborate through the use of writing. But written-response tasks are costly to administer and challenging to score because they rely on human scoring. One alternative to the manual scoring process is to integrate computer technology with writing assessment using automated essay scoring (AES). An AES system is a computer program that builds a scoring model by extracting linguistic features from a written-response prompt that has been pre-scored by human graders. Next the program uses machine learning algorithms to map the linguistic features to the human scores. Using this mapping, the computer can then be used to classify (i.e., score or grade) the responses for a new group of examinees.

Research outcomes have demonstrated consistent results: AES systems yield scores that agree with human raters at a level as high, if not higher, than the agreement among human raters themselves. Therefore, AES provides a variable method for scoring written-response tests that complements the current use of selected-response (e.g., multiple choice) testing. This session will present results on the application of AES for scoring clinical decision-making questions required as part of a high-stakes medical licensure examination administered to both English- and French-speaking examinees. The clinical decision-making question requires examinees to write a concise and clear written response to a clinical problem-solving prompt. Currently, these questions are hand scored by human graders.

In the study presented in this session, 8,007 English- and French-speaking examinees responded to 8 different clinical-decision making questions. Their responses were scored using an AES system and with human graders. An agreement rate of 95.4% was produced between the computer and human scoring across both language groups. AES offers educators many benefits for scoring constructed-response tasks, such as improving the consistency of scoring, reducing the time required for scoring and reporting, and minimizing the costs of scoring. Based on this research, the presenters of this session will describe how AES can be used to score a unique item type — clinical decision-making questions — and how it can be used to score the written responses from examinees in multiple language groups.

PRESENTERS:

Mark Gierl *University of Alberta*
Andre De Champlain *Medical Council of Canada*
Andre-Philippe Boulais *Medical Council of Canada*

Innovation Showcase 3

Innovation in Testing – Business Strategy Simulator

Room: Ambassador Foyer
Practice Area Division(s):
Topic: Program Design, and Test Administration and Delivery Models
Session Type: Innovation Showcase

Gamification – the word has been doing round in the learning circles as well as testing. The idea behind Gamification in learning/testing is simple – to deliver and deploy learning/testing knowledge or skills via a game – making the mundane experience a fun and interactive process which will also bring out the full potential of the user. This presentation will bring in a brand new perspective to testing business acumen via gamification.

How it is possible to take the whole process, of analysing a business problem statement, data interpretation, investigation, problem solving and guess-estimation, business strategy planning, online and to make it interactive and how to scientifically measure the above mentioned skills. The presentation will also talk about how the user can be made to experience being in the shoes of a CEO, or a Consultant, or a Marketing Manager – emulation based testing. Snapshots of the platform in action, performance measurement and Case Studies will also be displayed. The presentation will include details on program design, philosophy and the logic behind it, scoring logic and weight-age, interpretation of results, automating the above process, integrating all business angles to one framework (Finance, Marketing, HR, Sales and Executive Management), how to deliver it to candidates in a secure and authentic fashion.

PRESENTER:

Ketan Kapoor *Mettl, Inc.*

Innovation Showcase 4

The Generated Response Option Item

Room: Ambassador Foyer
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Innovation Showcase

The multiple choice item is a fixture on the standardized testing landscape. This is so, in part, because of relative ease with which multiple choice items can be developed, administered, and scored, and the information they provide relative to testing time. Despite these benefits, the multiple choice item is not without limitations. It is often justly criticized for reducing the cognitive task to one of selection instead of generation, a reduction that can sacrifice fidelity to the construct being assessed. Additionally, its quality is dependent on subject matter experts' ability to produce compelling distractors, a challenge that, when unmet, results in increased costs as items must be revised or discarded.

The short answer item, a common alternative to multiple choice, overcomes these limitations while introducing a few of its own. In its favor, the short answer item allows examinees to generate their own responses, simultaneously maintaining fidelity to the construct and simplifying item development by eliminating the need for distractors. While a short answer item may prove more challenging to an examinee, it is not for lack of understanding of the item type. The primary challenge comes in scoring. By not fixing the universe of possible responses, an item will frequently have multiple variations on a correct response, all of which must be identified in order to properly score the item. In practice, this challenge is met either by having humans manually score the items or by using an automated scoring procedure. Both are sources of measurement error and expensive to implement.

We offer an alternative approach. By predefining a universe of plausible responses and employing simple search algorithms, it is possible to take from the best of both of these items types to produce a new one without the limitations of either. Examinees interact with an item by generating text, which is matched against stored plausible responses, which are, in turn, listed as options for the examinee to select. On a geography item, for instance, one could type "guinea" and be presented with selection options of "Guinea", "Guinea-Bissau", "Equatorial Guinea", and "Papua New Guinea". This approach requires generative processes while allowing for instant, error-free scoring. Our presentation will offer an interactive demonstration of this new item type and how it works.

PRESENTERS:

Robert Cook *American Board of Internal Medicine*
Jerome Clauser *American Board of Internal Medicine*

Innovation Showcase 5

Generating Test Items Pre-Aligned to Common Core Standards

Room: Ambassador Foyer

Practice Area Division(s):

Topic: Testing, Measurement, and Psychometrics

Session Type: Innovation Showcase

Testing companies now require large numbers of test items to operate their assessment programs. With formative testing, not only are large numbers of items required, but the specific knowledge and problem solving skills to use must also be defined. Item generation combines cognitive modeling with computer technology to allow testing companies to produce large numbers of test items that measure specific knowledge and skills.

This session will present the results of a study that demonstrates how a modification to the cognitive modeling process can be used to produce large numbers of items that meet the Common Core Standards to facilitate item generation for formative testing. In the study, content experts will compile a list of appropriate Common Core Standards that were either related to a parent item or could be expressed in the parent item through modifications. The list of standards will be selected so that they spanned different content domains and grade levels. Then, during the cognitive modeling process, content experts will define how each standard could be expressed within the cognitive model. The expression of a standard is represented by the use of elements and constraints within the cognitive model. During item generation, a vector of codes is used to represent the combination of standards as expressed in each item. Using this process, items can be generated with predefined knowledge and skills required to solve each item. To-date, twenty models have been created using this process for Common Core Standards in mathematics across four grade levels.

This session will demonstrate how information can first be captured from content specialists, how content is then translated onto generated test items, and how the generated items can finally be assembled into formative tests based on the corresponding standards. Public accountability has been a driving force for more frequent testing to meet different purposes. But item development for formative tests requires a substantial amount of planning, indexing, and categorizing required knowledge and skills within a specific content area. This session will demonstrate how a new method of producing test items that involves collaboration between content expertise and technology can generate large numbers of items that meet specific Common Core Standards. This development method can be easily replicated to meet the test item needs in other content areas.

PRESENTERS:

Hollis Lai *University of Alberta*

Jim Hogan *ACT, Inc.*

Mark Gierl *University of Alberta*

Innovation Showcase 6

International Differences in Personality: Smaller than Occupational Differences

Room: Ambassador Foyer

Practice Area Division(s):

Topic: Testing, Measurement, and Psychometrics

Session Type: Innovation Showcase

Differences in translations of a personality questionnaire could arise from the translation process, cultural differences and different samples. These issues can be complicated further if the questionnaire includes questions designed to obscure their target from the test taker. This is the case with the 16PF® Questionnaire. The publishers of the 16PF assessment wanted to learn more about these differences, and thus embarked on a study to establish the equivalence of a range of different 16PF languages (e.g. UK, French, Dutch), using different English language versions as a benchmark (e.g. UK and US), and to see if the creation of a single, international norm group for the 16PF was possible. Along the way, a comparison of differences in language/cultures with in different occupational groups was also studied.

Study results show some surprising conclusions – that the differences between languages are not due to translation issues, but due to genuine cultural differences. These differences, however, were small when compared with two other yardsticks: the Standard Error of Measurement and again, surprisingly, occupational group differences.

PRESENTERS:

Pamela Becker *IPAT, Inc.*

Penny Moyle *OPP*

Innovation Showcase 7

Breadth is Better: A Cross-Industry Assessment Platform

Room: Ambassador Foyer
Practice Area Division(s):
Topic: Data Management
Session Type: Innovation Showcase

The majority of test development and delivery platforms are built specifically for a single organization, or at least for a single ATP Practice Area Division. However, the vast majority of functionality in test development and delivery is the same for all Divisions: follow best practices in setting a test blueprint, author and review items, assemble test forms, publish and deliver, provide reports, and manage results. This session will feature a comprehensive system that the presenters have developed that manages these processes in a way that is applicable to all Divisions. The system accomplishes through its focus on best practices in test development and psychometrics and implementing those directly into its functionality, rather than developing a system based on a certain type of test. The presenters will discuss the various applications of this system through discussion of the multiple clients from all ATP Divisions using the system, including:

- Education—school districts, non-profit organizations, and for-profit K–12 assessment companies
- Certification/Licensure—non-profit credentialing organizations and healthcare companies with in-house certifications
- Industrial/Organizational—preemployment selection organizations, 360-degree assessments
- Clinical—Addiction survey, clinical psychopathology inventory

The presenters will discuss technology, design, psychometrics, and cross-industry considerations related to their system in this innovation showcase.

PRESENTER:

Keith Morical *Assessment Systems*

Innovation Showcase 8

What is Workforce Skills Credentialing?

Room: Ambassador Foyer
Practice Area Division(s):
Session Type: Innovation Showcase

Come learn about and contribute to ATP's newest division related to Workforce Credentialing. The scope of workforce credentialing is growing and the associated benefits of measuring candidates for credentials is gaining interest – come add your voice to the discussion.

PRESENTER:

Dan Rinn *NTT, Inc.*

Innovation Showcase 9

The World is VUCA for Public Sector Assessment and Selection!

Room: Ambassador Foyer
Session Type: Innovation Showcase

Against the backdrop of these Volatile, Uncertain, Complex and Ambiguous times, and in order to do battle in the global War for Talent, the international public sector has realized it needs to stand up and be counted in the world of recruitment and testing! By setting up this dedicated Special Interest Group (SIG) as part of the world's leading testing association, members will be able to help each other meet the challenges they face.

The testing community can play a significant role in helping public sector bodies meet these challenges, not least because they need to become leaders in high quality, fair and cost effective assessment tools as well as in modern technology. CBT, IBT, psychometric, language and skills tests as well as videos, animation, serious gaming tools and other innovations - usually seen as the preserve of the private sector - need to be taken up by the public sector if it is to have attractive, modern, efficient, accountable, fair and cost-effective procedures.

This showcase will present to you what this SIG has achieved so far in its infancy, who the organizations are that have signed up to the challenge – the UN, NATO, European Union, national governments across the globe to name but a few, and where we feel the world of testing can help.

PRESENTERS:

Paula Chadwick *European Personnel Selection Office*
David Bearfield *European Personnel Selection Office*

Reception with Exhibitors & Innovation Showcases

MONDAY, MARCH 2, 2015 • 5:00 PM – 7:00 PM

Innovation Showcase 10

ATP's New Global Structure

Room: Ambassador Foyer
Session Type: Innovation Showcase

When you join ATP you are part of an International network of members committed to advancing the integrity of assessment services and products and their value to society. Learn about ATP's new structure as a Global Association! Regional Divisions, Global Practice Area Divisions, Committees and Special Interest Groups provide added networking opportunities for members and provide a forum for the unique regional needs of assessment professionals in different geographic areas.

PRESENTERS:

Lauren Scheib *Association of Test Publishers*

Pam Patel *Association of Test Publishers*

Innovation Showcase 11

How to Volunteer for an ATP Conference Committee

Room: Ambassador Foyer
Session Type: Innovation Showcase

Thought about volunteering for ATP but were not sure of what the opportunities might be or how the process might work? Stop by to discuss volunteering. It is a great way to contribute to the profession and to network with colleagues throughout the testing industry.

PRESENTERS:

Erwin van Schaffelaar *Cito*

Eric D'Astolfo *Pearson*

Reception with Exhibitors & Innovation Showcases

MONDAY, MARCH 2, 2015 • 5:00 PM – 7:00 PM

Innovation Showcase 12

Tips and Insights for Submitting Sessions for ATP

Room: Ambassador Foyer
Session Type: Innovation Showcase

ATP continues to attract great sessions for its conferences, and we would like yours to be among those selected! Whether you have had sessions submitted, or you are thinking about submitting an abstract for 2016, this showcase presentation will walk you through the steps required to develop and submit acceptable sessions. You will learn how to write a great abstract, write learning objectives, from your session around the theme of Innovations in Testing, as well as get tips for navigating through the submission process.

PRESENTERS:

Ashok Sarathy *Graduate Management Admissions Council*
Rachel Watkins Schoenig *ACT, Inc.*

Breakfast / Workshops

TUESDAY, MARCH 3, 2015 • 7:30 AM – 6:00 PM

7:30 AM – 6:00 PM

Speaker Lounge Open

Room: Mission Hills Boardroom

Are you presenting at the Innovations in Testing? Stop in to practice your presentation, collaborate with co-presenters, or just rest for a while. This room is open to all conference presenters.

7:30 AM – 8:15 AM

Security Committee Breakfast Meeting – All are Welcome!

Room: Oasis Den

Join the ATP Security Committee leadership to hear about recent progress, current and future initiatives, and the many ways to get involved with the committee. This is a great chance to learn more!

Clinical Practice Area Division Breakfast Meeting – All are Welcome!

Room: Celebrity Ballroom A

Join the ATP Clinical Division for an update and discussion related to the current goals and activities of the division.

7:30 AM – 8:30 AM

Breakfast with Exhibitors

Room: Ambassador Ballroom

8:30 AM – 10:00 AM

Building Assessments with New Technologies: The Good, the Bad, the Cloud, and the Free

Room: Celebrity Ballroom C

Practice Area Division(s):

Topic: Data Management

Session Type: Workshop

Individuals in the assessment industry are constantly asked to reduce costs and development time. Many have cut resources to the bone and need to find other solutions to reduce costs. One of the most costly aspects of exam development is subject matter expert (SME) travel during item development, item review, and Angoff sessions.

This session will focus on the technologies, tools, and cloud services that can be used to collaborate in various phases of content development. The content of the session will explore remote connectivity solutions, content storage solutions, collaboration tools, secure cloud storage, and other technology solutions used to develop content with SMEs located all over the world. This session will introduce attendees to innovative methods of using common software and to innovative software, services, and related processes that can be used to add efficiency to any development model. This session will provide useful information to all individuals looking to reduce content development costs, increase the speed of content output, or reduce delays in collaboration. Prior knowledge of technical solutions is not required for this session; presenters will ensure that baseline knowledge is established with all attendees.

PRESENTERS:

David Elfassy *Netlogon*

Don Tanedo *Microsoft*

Innovation Challenge

Room: Celebrity Ballroom D-E
Practice Area Division(s):
Topic: Business Strategy and Operations
Session Type: Workshop

It's different. It's engaging. It's edible. We're not kidding! Join us for 90 minutes of interactive fun as we talk about how we can leverage strengths within our own organizations to better meet customer needs and more successfully innovate in new product and market spaces. Come with a 4 person team from your organization, or join us in 1s, 2s or 3s and we'll match you up at the workshop. You'll come away with some new ways to challenge your thinking and build great teams within your organization to tackle the emerging challenges in testing!

PRESENTERS:

Rachel Watkins Schoenig *ACT, Inc.*

Ashok Sarathy *Graduate Management Admissions Council*

Trust But Verify: Exam Integrity Audits Made Easy Through The Adoption of Technology

Room: Celebrity Ballroom F-G
Practice Area Division(s):
Topic: Security
Session Type: Workshop

The pace of technology is ever increasing and new tools become available, both to the industry and to those seeking to profit from fraudulent activities. If you've ever had to investigate an incident of possible cheating, or sourced a complaint of exam failure due to environmental distractions, you know supporting evidence is imperative to resolution.

A technologist will provide insight into the various technical options available for incident asset recovery and allegation verification on kiosk and online exam delivery and proctoring platforms. We will also hear from a regulatory agency and licensure program director using both traditional testing centers and kiosk technology to proctor Illinois Department of Public Health Exams. He will share some incident reporting resolution practices applied in actual occurrences.

PRESENTERS:

Lily Soto *Innovative Exams*

Ron Rodgers *Continental Testing Services (CTS)*

How to Identify, Develop and Implement Innovative Items

Room: Celebrity Ballroom H
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Workshop

In the last few years, interest in using innovative item types has flourished and many smaller exam programs that were formerly invested solely in multiple choice items are now considering a fundamental change in their assessment.

This growth in interest is fueled by the availability of innovative item types as part of existing CBT software and by their successful use in larger exam programs. However, there is no guarantee that an alternative item type will be successful every time it is used. When exam programs add innovative items without sufficient preparation, the quality of the new items can be inadequate, and may contribute to an unfortunate decrease in the quality of the exam. Therefore, a thoughtful, well-planned approach for getting started with innovative item types is the best way to reduce the risks associated with new item types, and to increase the likelihood that the potential benefits they offer will be realized.

This workshop will provide attendees with a plan for designing and developing innovative item types. The workshop content will be aimed at measurement professionals who are new to the use of these item types. The innovative item types to be covered during the workshop include: hot spot, short answer, drag-and-drop, and audio item types.

The information will be accompanied by small group activities structured around the development of the four innovative item types. While these item types have been effectively used on numerous exams, it is also true that every item type has certain strengths and weaknesses. These will be addressed along with guidelines for selecting optimal item types for measuring specific content areas and cognitive levels. In addition, these four item types will be used to discuss a range of test development considerations that can arise with innovative item types.

Workshop participants will engage in the development of innovative items and will have the opportunity to discuss their experiences openly through facilitated group discussion.

By the end of the session, attendees will have a strong understanding of all four innovative item types presented during the workshop. Workshop participants will also have a more complete understanding of the development process for innovative item types, as well as ideas for how to select and implement innovative item types in their own examination programs.

PRESENTERS:

Cynthia Parshall *CBT Measurement*

Adrienne Cadle *Professional Testing, Inc.*

FUNDAMENTALS OF TESTING

Exam Security Incident Response Workshop

Room: Oasis Ballroom 1-3
Practice Area Division(s):
Topic: Security
Session Type: Workshop

Now more than ever, test sponsors must employ sensible strategies to identify and respond to exam security incidents that may threaten the integrity of their examinations. During this interactive workshop, panel members will facilitate an interactive forum to discuss effective strategies that test sponsors can implement to identify and respond to the most frequently encountered categories of exam security incidents experienced in high-stakes examination programs, including higher education admissions, licensure, certification, and statewide standardized education assessment.

First the presenters will identify useful resources for attendees to draw upon when responding to security incidents. Next, the presenters will introduce key concepts that should assist attendees in working through a series of 4 table top exercises relevant to the learning objectives. These exercises will consist of small group discussions of "real world" case studies that illustrate the key concepts. The exercises will require participants to: (1) decide which incidents to investigate and what resources to allocate to the investigation; (2) devise a response and investigation plan to gather sufficient information and evidence to make decisions about next steps regarding test scores and examinees related to the incident; and (3) generate a key lesson learned from the exercise.

Following the completion of the exercises, the presenters will answer questions about major issues relevant to the case studies that test sponsors are presently dealing with in exam security investigations.

PRESENTERS:

Marc Weinstein *Dilworth Paxson LLP*
A. Benjamin Mannes *American Board of Internal Medicine*
Aimee Hobby Rhodes *CFA Institute*
John Fremer *Caveon Test Security*

The Top 5 Mistakes in Performance Testing (and how to avoid them)

Room: Oasis Ballroom 5-6
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Workshop

Human-scored performance items offer measurement options beyond those of standard, selected-response item types, but require careful planning. This session will identify the most common mistakes made during their development and administration, as well as practical, real-world strategies for how they can be avoided. A case study and interactive exercise will accompany the presentation.

Methodology:

- The presenters will survey the top 5 mistakes in performance testing. Scott Russell and EW Looney will provide conceptual information behind each mistake, while Jack Terry will provide real world examples for dealing with these mistakes.
- Presenters will engage in a question and answer period.
- An interactive exercise will be conducted in which audience members will be given the opportunity to evaluate and discuss the practices of a fictitious testing organization.

PRESENTERS:

Scott Russell *Alpine Testing Solutions*
E.W. Looney *BrightLink*
Jack Terry *National Board of Examiners in Optometry*

Workshops / Hackathon

TUESDAY, MARCH 3, 2015 • 8:30 AM – 10:30 AM

Deepen your Knowledge of Depth of Knowledge (DOK): A Crash Course in the Appropriate Use of DOK

Room: Polo Room
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Workshop

In the late 1990s, Dr. Norman Webb developed the depth of knowledge (DOK) framework for analyzing content complexity of curriculum standards and assessments. The DOK system was designed for use by content experts who have completed training in the appropriate use of the system. An assigned DOK is intended to be a mean calculated from the input of multiple trained reviewers. Reliability is attained through calculation of the consistency of agreement of a group of reviewers. For more informal use, the DOK framework is a valuable tool for understanding the complexity of learning objectives and associated assessment items.

People often compare DOK with Bloom's taxonomy. Both are language systems that are used to describe content or cognitive complexity—but the two systems have important differences. Bloom's taxonomy has long been, and continues to be, a useful tool for many. There are, however, some major challenges to using the taxonomy from the perspective of current understandings of learning. One challenge is the organization of levels of complexity as a taxonomy that imposes an increasing value to the higher levels. Another challenge is that Bloom's taxonomy is centered in actions or verbs, but we know that complexity depends on additional factors, including the structure of the subject matter and student characteristics, including prior knowledge. The taxonomy also does not provide a good rationale for the number of levels—six—and why this number is needed for its intended purposes. The DOK system, driven by curriculum needs, addresses these concerns—particularly for analyzing curriculum standards and assessments.

Following a 50-minute overview of the DOK framework, attendees will have the opportunity to work in small groups and practice applying DOK to sample items. Small- and large-group facilitated discussions will help attendees develop and strengthen their understanding of DOK and their confidence in using this valuable tool to evaluate content complexity.

PRESENTER:

Sara Christopherson *Wisconsin Center for Education Products and Services (WebbAlign)*

Room: Rancho Mirage Room
Session Type: Workshop

Prior Registration Required

Show the ATP community your real skills and launch a brand new certification program. In this master workshop class, your voice will be heard by three top notch industry leaders, Marten Roorda (Cito), Clarke Porter (Pearson), and Peg Jobst (GMAC).

Six teams will be created to challenge each other in a collegial learning environment with participant skills ranging from newcomer to expert. You will have an opportunity to:

- collaborate with your colleagues
- learn about and create a high level certification program business plan
- learn from industry experts
- present the plan in a fun environment for critical review

10:00 AM – 10:30 AM

Break with Exhibitors

Room: Ambassador Ballroom

Teaching an Old Dog New Tricks: Using Principled Assessment Frameworks to Aid in Standard Setting

Room: Celebrity Ballroom C
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Workshop

Principled Assessment Frameworks (PAFs), like Evidence-Centered Design (ECD), offer the opportunity to innovate test development and validation practices. PAFs focus on the properties of measurement scales and the Knowledge, Skills, and Abilities (KSAs) associated with differential levels on the scale. The College Board has used ECD to guide the redesign of several Advanced Placement (AP®) exams. The current work seeks to extend the use of PAFs in this regard. As the purpose of standard setting is the categorization of examinees into distinct groups, this presentation discusses the ways in which incorporating the infrastructure of PAFs into existing standard setting practices can aid both facilitators and panelists in establishing valid cut scores for examinee classification.

The components of PAFs (notably Construct Maps and Task Models), focus on the clear delineation of proficiency claims, the requisite evidence that must be demonstrated by an examinee to support those claims, and the components of tasks that allow for the elicitation of said evidence. Specifically, Construct Maps (CMs) are visual representations of the alignment of proficiency claims, tasks, and the score scale. Additionally, Task Models (TMs) describe item features that provide evidence in support of proficiency claims. Through the use of CMs and TMs, it is possible to provide two concrete benefits to facilitators and panelists in standard setting. First, CMs and TMs aid in clearer understanding of the hypothetical “borderline examinee” on which panelists are to base their judgments. Second, TMs aid in bridging the gap between Achievement Level Descriptors (ALDs; which are typically written to describe “the big picture” of examinee KSAs) and the test items that are the unit of analysis for panelists in test-centered standard setting methods.

This presentation will be of interest to anyone looking to provide support to panelists as they engage in the recommendation of cut scores. Currently, standard setting for the AP® exams uses a Modified Angoff approach, but the work herein is applicable to other standard setting methods. Discussion will focus on major issues encountered and lessons learned. Examples of Task Models and a Construct Map will be provided in the context of how this information may be used by panelists.

PRESENTERS:

Matthew Burke *The College Board*
Robert Furter *UNC Greensboro*
Pamela Kaliski *The College Board*
Deanna Morgan *The College Board*

FUNDAMENTALS OF TESTING

Test Security on a Dime: Tips and Tools for a Cost-Effective Test Security Program

Room: Celebrity Ballroom F-G
Practice Area Division(s):
Topic: Security
Session Type: Workshop

An effective test security program can mitigate threats to assessments and ensure value and integrity in the results for stakeholders and examinees. This interactive workshop session will explore cost-effective tools and suggestions for developing, implementing and maintaining a test security program that is appropriate for different types of assessments. The workshop will include an interactive discussion on processes and tips to identify the most concerning threats to a program and how to mitigate those threats with limited resources.

PRESENTERS:

Camille Thompson *ACT, Inc.*
Rachel Watkins Schoenig *ACT, Inc.*
Bob Whorton *NCEES*

Don't Just Check the Box - Authoring TEIs Made Easy

Room: Celebrity Ballroom H

Practice Area Division(s):

Topic: Testing, Measurement, and Psychometrics

Session Type: Workshop

The Race to the Top Assessment Program and the introduction of the Common Core State Standards has resulted in an increased demand for new technologies and features in assessments. One of the most significant implications for online assessment has been the debut of technology enhanced item types. (TEIs).

TEIs are question types which go beyond the traditional selected-response or constructed response collection and instead require specialized interactions from test takers e.g. ordering a group of elements chronologically by dragging and dropping, picking out key words in a paragraph by highlighting, manipulating graphs & charts so that they match a stated function, etc.

While the use of TEIs have many benefits, there tends to be a higher cost involved in creating TEIs than in creating traditional questions. This is generally because a significant degree of technical expertise is required to create these more advanced question types.

In this session, we look at how to lower the bar for assessment authors - moving from a developer to subject matter expert with little or no technical training.

We will also discuss some of the common challenges we have encountered when creating interactive, online assessments:

- requirements that online tests be the same as the paper and pencil test and associated tradeoffs
- available technology features driving item design
- creation of new learning scenarios with poorly designed TEIs
- metadata and interoperability requirements

Join us for a fun and informative interactive session and see for yourself just how easy it can be to create complex TEIs from scratch. No need to create on paper or in Word or Excel. Create, review and publish directly from one authoring environment.

No prior experience required, but bring your laptop if you want to get involved!

PRESENTERS:

Rachel Murray *Learnosity*

Jennifer Lawrence *Houghton Mifflin Harcourt (Riverside)*

Successfully Taking Employment Assessments into the Global Market Place

Room: Oasis 5-6

Audience:

Session Type: Workshop

This workshop will address three of the five P's of Marketing in the context of going global with assessment tools utilized for employment purposes: Product (e.g., content, comparative psychometrics, and translations), Place (e.g., country/cultural testing practices/compliance, assessment trends/opportunities, and platform/technology considerations), and People (e.g., HR practitioners/end users, local consultants, and partners/distributors). The two remaining key aspects of marketing, Price and Promotion, would be woven into the presentations as relevant.

Each of the workshop facilitators will lead the presentation on one of the core topics drawing on unique professional experiences and expertise along with pertinent research and case studies. Attendees will have opportunities to apply the concepts and insights gained to their own global assessment aspirations through small group exercises as well as to share their knowledge and perspectives by discussion of best practices and lessons learned along the way.

PRESENTERS:

Gary Behrens *General Dynamics IT (Moderator)*

Eugene Burke *CEB/SHL*

Penny Moyle *OPP*

Ryan Ross *Hogan Assessment Systems*

We're Not in Kansas Anymore: Supporting International Expansion with Best Known Security Controls

Room: Oasis Ballroom 1-3
Practice Area Division(s):
Topic: Security
Session Type: Workshop

This workshop is for individuals generally interested in new approaches to supporting global growth, but will focus specifically on exam security in three tremendously promising markets: China, India and Nigeria. The opportunities for testing internationally are clear. Exam security plays a critical role in delivering good quality growth. This workshop will present exam security controls which must be implemented to preserve the integrity of your exams in these places, and generate increased confidence in test administration.

During the workshop, an approach to quantifying operational risks will be discussed. A process for using risk assessments to inform expansion planning will also be introduced. The presenters will provide their insights and solutions that address key risks. Workshop presenters include experts who have helped testing companies overcome difficult security challenges, and achieve substantial growth. In the spirit of cooperation and collaboration, participants will be asked to interact with the Presenters and each other.

PRESENTERS:

Dennis Maynes *Caveon Test Security*
James Tunkey *IonAsia*
Aimee Hobby Rhodes *CFA Institute*

Measuring, Monitoring and Improving the Quality of Item Pools

Room: Oasis Ballroom 4
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Workshop

It is critical for successful testing programs to maintain item pools that are balanced against content and statistical specifications in test plans, and large enough to support test administration and security demands. A plan to ensure the quality of the pool is critical to meet this goal. The organizations leading this session have either maintained the rigor of their own high stakes assessments or managed the review of pools for clients in education, certification, and licensure.

Pool quality can be evaluated by several criteria: balance (deviation from target at every blueprint level); depth (variants of items available at every level); currency (reflects current best evidence); blueprint specificity (consistently produce content and statistically acceptable exam forms); focus (items measure appropriate cognitive ability); and statistical attributes (ability to meet desired statistical characteristics).

After collecting these data, focus shifts to improving pool quality and size. Tactics included targeting item writing assignments to specific shortages and item modeling to increase pool depth. Strategies to increase pretest capacity were also used, including: a methodology for estimating IRT parameters from CTT statistics, adjusting statistical screening criteria, and using not-yet-pretested items on live exams. As the cost of generating items increases, it is important to have a strategy for keeping high quality items in the pool.

To assess pool quality, representatives from the first program have used tools such as Automated Test Assembly to determine how many non-overlapping, balanced forms could be created from pools.

Representatives from the second program have worked with partners in education and credentialing to implement policies and procedures to ensure pool quality and identify long-term needs for their item pools. Tools used have included data visualization techniques that have aided in communication with key constituents.

The session will have three parts:

- Presenters will share their experience with measuring, monitoring, and improving pool quality
- Participants will work in small groups to: discuss the presentation and share their experiences; brainstorm about pool metrics that might be useful; and brainstorm about other approaches to pool improvement and maintenance
- Each group will report on key points from their discussion

PRESENTERS:

Jeremy Dugosh *American Board of Internal Medicine*
Louis Grosso *American Board of Internal Medicine*
Andrew Wiley *Alpine Testing Solutions, Inc.*

Innovations Hackathon Part 2

Room: Rancho Mirage Room
Session Type: Workshop

Prior Registration Required

This is a continuation of the Innovation Hackathon – Part 1 from the earlier Workshop timeslot. Participants must have attended Part 1 in order to participate in Part 2.

The following Product Demonstrations will take place in Celebrity Ballroom B. During the Workshop timeslot, you may choose to see all of these demonstrations.

Product Demonstration 1
Simulation Development with Hammer

Room: Celebrity Ballroom B
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Product Demonstration

Simulations can be a very effective tool in performance-based testing. However, historically simulation development for assessment has been fraught with challenges, not least, significant cost implications and now the deprecation of Flash as a suitable tool for simulation development.

This presentation will provide an overview of such challenges, recommendations on how to address these challenges, along with a demonstration of Hammer, an HTML5-based simulation development tool.

Whether new to performance-based testing or experienced with it; whether considering Hammer for development or not; program managers, content developers and psychometricians will all gain insight on the successful development of simulations for assessment.

PRESENTER:

Sinéad Hogan McSweeney *Certiport*

Product Demonstration 2
MetaCat® SaaS Assessment Platform

Room: Celebrity Ballroom B
Practice Area Division(s):
Topic: Data Management
Session Type: Product Demonstration

The MetaCat Assessment Platform, by MetaLogic, Inc., offers a powerful, comprehensive assessment platform as a hosted SaaS service or with on-site installation. Complete with technology-enhanced item types, rigorous workflow options and fine-grained security, MetaCat's modular-based pricing can significantly reduce technology costs as well as increase efficiency in the assessment development and review processes. MetaLogic has been hosting item banks in the K-12 and higher education markets since 2000, and has provided online testing to millions of K-12 and higher education students since 2006.

PRESENTER:

Sara Moshman *MetaLogic, Inc.*

Product Demonstration 3
Curious About Taking an Online Proctored Exam?

Room: Celebrity Ballroom B
Practice Area Division(s):
Topic: Test Administration and Delivery Models
Session Type: Product Demonstration

A “live” online exam candidate experience will be presented. Our team will consist of the narrator, proctor and a client who we met at ATP 2012. To create a true experience, we will have a live proctor available to answer questions from delegates. The narrator will demonstrate with the assistance of the proctor the 3 steps to take an exam online: registration, system check and starting the exam.

PRESENTER:

Julia Chapelle *Loyalist Exam Services*

Product Demonstration 4

Not Ready To Switch Over To Online Proctoring Quite Yet? Learn About The Kiosk Proctoring Technology and How It Differs

Room: Celebrity Ballroom B
Practice Area Division(s):
Topic: Test Administration and Delivery Models
Session Type: Product Demonstration

There's a full spectrum of remote proctoring options. Some offer more security and some offer a wider reach. We'll demonstrate the Kiosk Proctoring Technology and how it differs greatly from what many generalize as "remote proctoring" or "online proctoring." In fact, kiosk proctoring technology adds sophisticated security measures that raise traditionally accepted security standards.

PRESENTER:

Lily Soto *Innovative Exams*

Product Demonstration 5

Scorpion: A Collaborative Exam Review Tool

Room: Celebrity Ballroom B
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Product Demonstration

Great tests require great collaboration from great people. Scorpion (testscorpion.com) opens up new and secure ways for project managers to collaborate remotely with subject matter experts at every stage of exam development. From blueprinting to forms assembly, Scorpion allows you to collect targeted feedback about your exam-to-be and leverage its powerful features to increase exam quality with an ease you never thought possible.

Depending on each audience's preference, the presenter will guide attendees through one of three tasks: blueprinting, item review, or forms assembly.

1. For blueprinting, the presenter will invite attendees to use Scorpion's simple review interface to rate the importance of a few sample objectives. The group will then use this data to automatically assign item counts.

2. For item review, the presenter will import a few entertaining trivia items, and the group will use Scorpion to review the content of the items and rate their difficulty according to multiple proficiency levels. This innovative approach to standard setting is known as Information Integration Theory (IIT), and attendees will have the opportunity to see how the process validates itself in real time as data is collected.

3. For forms assembly, the presenter will demonstrate how easy Scorpion makes it to set multiple cut scores at once and create test forms that are balanced on both content and difficulty.

PRESENTER:

Nathaniel Foster *Caveon Test Security*

Product Demonstration 6

A Statistical Software to Conduct Data Analysis for Test Security

Room: Celebrity Ballroom B
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Product Demonstration

Statistical analysis of test results is the most widely used approach employed by test sponsors. Different statistical methods can be used to capture the signs of security breaches and to evaluate the validity of test scores. The proposed software is a statistical tool that provides systematic and comprehensive analyses in test security. This software package provides applications of a variety of existing methods for analyzing test results and detecting cheating behaviors. The package can be used for: (1) screening spuriously low scores, (2) identifying potentially compromised items and identifying examinees who had an advantage in testing from preknowledge of those items, (3) identifying aberrant responses through use of several person-fit indices, and (4) identifying pairs or clusters of examinees who present similar response patterns through use of several IRT and non-IRT similarity indices.

PRESENTER:

Jiyoon Park *Federation of State Boards of Physical Therapy*

Product Demonstration 7

Multilingual Language Test Development: Best Practices and Unique Challenges

Room: Celebrity Ballroom B
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Product Demonstration

This presentation will describe, in case study format, best practice during the needs analysis, specification creation, test development, rater training, field testing, data analysis, and launching of a suite of parallel language proficiency tests for use in situations where bilingualism or multilingualism is a mandated requirement by job description, organization policy, or government legislation. The creation of language assessments for bilingual or multilingual populations imposes unique requirements on test designers and stakeholders to ensure balanced collaboration, access, and representation by all relevant language groups at key points during the testing process.

When testing language proficiency it is insufficient to translate test items from one language to another and assume equal performance of the test items and test forms. Thus, this presentation will advocate for the adoption of two guiding principles in the creation of language proficiency tests for bilingual or multilingual environments: 1) Parallel yet separate, and 2) Open and equal access. During the presentation, attendees will learn how and when language-specific test development teams operated separately and when and why they came together at specific points to share progress and ideas for mutual benefit.

PRESENTER:

Lauren Kennedy *Second Language Testing, Inc.*

Product Demonstration 8

Good Practices for Securing IP for ONLINE Exams

Room: Celebrity Ballroom B
Practice Area Division(s):
Topic: Security
Session Type: Product Demonstration

Loyalist Exam Services understands that organization's exams are an asset and must be protected. This session will provide a presentation of Loyalist Exam Services' Online Exam Process that provides exams securely in an online environment. Specifically the presenters will address the following security concerns and best practices related to:

- Authentication
- Cheating
- Proctors/Invigilators

Loyalist Exam Services will provide evidence for the success of their practices by detailing their success in data management, number of exams processed, number of years in business, and continual growth in the market.

PRESENTER:

Mike Miller *Loyalist Exam Services*

Lunch / Fishbowl Session

TUESDAY, MARCH 3, 2015 • 12:00 PM – 2:15 PM

12:00 PM – 1:00 PM

Certification/Licensure Practice Area Division Luncheon – All are Welcome!

Room: Oasis Den

Join the Certification/Licensure Division leadership as we share our recent progress and future initiatives across the 10 subcommittees – innovative item types, accessibility, recertification, globalization, value proposition for stakeholders, business/operations, lower volume exams, remote proctoring, legal and policy, and badging.

12:00 PM – 1:15 PM

Lunch with Exhibitors

Room: Ambassador Ballroom

1:15 PM- 2:15 PM

Evaluating the Psychometric Properties of Generated Test Items

Room: Celebrity Ballroom A

Practice Area Division(s):

Topic: Testing, Measurement, and Psychometrics

Session Type: Fishbowl

Educators require large numbers of test items to conduct assessments. But item development is a time and resource intensive process. Automatic item generation integrates cognitive modelling with computer technology to systematically generate items. To-date, however, items generated using automated procedures have not been administered to examinees.

We present item analysis results from the first empirical study designed to evaluate the psychometric properties of generated medical test items. A three-step process was used to generate medical test items. Four hundred and fifty-five Canadian and International medical graduates responded to generated items embedded in a larger test form. The item difficulty, item discrimination, and index of discrimination estimates were calculated for the correct and incorrect options using indices from both classical test theory and item response theory. The item analysis results for the correct option suggests that the generated items measured examinees across a range of ability levels while, at the same time, providing a consistently strong level of discrimination for each item. Results for the incorrect options reveal that the generated items differentiated the low from the high performing examinees. One important concern with the application of AIG stems from the nature of the generative process where it is feared than an item model would yield cloned test items (i.e., generated items that look alike) which perform similarly in the item analysis. The results from our study provide evidence that this concern is unfounded.

A single model produced the generated items administered in our study. Despite this single point of creation, the generated items displayed a variety of difficulty levels ranging from very easy (one generated item had a difficulty level of 0.95 indicating that 95% of the examinees got the item correct) to very hard (another generated item had a difficulty level of only 0.15). The results from our study help demonstrate that a single item model can be used to produce large numbers of statistically diverse items.

PRESENTERS:

Mark Gierl *University of Alberta*

Hollis Lai *University of Alberta*

Andre-Philippe Boulais *Medical Council of Canada*

Andre De Champlain *Medical Council of Canada*

Scoring Alternative Item Types: There's Many a Slip Between the Cup and the Lip

Room: Celebrity Ballroom B
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Breakout

Many testing agencies are expressing interest in closed-form items that are not multiple-choice, and many such item types are available. These item types include drag-and-drop, drag-and-drop matching, hotspot, and multiple answer multiple choice, to name a few. As one might expect, these additional item types present additional scoring possibilities.

This presentation will focus on six scoring models presented by Muckle, Becker, & Wu (2011). These models are Dichotomous, Polytomous trapdoor, Two-point trap door, Negative scoring, Multple true/false, and Positive count. We will then cross the item types with the scoring models and discuss the implications for scoring, some of which are more apparent than others, followed by considerations for setting a passing standard when these items and models are used.

PRESENTERS:

James Penny *Castle Worldwide, Inc.*
Robert Johnson *University of South Carolina*
Andrew Dwyer *Castle Worldwide, Inc.*

FEATURED SPEAKER SERIES

Identifying Sources of Construct-Irrelevant Variance in Performance Testing

Room: Celebrity Ballroom C
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Breakout

Performance testing occupies a prominent role in licensing and certification, with over one-third of credentialing agencies using methods such as oral exams, work samples, and computer-based simulations. The interest in performance testing stems from a desire to have authentic assessments that more directly portray the tasks encountered in the real world. While it is evident that performance tests can make a positive contribution to score validity, Messick (1994) cautions that the perceived directness of performance tests does not forgive the need to investigate potential sources of construct-irrelevant variance (CIV).

In this presentation, the presenter will first identify several possible sources of CIV in performance tests, and then summarize a sampling of studies that document the presence of CIV (e.g., examinee likeability; ambiguity in scoring keys). Much of this session will focus on practice effects – or the tendency for scores to increase throughout the testing day and as the examinee gains familiarity with the testing format. Several studies published in the past several years indicate that such effects are present on both computer-based simulations and simulated work samples in medicine. For example, studies have found that:

(a) performance improves throughout the testing day due to a notable practice effect, and this effect is larger for examinees with less experience with performance testing formats;

(b) score gains for repeaters on performance tests are about twice the magnitude as score gains on multiple-choice tests; meanwhile, there is no advantage for examinees who see a few of the same performance tasks on two occasions; and

(c) repeaters' scores on the second attempt exhibit better internal consistency, a more meaningful factor structure, and correlate more highly with external criteria than scores on their first attempt.

Collectively, the results of these and other studies suggest that performance tests, while arguably providing a more authentic assessment, also are influenced by CIV in non-ignorable ways. After discussing some of the specific ways that CIV impacts score interpretation, the presenter will suggest possible strategies for mitigating its effects.

PRESENTER:

Mark Raymond *National Board of Medical Examiners (NBME)*

Maximizing Exam Security Through & Susceptibility of Various Item Types

Room: Celebrity Ballroom F
Practice Area Division(s):
Topic: Security
Session Type: Breakout

Unintended advantage, regardless of source—exposure, prior knowledge, or cheating—presents persistent and pervasive threats to the validity of the interpretation and use of test scores resulting from certification exams. Research on various data forensic techniques has primarily focused on the susceptibility of selected-response items to exposure, collusion, piracy, and other types of test fraud. However, the content domain and delivery mode of IT certification exams has led to the increased development and use of performance item types, such as simulations. Some argue that these item types are memorable; while others contend that exposure has less impact.

The growing use of performance items could help to mitigate security concerns, as the structure of these items demand that candidates "actually demonstrate their knowledge or skill" and allow for the incorporation of "a variety of other assessment approaches" on a singular exam (Cizek, 1999, p. 168). However, there is currently a dearth of research focused on the security of performance items. With an increase in the use of these items, it is particularly important to understand the nature of the differences between item types, their respective impact on the relative security of exams, and ways to maintain quality exams in situations of known compromise.

This presentation will further existing research by using data forensic techniques (moving averages, item parameter drift, and differential person functioning) on performance item types as well as selected-response item types. Specifically, the relative security of performance items will be compared to traditional (e.g., multiple-choice) and less traditional (e.g., a hot spot graphic) selected response items that may vary in their memorability. This presentation will attempt to generalize these findings across multiple information technology certification exam programs to help practitioners better understand the susceptibility of various item types to exposure, determine next steps should their examinations display patterns of behaviors indicative of security issues based on item types, and lead to better decisions based on improved processes that would strengthen the validity of the interpretations of the results of these examinations. Cizek, G. J. (1999). Cheating on tests: How to do it, detect it, and prevent it. Mahwah, NJ: Lawrence Erlbaum.

PRESENTERS:

Lisa O'Leary *Alpine Testing Solutions*
Russell Smith *Alpine Testing Solutions*
Patrick Irwin *Cisco Systems*
Liberty Munson *Microsoft*
Sara Rupp *CompTIA*

FUNDAMENTALS OF TESTING

Testing Accommodations 101

Room: Oasis Ballroom 4
Practice Area Division(s):
Topic: Legislation, Policy, and Accessibility
Session Type: Breakout

Professional entrance, licensure and certification exam programs frequently receive requests for accommodations from candidates with disabilities. We'll present a basic, understandable model of "best-practices" for reviewing, analyzing, and decision-making of requests for testing accommodations, the legal underpinnings of this process, and common mistakes to avoid. The presenters will also discuss what makes a special arrangements request "reasonable" (or not), and why.

Following the passage of the ADA Amendments Act of 2008, testing organizations have seen a significant rise in requests for test accommodations. This session will provide a basic overview of the legal definition of who is considered "Disabled" and how this is defined differently in the law for adults versus children. The presenters will also discuss recent court decisions and settlements that may impact on the accommodations approval process. This session will continue with an overview of best practices for reviewing requests for accommodations, including appropriate guidelines for candidates and their evaluators, timeliness of the review process, the fundamentals of a sound appeal process for candidates who are denied, and the use of external experts. Some accommodations requests may be made by demanding or aggressive candidates and their advocates, who believe that the revised law gives them free reign to demand whatever accommodations they choose. Often, these candidates and their advocates are unclear about the differences between classroom modifications and reasonable testing accommodations.

The presenters will discuss the differences between classroom vs. testing accommodations, and why some accommodations requests may not be reasonable, because they are so extreme that there are concerns about test security or the validity of the test results.

PRESENTERS:

John Hosterman *GED Testing Service*
Kendra Johnson *Graduate Management Admission Council*

FUNDAMENTALS OF TESTING

Design Differential Item Functioning (DIF) Analyses for Certification/Licensure Programs: Values, Practical Considerations, and a Real-World Solution

Room: Celebrity Ballroom H
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Breakout

This session will discuss how to design Differential Item Functioning (DIF) analyses and utilize DIF results to increase item quality for high-stake certification and licensure programs. DIF refers to the difference in item functioning across groups who are similar in the attributes measured by the exam. Items showing DIF may provide threat to score validity, and DIF analyses are a set of tools to identify such items, which are particularly useful in pre-test item screen and/or evaluating items containing context or situational judgment. Carefully designed DIF analyses will enhance item quality and add credibility to certification/licensure programs.

In this presentation, fundamental meanings of DIF and major types of DIF analysis approaches will be briefly introduced at the beginning. Practical considerations to design DIF analyses are then discussed in details, focusing on issues and problems practitioners may encounter in designing DIF analyses for certification/licensure programs, including (1) whether DIF analyses are appropriate for your program, (2) how to define DIF subgroups, (3) how to deal with multi-racial/ethnical status and design/administer survey effectively, (4) different DIF methods to choose from and factors to consider, (5) how to evaluate DIF results: significance test versus effect size, (6) how to involve subject matter experts (SME) to evaluate identified DIF items, (7) how to weigh statistics versus SME reasoning, and (8) how DIF analyses fit into the overall picture of item analysis.

In the last part of the presentation, a practical solution of conducting DIF analyses will be illustrated using data of a large-scale licensure exam, the Pharmacy Technician Certification Exam. The presenters will walk the audiences through practical issues and weigh various factors, propose a DIF analysis model integrating statistics and subject matter reasoning, show results using real data, and discuss pros and cons of the design. This proposed approach will provide a hands-on solution to design DIF analyses for certification/licensure programs and show how DIF results can be utilized to enhance item quality for certification and licensure programs.

PRESENTERS:

Xiaohui (Dawn) Zhao *Pharmacy Technician Certification Board*
Kristopher Merkler *Pharmacy Technician Certification Board*
Nadine McBride *Castle Worldwide, Inc.*

The Connection Between Socializing Professional Skills and Career Advancement

Room: Mirage Room
Practice Area Division(s):
Topic: Business Strategy and Operations
Session Type: Breakout

Publishing certifications to online destinations provides equal benefit to credential issuers and earners. For credential issuers, making the details of a learning achievement available online increases program recognition while reducing instances of individuals fraudulently leveraging your brand.

For credential earners, the inclusion of certifications on a professional profile can garner twice the number of views as those that do not. And socializing skills over popular online destinations can lead to a more robust professional network as well. Combined, this increased visibility can increase the odds of career advancement and hire-ability.

Join industry experts from LinkedIn, IBM, Microsoft and Pearson VUE as they discuss the greater impact of socializing certifications on career advancement, as well as methods for enabling the sharing of credentials online. Session attendees will come away with a greater understanding of (and appreciation for) the impact that their program's credentials can make on an individual's ability to get noticed and get hired through digital representation.

PRESENTERS:

Mike Kim *LinkedIn*
Kristin Gibson *IBM*
Kerri Davis *Microsoft*
Clarke Porter *Pearson VUE*

Implementing an Open-source Multi-State Assessment Platform for a K12 Consortium – A View from the Trenches

Room: Celebrity Ballroom G
Practice Area Division(s):
Topic: Test Administration and Delivery Models
Session Type: Breakout

Since the Race to the Top initiative was introduced in 2009 – along with $4.35 billion dollars in federal grant money, the hot topic in education circles has been the Common Core State Standards and the consortia funded to implement new assessments aligned with them. Most states are participating in the two largest consortia – Partners for Assessment of Careers and College Readiness (PARCC) or the Smarter Balanced Assessment Consortium (SBAC). Some states are creating their own assessments, whether for technical, budgetary or other reasons. However, those assessments will still align with the CCSS or their own state standards that closely mirror the CCSS. The debate is not about the need for synchronized standards across states that define college readiness or high school graduation standards. It's whether the CCSS are the solution.

The National Center and State Collaborative, a lesser known consortium made up of 22 states and five centers, has taken on a greater challenge. Their goal: to provide CCSS-aligned to the 1% of the student population most underserved – those with significant cognitive disabilities. The NCSC project's mission is to ensure that students with the most significant cognitive disabilities achieve higher academic outcomes and leave high school ready for post-secondary options. NCSC also remained committed to an open-source solution that can be maintained and administered to as large a volume of students required without licensing costs.

Aside from just delivering the typical multiple choice, short answer and constructed response computer-based tests in Math and English Language Arts, the new NCSC Assessment Technology System had to be developed, designed and user-tested to work with multiple assistive devices and accommodations. Breakthrough Technologies, with CTB/McGraw-Hill, customized the open-source platform for delivery of these specialized tests, and created an open-source Teacher Administrator Portal to launch assessments and manage students, classrooms and schools. Piloting these new assessments in spring and fall 2014 demanded intense coordination between Breakthrough, CTB and the NCSC member states. In the end, multi-state test delivery and results export were achieved successfully, but not without encountering technical and logistical challenges. Those challenges and how they were overcome will be the focus of this session.

PRESENTERS:

Kate O'Connor *Breakthrough Technologies, LLC*
Doug Wilson *Breakthrough Technologies, LLC*
Terri Sage *CTB/McGraw Hill*

Back to the Future Planning: Envisioning I-O Assessments in 2025

Room: Oasis Ballroom 5-6
Practice Area Division(s):
Topic: Business Strategy and Operations
Session Type: Fishbowl

In any industry, dramatic change occurs almost every decade in society, technology, and business processes that financially benefits those companies that can accurately predict such change. The purpose of this session is to share a futuristic perspective of I-O assessment systems in 2025. While the focus of this session is on future-oriented strategic planning for I-O test publishers, the planning methodologies discussed will be relevant to all test publishers.

Three highly qualified participants will provide a futuristic vision of I-O assessments based on using a variety of future-oriented strategic planning tools, including (a) back-to-the-future scenario planning, (b) trend-tracking analyses to document megatrends and micro-trends, (c) "Blue Ocean" strategic planning exercises, and (d) anticipating disruptive paradigm shifts and change, among other future-oriented strategic planning methodologies. Each of the participants has approximately 30 years of experience with long-term strategic planning in the assessment industry. John Jones will focus on the possible emergence of new I-O psychology assessment constructs and the application of these solutions in both primary and peripheral I-O market spaces. Robert McHenry will focus on the future of global assessment with a special focus on emerging international testing applications and technologies. Wayne Camara will discuss the future of the school-to-work transition and how a new class of assessments can evolve to ensure a near seamless transition that yields a more competitive workforce.

Back-to-the-future strategic planning, made popular at Apple Computer, will be relied upon by each participant. This planning methodology requires the participants to envision the I-O assessment industry structure and opportunities for the year 2025. They will base their forecasts on a variety of factors, including consulting with futurists in their networks, and anticipating various external factors that will directly impact employment testing (e.g., the economy, laws and regulations, technological advances, and the state of the industry). The participants will then trace a route back from the future to the present to demonstrate that such changes are conceptually possible. The participants will make the case that I-O test publishers need to seriously consider such futuristic trends if they want to remain competitive. Attendees will be strongly encouraged to contribute to the discussion.

PRESENTERS:

Robert McHenry *Executive Chairman, OPP Ltd.*
John W. Jones *Vice President, General Dynamics IT*
Wayne Camara *Senior Vice President, ACT*
Alan Frost *Founder, OE Vision Works*
Gary Behrens *Director, General Dynamics IT (moderator)*

Blueprint Classification Schemes – More Than Just a Sequence of Characters

Room: Oasis Ballroom 7
Practice Area Division(s):
Topic: Program Design
Session Type: Breakout

Classification schemes have been around since at least the time of the ancient Greeks. There has been considerable deliberation over how to classify as the sciences of Taxonomy and Ontology (Information Science) have evolved. In assessment we have used classification schemes to organize and describe the work of professions. While this may seem like a dry and basic topic, the way we organize our thinking about a profession is the first step in the process of assessment and has significant implications for the rest of the process. This session will discuss current approaches and introduce some considerations from the evolving field of Ontology which may inform some new directions.

There are several common approaches to Test Blueprint Classification schemes. Presenters will discuss survey results summarizing the various ways that credentialing organizations currently conceptualize practice analyses and blueprint organization as well as organizations' perception of challenges in using current organization schemes. Common issues raised in blueprint development will be discussed; poly dimensional vs mono dimensional coding; advantages and disadvantages of different types of notation; user friendliness for different groups of stakeholders; uses of the blueprint beyond item writing and test construction; association of an organization's classification scheme with other schemes in the profession.

A final group discussion will be facilitated focused on two topics. How does our scheme for organizing information inform the way we conduct our practice analyses and blueprint development? Do we constrain ourselves with assumptions about structure? Does the relatively new field of information architecture, with its emphasis on user friendliness and ease of retrieval advance our thinking about how we present our classifications?

PRESENTERS:

Mary Browne *National Board of Osteopathic Medical Examiners (NBOME)*

Erik Meyer *American Board of Pediatrics*

FEATURED SPEAKER SERIES

Assessing and Developing Versatile Leaders for a VUCA World

Room: Polo Room
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Breakout

Things are so tumultuous in today's fast-paced operating environment that the business world has adopted the U.S. Army War College's description of the modern battlefield: Volatile, Uncertain, Complex, and Ambiguous. While that may exaggerate the case somewhat, it is no exaggeration that companies need versatile leaders who can read and respond to rapidly changing conditions like never before. In this session, the presenter will describe an approach to executive assessment and development that has been refined over the last 25 years through an iterative cycle of application and research and that is perfectly suited to the times. The focal point is helping executives become more versatile leaders, and it is driven by advances in three inter-related areas: the practice of leadership, the assessment of leadership, and the development of leadership. To that end, the following will be presented:

- a new model that portrays versatility as a mastery of opposites and represents the tensions and tradeoffs that make leadership a balancing act;
- a new tool for assessing versatility in these terms that includes innovative features in measurement and design that have been patented; and
- an associated method for learning and development that combines the outer work of behavior change with the inner work of mindset change and integrates techniques from behavioral, cognitive-emotional, and performance psychology.

PRESENTER:

Robert Kaiser *Kaiser Leadership Solutions*

FEATURED SPEAKER SERIES

Drivers for Innovation and the Changing Context for Testing and Assessment Solutions

Room: Oasis Ballroom 2
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Workshop

This session explores the challenges and opportunities for innovation in testing provided by the rapid changes in the educational and employments worlds. Increasing interests in gamification, demands for ever more efficient assessments deployed on mobile devices, open source technologies, MOOCs as well as crowd sourcing in education are all factors shaping the future for testing and assessment.

These trends and examples of how assessment can leverage them to drive innovation are discussed by two experienced testing professionals:

- Craig Mills, who is Vice President of Research at CTB McGraw-Hill Education, will discuss how trends in the educational space are driving new approaches to the development of tests as well as new needs for testing in areas such as the verification of learning.
- Eugene Burke, Chief Science & Analytics Officer at CEB, will discuss how the move to unproctored Internet testing, growing emphasis on the candidate experience as well as the drive towards more natural assessments aimed at Millennials are shaping the landscape of employment testing globally.

This session reflects a period of significant change in measurement driven by different expectations of test users and changes in education and the workplace. Many users expect more understanding of what test performance represents about a person's ability than is provided in a single score. As technology, outsourcing, and other changes in the workplace eliminate work, jobs that were formerly routine now require higher order thinking skills and licensure and certification tests need to change to reflect those new requirements. Demands for more realistic assessment scenarios, efficient testing, and convenience for test takers also affect test design, development, and delivery.

Session attendees will have an opportunity to reflect on these changes and what they mean for maintaining the relevance and the value of valid measurement in a rapidly changing world for testing.

PRESENTERS:

Eugene Burke *CEB*

Craig Mills *McGraw Hill Education CTB*

Test Security Tools and Measuring Their Effectiveness

Room: Rancho Room
Practice Area Division(s):
Topic: Security
Session Type: Breakout

An effective test security program will help a test sponsor ensure score validity and protect test content by deterring and detecting threats. This presentation will discuss different test security tools implemented (such as electronic detection scanning devices, tiplines and photo upload) and techniques and metrics for meauring their effectiveness in mitigating test security threats. This interactive session will provide descriptions of the threats being addressed, the metrics and measures gathered for using tools such as electronic detection scanning devices, tiplines and photos and how this information is used to measure whether tools being implemented have an impact in the risks.

PRESENTERS:

Rachel Watkins-Schoenig *ACT, Inc.*

Raymond Nicosia *Educational Testing Service (ETS)*

Platinum Sponsor Sessions

TUESDAY, MARCH 3, 2015 • 1:15 PM – 2:15 PM

Professional Testing, Inc.

PLATINUM SPONSOR SESSION

Marketing Your Credentialing Program for Success

Room: Oasis Ballroom 1

Session Type: Sponsor Session

We see it every day—credentialing programs that have enormous potential but don't take full advantage of the outreach opportunities at their disposal! Whether starting from scratch, refreshing an exam, or building new committees, credentialing programs have plenty of opportunities to promote their offerings, improving the value of the credential and growing the pool of applicants. Having built highly successful credentials and through our work with diverse clients, we've learned several strategies to improve outreach.

Created for organizations that want to effectively promote the value of their credentials at every opportunity, this session will discuss a variety of approaches certification organizations can take to expand their reach, more comprehensively represent the full complexity of their programs, and reinforce their messaging and marketing strategies.

PRESENTERS:

David Cox *Professional Testing, Inc.*

Christine Niero *Professional Testing, Inc.*

Vincent Lima *Professional Testing, Inc.*

Certification Management Services (CMS)

PLATINUM SPONSOR SESSION

Developing Performance-based Panel Exams at Mach Speed

Room: Oasis Ballroom 3

Session Type: Sponsor Session

This session will present strategies to plan for and execute on accelerated development of performance-based panel exams. Session topics will include:

- Connecting the requirements for global analyses, content development, alpha testing, and beta testing
- Addressing business constraints
- Managing exam artifacts
- Operationalizing the development process
- Securing the process and its multi-stage outcomes

Attendees will discover new ways to strategically approach development of performance-based panel exams when business requirements have a role in driving completion dates.

PRESENTERS:

Kim Thayn *Certification Management Services (CMS)*

Beverly van de Velde *Symantec Corporation*

International Voice of Testing

ASSOCIATION OF TEST PUBLISHERS

MAKING THE WORLD SMALLER FOR TESTING

The vision of the Association of Test Publishers (ATP) is to become the “International Voice of Testing.” Although the ATP was founded as a trade association in North America in 1992, testing and assessment have been developing into a global community at a rapid pace. Consequently, the ATP expanded by adding a European Division (2005), an Asian Division (2011), and a founders’ group has met to organize an India Division (2012). Accordingly, the ATP is now poised to become recognized as the global trade organization for test publishers, thus making the world smaller for testing. Such expansion is appropriate, inasmuch as the language of testing and psychometrics is international and many issues faced by publishers are the same all over the world. Members of the ATP are truly part of a global community of testing, benefiting from the coordinated efforts of a strong organization that offers education, networking, advocacy, advertising, and industry advancement/protection activities on behalf of publishers worldwide.

Ignite Sessions

TUESDAY, MARCH 3, 2015 • 2:30 PM – 3:15 PM

Ignite Sessions & Exhibits Open

Room: Celebrity Ballroom D-E (while Exhibits are open in the Ambassador Ballroom)

Moderated by: Ashok Sarathy *Innovations in Testing Program Chair*
Rachel Watkins Schoenig *Innovations in Testing Program Vice Chair*

IGNITE 1

Social Networks and Credential Sharing - Enabling and Leveraging Social Sharing as a Marketing Channel

Room: Celebrity Ballroom D-E
Practice Area Division(s):
Topic: Business Strategy and Operations
Session Type: Ignite

This presentation is about how organizations can enable social sharing of their credentials and leverage them as marketing channels.

So much of our identities are shared online, yet most of our credentials are still stored as paper, hidden away in folders. There are still no good ways for learners to share their credentials in a verified manner socially. Traditionally, credentials are "illiquid" - they are owned by the issuing organizations and cannot be used easily by the learner to promote themselves or to discover opportunities.

This presentation explores a new perspective on credentials and social sharing. We present tools and methods that allow learners to own and collect their credentials, and to showcase them on social networks in a compelling way. We discuss different designs and online representations of credentials, and how to effectively showcase them without compromising the learner's private information yet also enabling legitimate requests (such as from a prospective employer or admissions office) for more detailed information.

On the credential issuer side, most organizations still do not know how to leverage social networks as marketing channels. Often, the best promoters of an organization's credentials and brand are their earners. How does an organization encourage their learners to share their credentials and amplify their reach? We discuss different methods that organizations can use for branding, marketing and acquiring users by leveraging the credential earner's social shares. We show examples of how credentials could be showcased in a beautiful and verified manner on popular social networks like Linkedin or on professional services sites like Care.com or ZocDoc.

PRESENTERS:

Eugene Woo *Parchment*
Sarah Kiley *Parchment*

IGNITE 2

Five Innovative Practices to Manage and Work with Big Data in the Testing Industry

Room: Celebrity Ballroom D-E
Practice Area Division(s):
Topic: Data Management
Session Type: Ignite

In the day to day operation of an assessment/testing program, we often encounter "Big Data". Initially, the term "Big Data" was a term used to describe the necessity of solving more disk space among computer engineers. As time has progressed, Big Data has expanded to various situations since large, complex, and multiple datasets continue to grow. Consequently, managing and working with Big Data has been challenging and requires innovative practices of mining, managing, and analyzing the data to produce meaningful and effective ways to communicate it. In this Ignite Session, the presenter will present five innovative practices that can be helpful to the testing industry when managing and working with Big Data.

1. Define and operationalize the question your organization seeks to answer
2. Involve multiple stakeholders when discussing Big Data (e.g., vendors, IT, decision makers, certification holders/SMEs)
3. Identify your data sources and develop a data dictionary (e.g., registration/candidate data, exam score data, financial data, testing vendor data, survey data, public data sets (Department of Labor)
4. Standardized multiple data sets among common fields (ID number, demographics)
5. Produce simple reports designed to answer the question your organization sought to answer

PRESENTER:

Manny Straehle *International Credentialing Associates*

IGNITE 3

Don't Ignore What is Already There: Can Auxiliary Information Improve Formal Assessments?

Room: Celebrity Ballroom D-E
Practice Area Division(s):
Topic: Program Design
Session Type: Ignite

New types of data collected during learning and training can seriously change the nature of formal assessments in the near future. Think about: speed data of the responses in a test, diagnostic information, data from dynamic learning systems, behavior in serious games, and feedback data from tasks and performance evaluations, prior test performances and background data from the respondent. All of these sources of information and data are related to the performance level but are often ignored in formal evaluation of students and employees. Using this information and enhancing and verifying it in a formal assessment could enhance the efficiency and effectiveness of the assessment and its reports.

This ignite session aims to present an overview of potential routes to aggregate information from multiple sources into the reporting of formal assessments. This session describes the characteristics of these types of data and especially how these sources of information can be combined in such a way that richer reports can be delivered about the performance of test takers, decisions are improved and formal assessments times can be reduced. IT and psychometric requirements of the aggregation of data are discussed as well as potential pitfalls in the use of combined data.

Test professionals and those responsible for program design could pick-up ideas for enhancing their assessments and a general audience will get a wider picture of future developments in assessment and the role of more formal and summative assessment in it.

PRESENTER:

Anton Beguin *Cito*

IGNITE 4

Reconceptualizing Score Comparability in the Era of Devices

Room: Celebrity Ballroom D-E
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Ignite

A traditional requirement in testing programs is that scores derived from tests given in different formats should be comparable. For example, if a test is offered in paper-and-pencil format and also online, then a score reported for a paper-and-pencil administration should be the same as the expected score on the online version of the same test. Persons at the same level of ability, per the measured construct, are expected to obtain the same score regardless of format. In the wake of a proliferation of devices of various kinds in schools and workplaces, this conceptualization of test equivalence is both reasonable and problematic.

It is reasonable because it expresses the fundamental requirement of construct-relevance and fairness in all testing. When important decisions are made as a consequence of test scores, it is not appropriate for expected scores to depend on the device on which individuals took an exam. This is especially true in settings where devices with potentially positive effects on scores are less available to those with fewer resources – for example, students in low-income districts.

At the same time, score equivalence conceived of as format independence is problematic today because of the increasingly intimate relationship between platform and performance. Students who do much of their course work on a given device will be better equipped to do well on a test administered on that type of device than in a different format. Moreover, individual preferences and choice can play an important role in inducing a fluency differential across devices, resulting in different optimal devices for assessing each individual.

In this Ignite Session, the presenters argue that traditional notions of score equivalence across test formats have been strained by the availability and utilization of new platforms for assessment, requiring a reconceptualization of score comparability that more explicitly lays out assumptions infeasible to test in every case, but critical nonetheless. Test score and device-dependence rating data from tests administered in different formats are summarized to illustrate the difficulties with claiming comparability in the traditional sense. Implications for exam policy and communication to examinees are presented.

PRESENTER:

William Lorié *Questar Assessment, Inc.*

FEATURED SPEAKER SERIES

IGNITE 5

The Dark Side of Employee Engagement

Room: Celebrity Ballroom D-E

Practice Area Division(s): Industrial/Organizational

Topic: Testing, Measurement, and Psychometrics

Session Type: Ignite

This session provides a fresh perspective on employee engagement, focused on three key points. First, it will examine an important but largely neglected aspect of engagement, which is its dispositional nature. In particular, it will discuss the degree to which variability in engagement can be attributed to bright and dark side personality characteristics. Second, the session will examine the idea of curvilinear links between engagement and performance – can you be too engaged to perform well? Third, the session addresses whether engagement truly drives productivity more than vice-versa. Practical implications, especially regarding strategic innovations concerning the assessment of employee engagement, will be considered.

PRESENTER:

Tomas Chamorro-Premuzic *University College London*

IGNITE 6

Automating Your Video Remote Proctoring

Room: Celebrity Ballroom D-E

Practice Area Division(s): Education

Topic: Security

Session Type: Ignite

Authenticity and fairness in testing need no introduction. Whether in hiring, training, distance learning, contests, or certification, authentic and fair tests have always been a high priority among all stakeholders.

This session will cover existing and popular methodologies in proctoring and ensuring fairness in tests. The session will be futuristic and will discuss:

- What tests and authentication are likely to look like a decade from now—from online biometric verification to audio/video proctoring and data protection;
- Innovations in the last 5 years in the remote proctoring process, including best practices, popular tools, and related advantages;
- Automating the process of video/remote proctoring and how suspicious behaviors can be detected by machine via image processing and related algorithms;
- Digital-age technologies in action in various scenarios; and
- Integrating remote proctoring technology with existing software infrastructure and products, including ease of implementation and related costs.

This session will also talk about live projects via case studies including pain points, solutions, and impacts seen as a result of implementing automated remote proctoring on top of existing testing infrastructure.

PRESENTER:

Ketan Kapoor *Mettl Inc.*

IGNITE 7

Multilingual Automatic Item Generation: Overview and Demonstration

Room: Celebrity Ballroom D-E
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Ignite

Testing agencies require large numbers of high-quality items that are produced in a cost-effective and timely manner. Increasingly, these agencies are also require items in different languages.

In this Ignite session we will present a methodology for multilingual automatic item generation (AIG). AIG is the process of using item models to generate test items with the aid of computer technology. We describe a three-step AIG approach where, first, test development specialists identify the content that will be used for item generation.

Next, the specialists create item models to specify the content in the assessment task that must be manipulated to produce new items.

Finally, elements in the item model are manipulated with computer algorithms to produce new items. Language is added in the item model step to permit multilingual AIG by adapting concepts from classical machine translation that can be used to facilitate the translation process during item modelling.

We will also illustrate our method by generating 360 English, 360 Chinese, and 360 Spanish educational test items produced in mathematics and science designed to measure outcomes from the Common Core Standards. Taken together, our description of a multilingual AIG methodology combined with a demonstration of the methodology will ensure that the attendees in our Ignite session have both a conceptual and a practical understand of how items can be generated in multiple languages.

PRESENTER:

Mark Gierl *University of Alberta*

Roundtable Discussions

TUESDAY, MARCH 3, 2015 • 3:30 PM – 4:15 PM

3:30 PM – 4:15 PM

Dessert with Exhibitors and Roundtable Discussions

Room: Ambassador Foyer

Enjoy dessert with the industry leading vendors in the exhibit hall while learning and sharing with your peers during one of the many Roundtable discussions. All discussions will be centered around the Ignite Session topics. Ignite presenters will facilitate discussions.

Roundtable 1
Social Networks and Credential Sharing – Enabling and Leveraging Social Sharing as a Marketing Channel

Roundtable 2
Five Innovative Practices to Manage and Work with Big Data in the Testing Industry

Roundtable 3
Don't Ignore What is Already There: Can Auxiliary Information Improve Formal Assessments?

Roundtable 4
Reconceptualizing Score Comparability in the Era of Devices

Roundtable 5
The Dark Side of Employee Engagement

Roundtable 6
Multilingual Automatic Item Generation: Overview and Demonstration

Roundtable 7
Automating your Video Remote Proctoring

Fishbowl Session

TUESDAY, MARCH 3, 2015 • 4:15 PM – 5:15 PM

Let's Eat Grandma!: Toward Better, Cheaper, and Faster Psychometric Editing Practices

Room: Celebrity Ballroom A
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Fishbowl

Let's eat grandma. No! Let's eat, Grandma. Let's face it, punctuation saves lives. Indeed, good punctuation can also save test items. The difference between a poor performing item and a strong performing item sometimes rests on the shoulders of a minor edit that powerfully unpacks meaning.

Thus, a sound psychometric edit, as industry veterans know, is a necessary component in the test development process. However, developers of assessments need psychometric edits to be performed faster and cheaper, and for a global audience—without compromising quality. What is the best way to meet these requirements? How can test development organizations foster processes and tools that bolster the editing process, further cutting costs?

This fishbowl session will create a forum for industry newcomers and experts to discuss best practices regarding psychometric editing processes, staffing, and the future of online tools, as the tools apply to streamlining the editing process, particularly for a global audience.

PRESENTERS:

Tara Williams *Caveon Test Security*

Kelli Foster *Caveon Test Security*

Our Heads Shouldn't be in the Cloud but Perhaps Our Tests Should be: Performance-Based Testing and Online Proctoring

Room: Rancho Room

Practice Area Division(s):

Topic: Program Design

Session Type: Breakout

Launching a global certification merging both performance testing and remote proctoring has the potential to reach an immense market. In this session we'll hear from a leading information technology certification organization on why they chose to launch a performance-based exam proctored on-line as opposed to a traditional multiple choice in a testing center.

Advances in cloud computing have made developing a sophisticated performance-based exam economical and less of a security risk; there is no item compromise or loss, they are unique and destroyed after scoring. Paired with real-time online live proctoring, this testing model has the potential to not only more accurately assess candidate knowledge, skill, and ability, but also provide an increased level of convenience and stress reduction for testing candidates.

We'll discuss the driving factors for this type of exam development, implementation, and launch.

PRESENTERS:

Ruben Garcia *Innovative Exams*

Clyde Seepersad *The Linux Foundation*

The Application of Scenario Based Testing (SST) in Large-scale Campus Recruitment Assessment

Room: Celebrity Ballroom C

Practice Area Division(s):

Topic: Test Administration and Delivery Models

Session Type: Breakout

Scenario Based Testing (SST) is designed to evaluate candidates on various situations that occur in a simulated workplace. When candidates complete SST, they are put in a situation very similar to an actual workplace to solve multiple tasks through different items. The skill level, which is the most important factor to evaluate whether the candidate fits the job, is being tested during this process. The result of the study shows that SST in the competitive selection has higher validity and predicts the actual performance of candidates in the future work more effectively.

For the first time in large-scale campus recruitment assessment, computer-based SST was used to assess basic work ability. In this 2013 study which took place in China, the number of examinees was more than fifty thousand. Attend this session to hear the results, and how SST overturns traditional assessment in a competitive selection process. Learn why SST should be the significant new trend in the assessment area.

PRESENTERS:

Ivy QiuShi Wang *Assessment Research Institute, ATA, Inc.*

Alex Tong *ATA, Inc.*

You Don't Have to be a Giant Testing Program to Have Scalable and Manageable Performance-Based Exams

Room: Celebrity Ballroom F
Practice Area Division(s):
Topic: Test Administration and Delivery Models
Session Type: Breakout

A quickly growing mid-size IT testing program developed an in-house performance-based testing program to better measure its candidates' on-the-job knowledge base and skill levels using a live instance of its software. After successfully testing the lab-based exams internally with a subset of its candidate base, the organization looked to expand to its global certification population. The goal was to provide a hands-on testing solution that would interact with its test delivery provider's environment. By doing so, this organization would be able to realize the operational benefits offered by its testing vendor such as registration and scheduling, full financial reconciliation, scoring, and reporting. The challenge lay in scaling and supporting the integration of a live software instance that could interact with an existing test delivery system located on a separate server.

Additionally, the testing program delivers exams outside of a traditional test center network using an internet-based testing solution. This provided the added challenge of maintaining internet connectivity between the live software and exam delivery servers, as well as restoration in case connectivity was lost since the servers are cloud-based. The IT testing program was able to work with multiple partners to build a technology solution as well as design a series of staggered gateway pilots to test the effectiveness of the solution. Through this solution, candidates are able to initiate the testing program's software, access dynamic exhibits and solve problems, and then toggle back to the live exam for response input. This is all accomplished while maintaining a secure environment that does not allow for item exposure or access to browsers or desktop applications. The end goal of the series of pilots is to scale the performance-based testing side of the program while only delivering in venues that are fully supportable.

Session attendees will hear from the testing organization and its partners about how the solution was designed, developed, and tested and the lessons learned at the various pilot stages.

PRESENTERS:

Terry Barham *SAS*

Trish Thomas *Internet Testing Systems*

Ralph Capasso *Skytap, Inc.*

Evolution of Psychological Measurement Models and their Applications in Practical Testing and Assessment

Room: Celebrity Ballroom G
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Breakout

Psychological measurement models have evolved for over a century. While classical test theory methods usually get the basic job done, like the evolution of any technology they are less refined and more limited than their successors. Still, in our experience with testing professionals and researchers alike, we have found a wide spectrum of perspectives on the relative utilities of the classical versus more modern models (e.g., CFA, IRT, and Rasch models). For some, venturing into these latent trait models would alone be a vast innovation over current practices. For others, these models are the current norm and innovation involves new and creative applications.

In this session the presenters will provide a broad, user-friendly overview of the 100+ year evolution of psychological measurement models, with an emphasis on the promises of latent trait models for testing and assessment practice. They will then draw on their experiences interacting with other testing professionals, clients, and organizational leaders to discuss some of the challenges sometimes faced in "selling" the implementation of these somewhat more mathematically complex models.

Placing Rasch models into focus, they will then address the benefits of using these models in common testing applications ranging from multiple-choice testing to partial-credit and rating scale scenarios, to analysis of standard setting judgments and subjective scoring of performance-based tests. Next, they will discuss the utility of Rasch models for building computer-adaptive measurement tools, in both standardized testing and cutting-edge applications where mobile devices are used for measurement in real or simulated performance environments, with immediate, dynamically-tailored feedback. The goal of discussing this latter application is to demonstrate new horizons for testing and assessment based on latent trait models, and hopefully to spawn creative thinking about these technologies in performance assessment and development environments.

The driving objective is to encourage increased use of latent trait models by painting a clear picture of what they can help to accomplish. The focus is not on technical details of the process that are mainly interesting to psychometricians, but on valuable outcomes that drive support for their increased use in a variety of testing and assessment scenarios.

PRESENTERS:

Greg Hurtz *PSI Services, LLC*

Ross Brown *PSI Services, LLC*

Case Study: How Human Grading Machines Standardize Performance Examinations

Room: Celebrity Ballroom H
Practice Area Division(s):
Topic: Test Administration and Delivery Models
Session Type: Breakout

It has been long recognized that traditional human-markers have multiple, negative effects on assessing observational and performance-based tests. The opportunities of integrated human/computer-based assessment not only offer better service to both the candidate and the exam-owner, they can help provide a more accurate picture of competence in increasingly complex examination settings.

This session explores the problems and impacts of traditional human-marking and the effects people can have on examination outcomes. A case-study approach, featuring the Australian Medical Council, will show how integrated computer-based assessments work in practice and how the new breed of human grading machines will create a more valid, reliable and smoother operating examination.

The audience will gain perspective on the use of performance items from representatives of an actual testing program that is currently employing them. Additionally, the members of the audience will be invited to participate in an interactive exercise in which they will be presented with fictitious performance item use cases and asked to apply the discussed concepts in order to provide an informed evaluation and suggest appropriate measurement strategies.

PRESENTERS:

E.W. Looney *BrightLink*
Peter Pascale *Pearson VUE*
Carl Matheson *Australian Medical Council*

FEATURED SPEAKER SERIES

The Political Economy of Assessment: Changing the Rules to Get the Assessments We Need

Room: Mirage Room
Practice Area Division(s):
Topic: Legislation, Policy, and Accessibility
Session Type: Breakout

Darling-Hammond et al.'s recent paper, "Criteria for High-Quality Assessment," distills much of the best recent thought on assessment into five guidelines that seem eminently reasonable as guideposts to the future of high-stakes testing. So why, then, do we face so many obstacles in introducing such next-generation, high-quality, assessments into our classrooms? Why are many testing industry professionals skeptical of the viability of any new high-stakes assessment that deviates too much from the status quo of both cost and features? The trivial answer is: economics and politics. But the problem is worth in considering in greater detail.

In this talk, the presenter will examine both the actors and, crucially, the institutional "rules of the game" that together constrain our actions as an industry. Building high-quality, high-stakes educational assessments is not impossible, but we won't get there by playing under the same rules and hoping things get better. We need to think carefully about why our tests look like they do and how to structure the politics and economics to be more conducive to a better future.

PRESENTER:

Sean (Jack) Buckley *The College Board*

FUNDAMENTALS OF TESTING

Learn Badging from the Leaders

Room: Oasis Ballroom 4
Practice Area Division(s):
Topic: Business Strategy and Operations
Session Type: Breakout

Leading credentialing programs have begun to adopt badging as a means to offer candidates digital representations of their credentials that can be broadcast across the web. However, with anything new there are numerous unknowns, or lessons to be learned. Representatives from Citrix, ISACA, HR Certification Institute and Juniper will join Pearson VUE for a panel discussion around the business drivers for exploring this new technology / emerging standard and what they have learned in their research / implementations.

Attendees will come away from the session with an understanding of what badges are (and what they aren't), considerations to make when adopting badges for your program, benefits of representing your credentials as a digital asset and overall lessons learned from industry leading credentialing programs.

PRESENTERS:

Jarin Schmidt *Pearson VUE*
Kpayah Tamba *Citrix*
Liz Burns *NetApp*
Karyn Waller *ISACA*
Linda Anguish *HR Certification Institute*

Combining Learning and Testing: Individual Learning Paths Through Testing Expertise in Learning Platforms

Room: Oasis Ballroom 5-6
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Breakout

With new technology, the availability of data on student learning increases. This leads to a world of new possibilities regarding the measuring, tracking or monitoring of proficiencies. Furthermore, it offers chances to guide student learning more efficiently and effectively by presenting individual learning materials at the right time in the right format.

In this session we present results of a pilot project in which a learning platform is developed in collaboration between Cito institute for Educational Measurement and the Infinitas Learning Group. In the development of the learning platform, educational learning and assessment materials are digitalized and psychometric expertise are used to present students with their own individual learning paths. Starting from rather traditional textbooks and test material from a Mathematics program, book chapters were digitalized and exercises as well as test items were made interactive and individual learning trajectories were designed. In the learning program the answers that students produce while learning and practicing are stored, aggregated, and analyzed. Based on these analyses individual advice on future learning is provided. This way more proficient students are presented with more complex and additional learning materials while less proficient students are presented with additional exercises and (remedial) instruction. Teachers are enabled to monitor progress of students and to indicate students that need additional help. By monitoring the learning more accurately, they can adapt their teaching to the needs of students.

During the session we will discuss the learning platform and discuss the individual components that are included. The session is of interest to test developers and professionals interested in program design and data management. The aim is to give an overview of the possibilities and challenges that will be encountered aggregating learning and testing content. Challenges from an IT-perspective as well as a psychometric perspective are addressed. Finally, this session will conclude with a perspective of possibilities that arise when learning and testing become more entwined.

PRESENTERS:

Anton Beguin *Cito*
Saskia Wools *Cito*
Jos Wilschut *Infinitas Learning*
Mark Molenaar *Cito*

FUNDAMENTALS OF TESTING

Subscores: Relevance and Reporting

Room: Oasis Ballroom 7
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Breakout

Candidates desire maximum information about exam performance, especially when failing an exam. Exam sponsors strive to provide information in accordance with professional standards and ensure that valid inferences can be made from provided scores. This session spotlights two IT certification programs and examines whether or not reporting subscores would be meaningful and how to report subscores to ensure appropriate inferences.

Recent research in subscores (Haberman, 2008; Brennan, 2012; Sinharay, 2013) provides tractable methodologies for assessment programs to determine the added value of reporting subscores. While candidates tend to want more information about their performance, the reporting of this information may not always be appropriate. It is critical that exam sponsors ensure that reported information leads to valid interpretations of performance. To that end, Zenisky and Hambleton (2013) provide a framework for the score report development process and illustrate relevant considerations for exam program owners.

This session includes two exam sponsors: one currently and one not currently reporting subscores to candidates. The exam sponsor not currently reporting subscores has received candidate feedback illustrating a desire to receive subscore information. This program will be used as a case study example to demonstrate how an exam sponsor can determine the appropriateness of reporting subscores for exams of varying sample sizes, expertise levels, and subscore categories (based on exam blueprint sections). The exam sponsor currently reporting subscores will provide details into their process for deciding which subscores to report and how they should be reported.

Session attendees will be provided with practical information and resources to assist in determining if subscores should be reported for their program and how those scores might be reported. Information regarding a framework for contextualizing the score reporting process will also be provided. Attendees may already be reporting subscores, or trying to determine if subscores would be appropriate for their program. No psychometric knowledge or experience is required to fully engage and digest this session.

PRESENTERS:

Jill van den Heuvel *Alpine Testing Solutions*
Lisa O'Leary *Alpine Testing Solutions*
Kenneth Salchow *F5 Networks*
Jessi Mielke *Esri*

FEATURED SPEAKER SERIES

Uncommon Measures Revisited

Room: Polo Room
Practice Area Division(s):
Topic: Test Administration and Delivery Models
Session Type: Breakout

Late last century, the National Research Council published Uncommon Measures: Equivalence and Linkage Among Educational Tests. The motivation for that report was the debate in the late 1990s between those who favored Voluntary National Tests as a means of assessing the educational progress of students across the nation and those who believed that statistical linkages among existing tests could be used to achieve that purpose. The volume examined the feasibility of linking the results of commercial and state tests to compare one student's achievement with national and international benchmarks, as well as with those of students in other places. The executive summary of that report (Feuer, Holland, Green, Bertenthal, & Hemphill, F.C., 1999) contains: 1. Comparing the full array of currently administered commercial and state achievement tests to one another, through the development of a single equivalency or linking scale, is not feasible. (p. 4) 2. Reporting individual student scores from the full array of state and commercial achievement tests on the NAEP scale and transforming individual scores on these various tests and assessments into the NAEP achievement levels are not feasible. (p. 4) 3. Under limited conditions it may be possible to calculate a linkage between two tests, but multiple factors affect the validity of inferences drawn from the linked scores. These factors include the content, format, and margins of error of the tests; the intended and actual uses of the tests; and the consequences attached to the results of the tests. When tests differ on any of these factors, some limited interpretations of the linked results may be defensible while others would not. (p. 5) 4. Links between most existing tests and NAEP, for the purpose of reporting individual students' scores on the NAEP scale and in terms of the NAEP achievement levels, will be problematic. Unless the test to be linked to NAEP is very similar to NAEP in content, format, and uses, the resulting linkage is likely to be unstable and potentially misleading. (p. 5)

Over the last decade and a half, there have been several developments related to linking test score scales. Many linkages have been performed, some sound and some unsound. In addition, there have been several technical developments with respect to linking scales (Dorans, 1999; Dorans and Holland, 2000; Pommerich & Dorans, 2004; Dorans, Pommerich, & Holland, 2007).

PRESENTER:

Neil Dorans *Educational Testing Service (ETS)*

New Security Trends in High Stakes International Examinations: Communicating with Regulators

Room: Celebrity Room B
Practice Area Division(s):
Topic: Security
Session Type: Breakout

Test and data integrity has always been an important aspect of any credentialing testing agency. Cheating in tests is not a new phenomenon and has existed as long as exams have been used for decision making. There are at least 5 explicit statements in the Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1999) for the prevention of cheating: Standards 8.7, 11.7, 13.10, 13.11 and 15.9. In recent years, due to various media coverage, the topic of cheating has been widely debated in various academic and public fora. The topic of cheating and test security in general has been extensively discussed in various publications. Cizek (1999), Wollack and Fremer (2013) and Geranpayeh (2014) are just a few examples. Media to have shown increasing interest in the security of examinations and how pupils, teachers, schools, test centers and prep teaching centers may be involved in cheating. It is not uncommon for the cheating stories to find their way into the news headlines. The latter has recently caught the attention of the politicians and policy makers who tend to use exam results for various high stake decision making such as teachers and school progress evaluation and immigration.

There are an increasing number of legislative and regulatory requirements on instances of suspected or confirmed cheating is to be handled by the testing agencies. In this connection, a number of government agencies around the globe have added security regulatory requirements for the assessment organizations as part of the conditions for recognizing their tests. The new security requirement conditions may not necessarily be driven by assessment experts, which in many instances may cause disruption to test delivery. We argue in this session that there has to be a collaborative dialogue on test security between the awarding bodies and the regulators. We bring two examples of such positive work in the US and the UK where dialogue between assessment experts and regulators has helped the latter to shape their security policies. We report specifically on new policies with regard to preventing cheating, investigating and managing the effect of cheating and maladministration and how to deal with cheating and maladministration once they are detected.

PRESENTER:

Ardeshir Geranpayeh *Cambridge Assessment*

Platinum Sponsor Sessions

TUESDAY, MARCH 3, 2015 • 4:15 PM – 5:15 PM

Yardstick

PLATINUM SPONSOR SESSION

AIG Free: Mass Production of Items for the Masses

Room: Oasis Ballroom 1

Session Type: Sponsor Session

One of the most promising technologies with the potential to transform the testing industry comes from Automatic Item Generation (AIG). AIG is a class of methodologies to mass-produce test questions. This technology has not yet been available to the broad testing community, forestalling its potentially transformative effects. This session focuses on the unveiling of a free web-based tool for AIG.

Participants in the session will first be introduced to the concept and application of AIG and then receive a complete demo of the tool. Invitations to a Beta launch of the software will also be provided so that afterwards, session participants can be among the first to access and use this technology, with their own content.

PRESENTER:

Greg Sadesky *Yardstick*

PSI Services

PLATINUM SPONSOR SESSION

Organizational Effectiveness Starts at the Top

Room: Oasis Ballroom 2

Session Type: Sponsor Session

Organizational effectiveness starts at the top. Effective, aligned, and passionate leadership is needed from both key executives and Board members to move the organization forward and keep it strong. Accordingly, successful organizations invest resources in the selection, development, and management of leadership talent. This session will incorporate a roundtable discussion among Association Executives, Industry leaders, assessment experts and members of the audience to explore and discuss challenges, considerations, and approaches for leadership talent management, with an emphasis on non-profit credentialing organizations and Boards as a governing body.

1. Attendees will be able to identify essential leadership competencies and characteristics.
2. Attendees will be able to describe key principles of leadership development.
3. Attendees will be able to describe the key considerations in performance management for executives and Board members.
4. Attendees will be able to discuss best practices for evaluating Board and executive performance.
5. Attendees will be able to use the principles, practices, and insights in developing action plans to enhance executive and Board effectiveness.

PRESENTERS:

Greg Becker *PSI Services LLC*
John Weiner *PSI Services LLC*
Susan Stang *PSI Services LLC*

Kryterion

PLATINUM SPONSOR SESSION

Hybrid Cloud-Based Test Delivery

Room: Oasis Ballroom 3

Session Type: Sponsor Session

The era of Client Server-Based testing, in which exams (including items and answer keys) are downloaded and stored at testing centers, is coming to an end as test sponsors seek to address the insatiable demand from brain dump sites for new test content and the profits generated through their unauthorized sale to prospective test-takers. The introduction of cloud-based test delivery represents a quantum leap forward in the never-ending battle to safeguard the intellectual property of test sponsors and reduce the millions of dollars spent each year to refresh exams because of over-exposure. Client Server technology was the logical choice in the 1990's when the Internet was still in its infancy. Two decades' worth of innovation, including browser capabilities and dramatically improved bandwidth, make it possible to stream virtually any test content, in real-time, to any internet-enabled device anywhere in the world with no software installation or content downloads required.

Kryterion, a pioneer in Internet-based test delivery, will lead a comprehensive review of cloud-based test delivery. We will demonstrate how a single platform and test package can be administered across a wide array of delivery modes tailored to the security requirements of each exam. Using this hybrid approach Kryterion currently supports; unproctored quizzes and practice exams, client proctored testing sessions, Kryterion proctored conference/event testing, proctored exams delivered at Kryterion Testing Network sites and online proctored exams with real-time streaming video. Each of these modes utilizes a single test delivery platform with all exam content securely maintained in the cloud.

PRESENTERS:

David Meissner *Kryterion, Inc.*
William Dorman *Kryterion, Inc.*

Receptions

TUESDAY, MARCH 3, 2015 • 5:30 PM – 7:00 PM

5:30 PM – 6:30 PM

Sponsor Thank You and Debrief Reception

Room: Oasis Den

All sponsoring organizations are invited to join the ATP for cocktails as we debrief 2015 and look ahead to 2016.

6:00 PM – 7:00 PM

Industrial/Organizational Practice Area Division Reception

Room: Hospitality Suite 1419

Join the I/O Division for this casual networking reception as you meet the Industrial/Organizational division leadership, learn more about the current initiatives and find out how to get involved.

Breakfast / Closing Keynote

WEDNESDAY, MARCH 4, 2015 • 8:00 AM – 10:00 AM

7:30 AM – 11:00 AM
Speaker Lounge Open
Room: Mission Hills Boardroom

Are you presenting at the Innovations in Testing? Stop in to practice your presentation, collaborate with co-presenters, or just rest for a while. This room is open to all conference presenters.

8:00 AM – 10:00 AM

Breakfast with Closing Keynote & ATP Updates

Room: Celebrity Ballroom D-E

In this session, the ATP Board Chair and CEO will explore the strong ties that define us as a community of testing professionals and where the ATP is headed next.

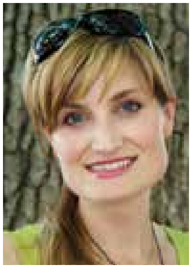
Breakfast will begin at 8:00 AM and the Program will begin at 8:30 AM

PRESENTERS:

Douglas F. Becker, PhD *HMH Riverside, 2015 ATP Board Chair*

William G. Harris, PhD *Association of Test Publishers, Chief Executive Officer*

Closing Keynote



PRESENTER:
Constance Steinkuehler, PhD

Associate Professor in Digital Media at the University of Wisconsin–Madison, Co-Director of the Games+Learning+Society (GLS) Center at the Wisconsin Institute of Discovery

Games for Learning Assessment

Over the last decade, games have been taken up as an important medium for behavioral and cognitive change. As immersive, interactive technologies, games enable new forms of engagement, personalization, and assessment. From balancing algebra equations to invading cells with viral DNA, from negotiating with competing stakeholders over local land use to conducting bench lab stem cell generation in order to save the planet from a zombie apocalypse, games for learning have become the go-to medium for giving students first hand yet designed experiences in areas that are otherwise obscure, dangerous, impossible, or simply bland.

Data exhaust from such learning technologies allows constant product iteration and improvement, tailored personalization for the user, and a rich data set for inferences about not only target domain understanding, skills, and dispositions but also second order variables such as persistence, self-regulation of attention, and play preference. In this keynote, Steinkuehler reviews the current landscape of games for learning and assessment, highlighting recent innovations in assessment design and delivery.

- What are the current opportunities in the game-based assessment space?
- What the barriers and frictions?
- And how will such new technologies fit within (or disrupt) our current educational assessment ecosystem?

The Critical Role of SMEs: Implications to Program Quality, Test Security & Business Assets

Room: Celebrity Ballroom A
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Fishbowl

Subject matter experts (SMEs) are in a unique position to provide expertise and guidance to a testing program's vision and strategic direction. SMEs help define content domains for testing programs and often serve as guardians of academic or professional knowledge bases. Their contributions are vital to the defensibility of exam programs, as they impact nearly every aspect of the test development process. SMEs participate in practice and content analysis, blueprint creation, item development, item promotion decisions, form approval, and standard setting—each of which has implications to validity, test security, and legal defensibility arguments. Given the potential breadth and depth of impact that SMEs can have on a testing program, it makes good sense from test development, psychometric, and business perspectives to invest wisely in their recruitment, training, growth, and retention.

This interactive fishbowl session will include presenters from three different types of organizations: academic, licensure/certification, and industry. The purpose of this session will be to share information and strategies to recruit, train, and retain SMEs. Presenters will outline approaches used by their own institutions, and the session's facilitator will encourage attendee participation and interaction with the presenters so that an authentic exchange of information will occur.

PRESENTERS:

Joy Matthews-Lopez *National Association of Boards of Pharmacy*
Julia Leahy *Prometric*
Ellen Julian *ARDMS*

Using Secure Item and Test Designs to Protect Your Tests

Room: Oasis Ballroom 4
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Breakout

The current testing environment is characterized by relentless attacks on our tests from all directions. Test items are stolen, shared, and sold at lightening speeds through Internet and social media channels. Individual test takers find ways to thwart the testing environment through hard-to-detect technology-based communication devices. Those proctoring the exams either turn their heads or assist in the capture of test content. It is critical that we use every tool at our disposal to blunt the effectiveness of these cheaters and test thieves.

This session will present a number of secure item and test design features; some new, some not-so-new, that can be used in today's technology-based testing world to enhance the security of high-stakes exams. Using a risk-analysis model, we will target the most dangerous types of test threats, and show how we can design items and tests as a realistic defense to the relentless attacks on our exams. Brief demonstrations of innovative designs will illustrate the security principles. Research will be presented to support these principles. Suggestions for when and how to practically implement those innovations will be provided and discussed with the attendees.

PRESENTERS:

Jamie Mulkey *Caveon Test Security*
David Foster *Caveon Test Security*
Steve Ferrara *Pearson School*

Breakout Session

WEDNESDAY, MARCH 4, 2015 • 10:15 AM – 11:15 AM

Designing a Standardized Admissions Test for Young Children Utilizing iPad Technology

Room: Celebrity Ballroom C

Practice Area Division(s):

Topic: Testing, Measurement, and Psychometrics

Session Type: Breakout

Effectively assessing young children has posed a consistent challenge for the testing community. The National Education Goals panel recommended in 1998 that standardized achievement tests should not be used for children younger than age 8. However, much of the research underlying these recommendations was conducted utilizing traditional (i.e., paper-and-pencil) standardized testing. Non-traditional tests have been developed to effectively assess young children, such as the Wechsler Preschool and Primary Scale of Intelligence (WPPSI). However, the WPPSI and similar assessments require a trained psychologist to administer the assessment, substantially increasing the cost of the assessment.

New technologies provide an opportunity for innovative assessments that can overcome the shortcomings of previous early childhood assessments. Tablet-based testing allows for instant feedback to students, greater testing flexibility, less proctoring costs, and greater options in question types. Additionally, test items can be made more interactive to keep young children interested in the assessment. This session will focus on an IPad admissions assessment for children aged 4 to 6 years of age. This assessment includes both ability and achievement sections across both verbal and quantitative domains, as well as multiple item types (e.g., touch answer, drag-and-drop) to vary the students' interaction with the platform. Test questions are presented with colorful artwork and characters to keep the students engaged throughout the testing session. Utilizing the iPad interface allows for the testing of young children without a trained psychologist present, substantially cutting testing costs.

In addition to showcasing the iPad-based early childhood admissions assessment, this session will review the effects of item type on examinee performance. Given the development of motor skills and cognitive capacity at young ages, some item types may be better representations of children's cognitive abilities than others. The session will also review the initial psychometric and validity studies of the new assessment. The accumulation of evidence for the effectiveness of this assessment program will support the use technology with a new population of test takers.

PRESENTERS:

Zhiming Yang *Educational Records Bureau*

Denise Mutlu *Educational Records Bureau*

Breakout Session

WEDNESDAY, MARCH 4, 2015 • 10:15 AM – 11:15 AM

Security & Transparency – Do You Know What You Purchased? Do They Know What They Sold You?

Room: Celebrity Ballroom F
Practice Area Division(s):
Topic: Security
Session Type: Breakout

Your organization has spent the time and money on building your brand, servicing your community, and developing your items and exams. You have chosen your delivery partner after an exhaustive and complete RFP. So how do you know you are getting what you think you bought? How does your delivery partner know what they are providing you in terms of quality and security? Where is the data – the evidence?

A test delivery audit program is essential – it is the most detailed, objective, and complete method to obtain the evidence and assurance you need that your exam content is safe, that candidates are being tested fairly, and that the delivery environment is professional with the security and quality of service you expected at the sale.

- What should you demand and expect from your partner?
- What are the components of a complete audit program?
- What do the results look like?
- How can you trust and verify the results?

This session will help you answer those questions. We will discuss, outline, and explain what makes up a good audit program:

- Secret shops
- Operational audits
- Remote audits
- Corrective and preventative action plans
- Center performance reporting systems
- Test center quality and security scorecards
- Reporting

The benefits of a good audit program are transparency, accountability, and continuous improvement. The output provides you the information to measure the return on your Security investment.

PRESENTERS:

Stephen Hennessey *Prometric*

Jared Zurn *National Council of Architectural Registration Boards*

Item-Writing Workshop: How to Properly Train Writers and Review Items

Room: Celebrity Ballroom H
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Breakout

Items are a huge commodity in educational assessment and certification/licensure. Finding gifted item writers and then training them properly is paramount to a successful assessment development process. Therefore, participants will partake in a mock item-writing workshop where they will undergo training and experience what it takes to write high-quality items in order to give them deeper insight into the process.

Proper training of item writers provides the best return on investment because the products need less editing and feedback. Additionally, strong writers can be retained for future endeavors while weaker performers are recognized and replaced more quickly.

In this workshop, participants will be trained on a particular topic and provided with sample writing assignments. They will then evaluate what makes a good item, as well as tips to writing good items, and participate in a discussion on methods for reviewing items and next steps. Participants will leave the workshop with a stronger understanding of how to properly train item writers and ideas for planning item review.

PRESENTERS:

Canda Mueller *Questar Assessment, Inc.*
Timothy Vansickle *Questar Assessment, Inc.*
Jennifer Dugan *Minnesota Department of Education*

Assessing Conversion Equivalence Between Computer and Paper Pencil Testing

Room: Rancho Room
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Breakout

Measurement conditions play a vital role in assessment. In this study, measurement conditions refer to the situation in which the same set of items is administered in different modes: Computer based test (CBT), and paper-and-pencil test (PPT). When a test is administered in different modes, the question of score equity arises: Do items indeed function the same way when they are administered in computer vs. in paper?

The issue of score equity between CBT and PPT has been extensively studied, yet no consensus has been reached. Some studies found no mode effects, whereas other studies detected mode effects. While the debating continues, none of the earlier studies examined practical equivalence of conversions: when mode effects are detected, should a separate conversion be reported for CBT and PPT each, or should a common conversion be reported?

The purpose of this study is to assess the practical equivalence of conversions, and to determine when it is defensible to report a separate conversion for the same form but administered under CBT vs. PPT.

Data Collection: Equivalent Groups Design
Data will be collected from an admission test. Test takers will be randomly assigned into two groups: One group takes the CBT, and the other takes the PPT. The scores obtained under CBT will be linked to those obtained under PPT using equivalent groups design.

Analysis Plan
Comparison of item statistics. Item statistics, such as P+ and r-bis, can be compared for individual items. Correlation of the item difficulty will be examined as well.

Comparison of Form Equivalence
We also need to evaluate form equivalence at the test level: Will test difficulty change large enough to result in different equating functions? We will link the CBT form to the PPT form. If the difference between the linking function and the identity function is large enough (0.5) and not due to sampling error (±2SEE), mode effects are suspected to exist.

Assessing the Practical Equivalence of Conversions
If steps 1 and 2 above suggest that mode effects are likely to exist, we will assess whether it is defensible to report a separate. If more than half of the test takers whose reported scores will be affected by the choice of conversion (e.g., unrounded scores differ by at least half of the reported-score unit), using a separate conversion for the altered conditions should be considered.

PRESENTERS:

Jinghua Liu *Secondary School Admission Test Board*
Linda Cook *Lakeview Consulting*

FUNDAMENTALS OF TESTING

The Economics of Item Development – Key Cost Factors Impacting Program Profitability

Room: Mirage Room
Practice Area Division(s):
Topic: Business Strategy and Operations
Session Type: Breakout

The test development industry has a unique challenge when it comes to developing valid, high quality examinations – particularly with the costly steps in most testing programs. Most business organizations need to implement a quality test development process while also managing to a specified budget. They are challenge to develop quality assessments, while managing to control costs. It is not unusual, however, for personnel involved with a testing program to have an unclear picture of the true out-of-pocket expenses and resource challenges that typically accompanying significant test development efforts.

This presentation focuses on the key cost factors within the test development process and outlines specific methodologies for reducing those costs. Costs are driven down to the most basic test development unit – the items. In order to get down to the per unit cost of the items (the largest asset any testing organization has) there are many factors that need to be taken into account such as the cost of Subject Matter Experts (SMEs), costs associated with meeting facilitation – such as travel, hotel, and food, and staff time needed to manage the entire test development process.

PRESENTERS:

Julia Leahy *Prometric*

Shannon Carter *Competency and Credentialing Institute*

Innovative Strategies for Efficient High Quality Item Development

Room: Celebrity Ballroom B
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Breakout

Regardless of setting - certification, licensure, employment, or education – test developers are faced with the challenge of reducing costs associated with item development. Even developing multiple-choice items can be costly and time-consuming. In this session, we will discuss strategies that represent innovations in the efficient, cost effective development of high quality test items.

The Armed Services Vocational Aptitude Battery (ASVAB) includes 10 subtests ranging from math to science. It is administered in an IRT-based adaptive format to thousands each year, making the creation of new item pools a constant requirement. In support of DoD's interest in reducing development costs, HumRRO conducted research to (1) examine the relationship between various item features and their associated item difficulty and item discrimination parameter values, and (2) identify model items associated with various difficulty items. The goal of this work was to enable item writers to design items that are targeted specific item difficulty levels, an important requirement for adaptive tests that target the full range of ability. Details of this research and the relationships between item attributes and difficulty will be discussed during this presentation.

Assessment Engineering (AE; Luecht) uses a formulaic approach to specifying item content, thereby simplifying the item writer's role in item development and enabling more precise targeting of item difficulty. As part of the ASVAB effort, key concepts associated with the process were applied in reverse to two subtests to determine if the resulting "formulae" could be used to facilitate future item development. Results and implications of this work will be presented.

NetLogon Technologies leverages a rapid item development approach that concentrates exam development activities into a shortened time frame. By combining item development with the technical review of the items, they been able to improve the quality, clarity, and technical accuracy of their items, reduce time required to develop items, and reduce subject matter expert related costs. Details of this process, pros and cons, and lessons learned will be described during this presentation.

Time will be set aside at the end of the session for audience members to ask questions and share their ideas for improving item development activities.

PRESENTERS:

Deirdre Knapp *HumRRO*

Richard Luecht *University of North Carolina at Greensboro*

David Elfassy *Netlogon*

Development and Validation of a Technology Enhanced Score Report

Room: Oasis Ballroom 5-6
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Breakout

The appropriate interpretation of score reports is fundamental to test validity, yet little research has been conducted to support the validity of score reports.

Contemporary report formats include paper reports, online reports with hyperlinks, and interactive reports that support secondary analyses using demographic filters. Some users may appropriately interpret these types of reports. However, score reports include relatively complex concepts and many users may not correctly interpret the results even when they are precisely described.

The current study investigated the use of digital media to support the interpretation of test results by embedding assessment literacy in a proof-of-concept, technology-enhanced score report targeted for teachers and parents. A parent report from a published, research-based interim assessment was modified for the study.

Developing the report prototype: A master TE score report comprised of graphics, animation and narration was constructed so that the elements could be easily customized to represent students' unique test results. Reported information included number/percent correct, scaled score, local percentile rank, and predicted performance on subsequent state tests.

Pilot and validation study: A pilot school was identified that had (a) recently administered the interim assessment and (b) would share the results of the interim assessment with parents in subsequent parent-teacher conferences. Two validation instruments were developed including (a) a test measuring understanding of reported concepts, information, and scores and (b) an interview protocol to assess parents' affective responses.

A quasi-experimental design was implemented. Two 3rd grade teachers "played" the TE reports and two 4th grade teachers shared paper reports with parents during parent-teacher conferences. Parents were interviewed and took a survey following the viewing.

Parents of twelve 3rd grade (treatment group) and six 4th grade students (control group) participated. Detailed results will be discussed in the presentation. This completed study investigated the use of technology to foster assessment literacy and the appropriate interpretation of test results. The use of technology to enhance reports may improve test validity, instruction that is guided by test results, and student achievement.

PRESENTER:

Daniel Lewis *McGraw-Hill Education/CTB*

Policies and Procedures for the Independent Review of Standard Setting

Room: Oasis Ballroom 7
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Breakout

A critical component of many testing programs is the standard setting process for determining cut scores. These cut scores are critical drivers of the success or failure of any testing program and require strong defensibility from a practical, psychometric, and legal perspective. Test publishers are required increasingly often to provide an independent review and verification of their standard setting activities. This session will review the appropriate policies and procedures for this evolving requirement and allow attendees the opportunity to discuss how this new practice can be incorporated into their own assessment programs.

The presenters for this session have worked as independent reviewers of standard setting activities, including a review of the National Assessment of Educational Progress (NAEP) and major certification and licensure programs. The presenters will share best practices that they have experienced when working with test publishers as well as scenarios that have not been as successful and why. The presenters will also discuss the critical features and components of a valid standard setting, using the framework advanced by Kane and the Standards for Educational and Psychological Testing. This approach details three general sources of validity evidence: procedural, internal, and external. Procedural evidence investigates the defensibility of the method used, the panelist qualifications, and effectiveness of the training. The internal consistency evidence looks for consistency in the results, perhaps with the inclusion of a second panel. External validity evidence can include the analysis of the standard setting results with a different method or can be tied to representativeness of the panelists.

The presenters will review each of these sources of evidence and discuss practical ways to ensure that data is collected for each of these sources during standard setting exercises. Discussion will also include how these validity sources strengthen the resulting standards in terms of a practical, psychometric, and legal perspective. Lastly, attendees will be given the opportunity to discuss their experiences with independent reviewers and the strategies that they have adopted in their own programs.

PRESENTERS:

Hillary Michaels *Human Resources Research Organization*

Michaela Geddes *Yardstick*

Identity Authentication: Tools to Legitimize Identities in Distance Learning

Room: Polo Room
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Breakout

Distance learning continues to grow at a rapid pace and can no longer be ignored. According to Towards Maturity's research, over 70% of companies are planning to implement mobile learning in the next two years. With this explosion of participation, it becomes paramount to ensure that credit for coursework is given to those that are actually completing the tasks. The presentation will demonstrate technology tools that exist to ensure accurate identity authentication in the distance learning programs of organizations, whether candidates are in the same city or across the world.

PRESENTERS:

Andrew Caldwell *ProctorU*
René Massey *Columbia College*
Bryan Leibhart *Columbia College*

ASSOCIATION OF TEST PUBLISHERS
MAKING THE WORLD SMALLER FOR TESTING

End-to-end Certification Program on the Cloud

Room: Celebrity Ballroom G
Practice Area Division(s):
Topic: Program Design
Session Type: Breakout

Certification programs today are not limited to Professional Institutes or Educational Societies –but are being hosted by Enterprises that specialize in a certain domain and are giving out third party endorsement to individuals for their professional knowledge and thereby increasing their employability.

In the age where everything from banking, shopping, learning and even dating has moved to the cloud, certification still has a mile to cover. Traditional setups and practices like physical classrooms for knowledge dissemination and testing under manually invigilation are still very entrenched in the market. These setups boast security and better quality of knowledge delivery but fall short on scalability and outreach which can be overcome with Online Implementations. With technology, pros of both the worlds – online and offline certification programs are being combined.

This presentation will talk about existing and innovative technologies that allow organizations to host and deploy all the processes in the Certification Program Management workflow – and the related best practices and dos/don'ts. All processes – Acquiring new users, Candidate management, Data protection, Candidate communication, Candidate Authentication, Online Payment gateways, Learning Management Systems, Content structuring and delivery, testing in a secured environment, Automated proctoring, Remote program management and Certificate delivery. Innovations in the space will be spoken about and how they can dramatically impact the success of the program. The presenter will also engage the audience with interactive Case Studies, discussions on how some very popular certification programs were conceptualized and implemented – what was done right and what could have been done better! Presenter will also talk about the numbers around the successful certification programs and also discuss critical and common mistakes committed even by the best.

The presenter will also discuss Critical Success Factors of a generic certification program – and how they are impacted positively by taking the whole process on the cloud. Fundamentals of testing will be covered for those who are being initiated into the industry. This will be a cross topic conversation which will also touch upon the following sub-topics listed in the Guideline document – Security, Data Management, Business Strategy and Operations.

PRESENTERS:

Ketan Kapoor *Mettl, Inc.*
Greg Tate *Optimizely*

Breakout Session

WEDNESDAY, MARCH 4, 2015 • 11:30 AM – 12:30 PM

Standard Setting Issues, Research, and Opportunities for Innovation

Room: Celebrity Ballroom B
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Breakout

The rapid evolution of technology has led to many advances in testing methodologies over the past decade, creating new capabilities for item development, banking, exam assembly, and delivery. Despite these advances, fundamental issues in testing remain. Perhaps chief among them is standard setting: how to best establish an appropriate threshold (cut score or passing standard) for examinee performance to meet the objectives of the examination program; e.g., verifying educational achievement, measuring entry-level professional/ occupational capabilities for a job, or certifying professional competence to protect the public.

The predominant measurement framework for credentialing exams and content-oriented selection tests (e.g., simulations, work samples, and knowledge tests) is criterion-referenced measurement (Glaser, 1963; Meherans & Lehman, 1984). Methods for establishing cutscores on criterion-referenced examinations have been around for several decades (e.g., Angoff, 1971; Ebel, 1972; Nedelsky, 1954), although more recent approaches, such as the Bookmark method, have emerged (Lewis, Mitzel & Green, 1996).

This session will review common standard setting approaches such as the Angoff and Bookmark methods, along with results of recent empirical research in applying these methods, lessons learned and implications for standard setting practice. The presenters in this session represent a diverse set of perspectives as psychometric executives, researchers, and practitioners. The presenters will also identify and discuss opportunities for innovation in standard setting practice; e.g., leveraging technology.

PRESENTERS:

John Weiner *PSI Services LLC*
Ellen Julian *American Registry for Diagnostic Medical Sonography*
David Paulson *American Registry for Diagnostic Medical Sonography*
Amin Saiar *PSI Services LLC*

Breakout Session

WEDNESDAY, MARCH 4, 2015 • 11:30 AM – 12:30 PM

What's in YOUR Quality Toolkit?

Room: Celebrity Ballroom C
Practice Area Division(s):
Topic: Business Strategy and Operations
Session Type: Breakout

It is crucial for every testing organization to consider how it assures the highest level of quality all along the value chain to maintain the confidence of test takers, test score users, regulators, and the public. While most testing organizations pride themselves on producing high-quality products and services, they may not be using the most efficient tools and practices to enable them to maximize process efficiency. In this session, leaders from ETS's Office of Quality will discuss how the introduction and application of tools such as a Quality Management System, Six Sigma, and the DMAIC process have strengthened and invigorated ETS's culture of high quality, allowing it to grow its business with confidence. The session will address such key questions as:

- What is a Quality Management System, and how does it help an organization meet client expectations in the most cost-effective manner?
- How can such a system be used to manage, track and analyze the effects of quality improvements?
- What is the Six Sigma methodology, and how can it be best used by a testing company to identify opportunities for improving quality?
- What is the DMAIC (Define, Measure, Analyze, Improve and Control) approach, and how can it be used to generate quality improvement solutions that are data-driven and replicable?
- What are some strategies for ensuring that the use of quality tools become "second nature" within the organization?

Leaders from ETS's Office of Quality will join with a leader from the ETS R&D/Psychometrics area to illustrate how the ETS Corrective Action/Preventative Action (CAPA) system utilizes Six Sigma methodology and tools from the DMAIC toolkit to analyze and control quality non-conformance issues that have occurred, and to assess the risk of incidents that might occur, allowing for proactive improvements. The panelists will present case studies to illustrate the business benefits achieved to date, which include higher client satisfaction and a reduction in "re-work". Audience members will be encouraged to participate by presenting their own challenges, resulting in a discussion of how the tools presented in the session can be leveraged to help improve processes in other organizations.

PRESENTERS:

Amy Schmidt *Educational Testing Service (ETS)*
Jane Borden *Educational Testing Service (ETS)*
Marlene Wolf *Educational Testing Service (ETS)*

The Evolution of Standardized Testing – Yesterday, Today and Tomorrow

Room: Rancho Room
Practice Area Division(s):
Topic: Legislation, Policy, and Accessibility
Session Type: Breakout

Given the heightened focus on standardized testing in today's media, it is not surprising that misperceptions exist about what these tests are designed to do (or not). One of the major misconceptions is that standardized tests are not "fair" and are "biased" against specific groups. No one can argue that tests demonstrate variable performance by specific geographic, gender, racial, and/ or ethnic groups. However, well-constructed tests do not create differences; they only measure differences. From a testing/psychometric standpoint, these performance differences do not make a test "unfair" or "biased." The discussion today related to standardized testing is being framed incorrectly, due to a lack of understanding behind the history of standardized testing. This presentation will focus on presenting a thorough chronological historical look at standardized testing origins and its current state today.

PRESENTERS:

Keith Wright *SSATB*
Jonathan Martin *Consultant*

FUNDAMENTALS OF TESTING

Burning Questions: What Every Exam Manager Wants to Know But Is Afraid to Ask

Room: Celebrity Ballroom G
Practice Area Division(s):
Topic: Data Management
Session Type: Breakout

You've gotten through the initial growing pains of getting your testing program off the ground, but now what? Now that your program has matured and the item bank has begun to grow, how do you further develop and keep track of your evolving bank of content to ensure that you have the most comprehensive and best-possible representation of the knowledge domain you are trying to measure?

In the daily grind of writing, editing, and managing test questions, there seems to emerge a common set of "nagging questions"—questions that exam managers never seem to have the time to sit back and critically evaluate. Does the ever growing to-do list in the back of your mind include any of the following questions:

- When is it appropriate to rework a question or just retire it? Is it worth putting in the extra effort to "fix" an item that is good conceptually, but flawed structurally?
- What is the best way to identify content gaps while combatting redundant items? What important topics are not being assessed in the item bank? What topics already appear in the item bank but are continually reappearing in new item pools?
- What considerations should be used in maintaining sources for referencing items?
- What is the value of using images and what is the best approach for developing a bank of images?
- Should my program consider adding item formats beyond multiple choice?

In this session, join the National Board of Certification and Recertification for Nurse Anesthetists (NBCRNA) and Castle Worldwide as they explore the often-neglected—but very real—challenges of item bank maintenance and the return on investment of maintenance activities. The presenters will draw from their own extensive experience in grappling with these very same issues and will provide guidance and resources for test developers and managers.

PRESENTERS:

Meredith Kuny *Castle Worldwide, Inc.*
Diana Patten *Castle Worldwide, Inc.*
Timothy Muckle *National Board of Certification & Recertification for Nurse Anesthetists (NBCRNA)*

Reasonable Accommodations of Selection Tests for Candidates with Disabilities Staffing the EU Institutions: From Policy to Practice

Room: Oasis Ballroom 4
Practice Area Division(s):
Topic: Legislation, Policy, and Accessibility
Session Type: Breakout

The European Personnel Selection Office (EPSO) delivers a staff selection service to the highest professional standards on behalf of the EU Institutions. Every year, thousands of candidates, from the 28 European Member States, take part in the EPSO selection process.

Considering the volume of candidates, the diversity of their backgrounds, cultures, nationalities, languages, … the legal obligations and EPSO's central role as the selection office for the EU institutions; it is essential that EPSO guarantees equal opportunities, treatment and access to all candidates regardless of their age, gender, nationality, disability, belief … Therefore, EPSO has developed an equality and diversity policy for all, based on an inclusion principle and ensuring non-discrimination in the selection processes.

Most specifically, EPSO has developed a reasonable accommodation policy and procedure for candidates with disabilities and special needs. This policy aims at identifying and reducing the obstacles and the blocking factors in the candidate's experience considering the candidate as the expert in his/her disability. The final goal is to allow them to fully demonstrate their abilities at each phase of the assessment procedure. (In order to define the reasonable accommodations, EPSO takes into consideration different elements: information on candidates' disability, their real difficulties in a selection context and the ways used by each individual to compensate, the type of tests (written, PC, oral), the type of expected answers (multiple choice, case studies, oral presentation,…), experts' consultation.)

This presentation highlights this step by step reasonable accommodation procedure, its challenges and difficulties related to psychometric requirements but also to the high diversity of the national standards and practices in terms of disability and accommodation.

Finally, it gives some very practical solutions to selection and recruitment staff in order to better ensure equality between all candidates regardless of their disability.

PRESENTER:

David Bearfield *European Personnel Selection Office*

FUNDAMENTALS OF TESTING

It's Not Just What You Know, But What You Can Do with What You Know that Matters

Room: Mirage Room
Practice Area Division(s):
Topic: Testing, Measurement, and Psychometrics
Session Type: Breakout

What makes an item hard? What makes it relevant to the real world? For many subject matter experts (SMEs) there is often a notion that the difficulty of a performance task (PT) relates solely to the content knowledge (i.e. rule) being tested with little thought for how that knowledge is to be used (skill level). To create items that get at what candidates know and can do, test developers can utilize a structured approach to item development. This session presents an approach that is extensible to most professional organizations and utilizes current ideas in psychology and psychometrics. Using the Revised Bloom's Taxonomy (RBT), the AICPA has developed a customized skills hierarchy for the accounting profession.

SMEs can be skeptical of thinking in anything other than the content rules that make up their practice. Thus, traditional item development is heavily content focused. However, practitioners can benefit from a system that focuses on what candidates are to do with said knowledge. A customized skills taxonomy has helped the AICPA rethink how it addresses the content and skills of the profession with stakeholders and item writers, alike. In particular, the AICPA has engaged external SMEs to assist in developing task statements for a survey that considers both content and skills associated with each topic and targeted at the right level of the skills taxonomy. This is to inform not just item development, but an overall practice analysis survey to ascertain exactly what is expected of the newly licensed.

This work is part of an Evidence Centered Design-based approach to link the skills of the profession to items anchored in the real world. This presentation will explore the RBT and how it was used/modified to identify and validate the requisite professional skills as well as the tools necessary to create items attuned to practice. This session will present the challenges and practical considerations encountered along the way. Through the process, the AICPA has learned a great deal regarding when to stick with, and more importantly, when to break the RBT rules to create a hierarchy that is both meaningful and practical for identifying competent candidates.

This presentation is applicable to any testing organization interested in better structuring item development to represent "real world" tasks rather than just testing the knowledge of their candidates.

PRESENTERS:

Joshua Stopek *American Institute of Certified Public Accountants (AICPA)*

Henrietta Eve *American Institute of Certified Public Accountants (AICPA)*

Matthew Burke *The College Board*

Accessing Web Resources During a High-Stakes Secure Exam: Is an Open-Book Exam Feasible?

Room: Oasis Ballroom 5-6
Practice Area Division(s):
Topic: Test Administration and Delivery Models
Session Type: Breakout

The literature addressing the relative merits and challenges with open versus closed-book exams is fairly limited. In general, the studies conclude that closed-book exams are better for learning and open-book exams reduce test anxiety. Few studies have examined "high stakes" assessment in medicine with and without web resources as the open-book. On the one hand, physicians argue that they use resources in real time and this makes practice safer as they are not relying on their memory. On the other hand, physicians need a strong working memory of medical practice in order to have efficient information gathering skills so they can provide high-quality patient care. There is little time during a patient visit or in a critical situation to use resources in real time. Either way, testing organizations must be able to ensure the public of the rigor and validity of the certificate and changing to an open-book exam needs further study.

In this session, we will first present the literature on six areas that need to be considered when making the decision about an open-book exam. These areas include: the purpose of the exam such as authenticity and relevance, the overall learning experience, exam performance and preparation, anxiety and enjoyment, psychometrics and logistics, and stakeholder perceptions.

We will then share our experience implementing web resources in a secure exam setting as a proof of concept exercise using a medical certification exam. We will review the system requirements, security management, credential management, external resource usage, and the user experience. We will discuss what research must be done through a pilot where users take a short version of the test, for the purpose of comparing several variations in the design such as type of resource provided, amount of additional testing time allowed, type of participant (e.g., age, gender), and type of question (e.g., recall, synthesis).

The goal of this session is to provide a framework for what needs to be considered in introducing open-book exams to a testing program. We will share our real-world experience using web resources as our open-book to understand the steps in the process to ascertain whether the assessment maintains its validity.

PRESENTERS:

Jane Zhang *American Board of Internal Medicine*
Rebecca Lipner *American Board of Internal Medicine*
Dwan Bovell *American Board of Internal Medicine*

Guaranteeing Fairness and Equality of Treatment Using Parallel Simulation Exercises in Large Scale Assessment Centres: Two Empirical Studies

Room: Oasis Ballroom 7
Practice Area Division(s):
Topic: Test Administration and Delivery Models
Session Type: Breakout

The European Personnel Selection Office (EPSO) delivers a staff selection service to the highest professional standards on behalf of the EU Institutions. Every year, thousands of candidates, from 28 European Members states, go through the EPSO selection process, which includes an assessment centre as final stage.

In every AC, 7 to 8 general competencies and 1 or more job related competencies are assessed twice, in different exercises. When fully deployed, EPSO logistics, staff and selection boards have the capacity to run 6 AC of 6 candidates each per day. Several hundreds of candidates are invited to AC, meaning that the AC periods may last weeks or even months. For this reason, parallel simulation exercises for written case studies, oral presentation and group exercises are developed and used over the whole AC period.

In this presentation, the first EPSO competition with an AC, realized in 2010, and involving 174 candidates (profile lawyer) is presented, job specific scores and general competency scores obtained in the written case study are analysed, and statistical tests are applied in order to check for possible biases related to the use of parallel simulations.

In another competition of 2012, involving 123 candidates (profile assistant in project management), the parallel simulation system in group exercises is also presented. In both cases, results confirm that the use of parallel simulation allows maintaining fairness and equality of treatment, even in extended testing periods.

Results are discussed and recommendations regarding the construction, the use and the legal aspects of parallel simulations are proposed.

PRESENTERS:

Paula Chadwick *European Personnel Selection Office*
Stephane Vanderveken *European Personnel Selection Office*

Breakout Sessions

WEDNESDAY, MARCH 4, 2015 • 11:30 AM – 12:30 PM

Working a Case: Best Practices in Conducting Exam Integrity Investigations

Room: Celebrity Ballroom F
Practice Area Division(s):
Topic: Security
Session Type: Breakout

When confronted with evidence of irregular activities like brain dumping, proxy testing, test tampering, or other forms of high-stakes cheating, test sponsors must investigate to determine the scope of activities, the extent of the potential harm to the test sponsor, and the identities of the participants involved. Now more than ever, test sponsors must be extremely vigilant in enforcing their security policies and take decisive action against those who compromise the integrity of their exams. Test sponsors can only achieve these goals by conducting thorough and effective exam integrity investigations that produce admissible and compelling evidence for later use in legal proceedings against cheaters and perpetrators of fraud. This interactive session will detail strategies for preparing for and conducting an exam integrity investigation.

PRESENTERS:

A. Benjamin Mannes *American Board of Internal Medicine (ABIM)*

Marc Weinstein *Dilworth Paxson LLP*

Workforce Credentials: The Employer Perspective

Room: Celebrity H
Practice Area Division(s):
Topic: Business Strategy and Operations
Session Type: Breakout

Join us for the inaugural invited speaker session of ATP's newest division, the Workforce Skills Credentialing Division. A panel of business and education leaders will provide attendees with real-world insight concerning the current gaps in the workforce assessment space, the global nature of workforce credentialing, and the need for valid and reliable measurement tools and credentials that can be verified. Attendees will have the opportunity to ask questions of business leaders and gauge the interest of business in the new division. Don't miss this historic ATP event!

PRESENTERS:

Dan Rinn *National Technology Transfer*

Kevin Paveglio *ECPI University Virginia Beach*

Speaker Index

ATP Members

ACT, Inc.
Academic Therapy Publications
A&D Resources
AICPA
Alpine Testing Solutions
American Association of Medical Assistants
American Institutes for Research (AIR)
American National Standards Institute (ANSI)
American Printing House for the Blind
Assessio International AB
ATA Testing Authority, Inc.
Assessment Distribution Services
Autodesk, Inc.
Australian Medical Council
Badgecert Inc.
Balochistan Testing Services
Bay State Psychological Associates, Inc.
Beijing Insight Management Consulting Co.
Board of Certification for Emergency Nursing
Breckenridge Institute
Brightlink
CA Commission on Peace Officer Standards & Training
Cambridge Assessment
The Cambridge Don
C++ Institute
Castle Worldwide, Inc.
Caveon Test Security
CASAS
CEB/ SHL Talent Measurement
Center for Credentialing & Education, Inc.
Center for Advanced Studies in Measurement and Assessment
Center for Applied Axiometrics
Center on Education and Training for Employment
Certification Management Services
Certnt, LLC
Chicago Medical Training Center
Cicek Svensson Consulting
Cisco
Cito B.V.
The Clark Wilson Group
The College Board
Collegepre
Comms Multilingual Ltd.
CSA Standards
CTB/McGraw Hill
Dansk Psykologisk Forlag
Data Recognition Corporation
Dilworth Paxson Llp
Directional Insight International
The Donath Group
Educational Records Bureau.
Educational & Industrial Testing Service
Educational Testing Service (ETS)
Eduquity Career Technologies Pvt. Ltd.
Edutech
EMC Corporation
Empleos.net
Employment Research & Development Institute
Engage Global Solutions
Enlight KS
Ericsson
Examity
Excelsoft
Executive Development Assessment Centre
General Dynamics Information Technology
Global Learning and Assessment Development
Golden LLC
Graduate Management Admission Council
Hewlett Packard
Hogan Assessment Systems
Hogrefe Publishing
IDS Publishing Corporation
IELTS USA
Industrial Psychology International, Ltd.
Innovative Exams
Institute for Personality and Ability Testing
International Credentialing Associates
Internet Testing Systems
Japan Third Party Co., Ltd.
Joint Commission on Allied Health Personnel in Opthalmology (JCAHPO)
Joint Commission on National Dental Examinations (JCNDE)
Kryterion Inc.
Learndirect
LinkedIn
Linux Professional Institute
Lotus/IBM
Loyalist Exam Services
Measured Progress, Inc.
Measurement Incorporated
MeritTrac Services Pvt. Ltd.
Metalogic, Inc.
Meta Metrics, Inc.
Mettl, Inc.
MetriTech, Inc.
Microsoft Corporation
Morris & McDaniel, Inc.
Mountain Measurement, Inc.
National Accountant Assessment & Certification Center, PRC
National Center on Educational Outcomes, University of Minnesota
Naklada Slap
NATA Board of Certification, Inc.
National Technology Transfer
National Testing Service - Pakistan
National Council of State Boards of Nursing
NCCPA
Netlogon Technologies, Inc.
Nichols & Molinder Assessments
NPC Inc.
Open Assessment Technologies S.A.
Organization Analysis & Design
Pakistan Librarian Welfare Organization
Pearson Assessments
Pearson Educational Measurement
Pearson VUE
Peoplecert International
Performance Assessment Network (PAN)
Performance Programs, Inc.
Performanse
Personnel Systems Corporation
Pharmacy Technician Certification Board
Prevue Assessments
ProctorU, Inc.
Professional Credential Services, Inc.
Professional Examination Service
Professional Testing, Inc
Profiles International, Inc.
Prov, Inc.
Prometric
PSI Services LLC
Psychological Services Bureau, Inc.
Psychometrics Canada Ltd.
Qpercom Ltd.
Questar Assessment, Inc.
Questionmark Corp.
Ramsay Corporation
Randa Solutions
The Riverside Publishing Company
RM Education
SAP AG
Scantron
Shenzhen Sea Sky Land Technoogies
Sindh Testing Service Pakistan
Sigma Assessment Systems, Inc.
Software Secure, Inc.
Strasz Assessment Systems
Talentum By Projectoneto
Targeted Testing, Inc.
Testing Center for Technological & Vocational Education
Tests.com
Tests for Higher Standards
Turning Technologies
UBC Real Estate Division, Sauder School of Business
Uw Center for Placement Testing
Western Psychological Services
World Class Instructional Design (WIDA) at WCER
Wiris Math
Winslow Research Institute
Wohl Communications
Wonderlic, Inc.
Zoomorphix Systems

Getting connected and staying engaged with colleagues that you meet here is simple.

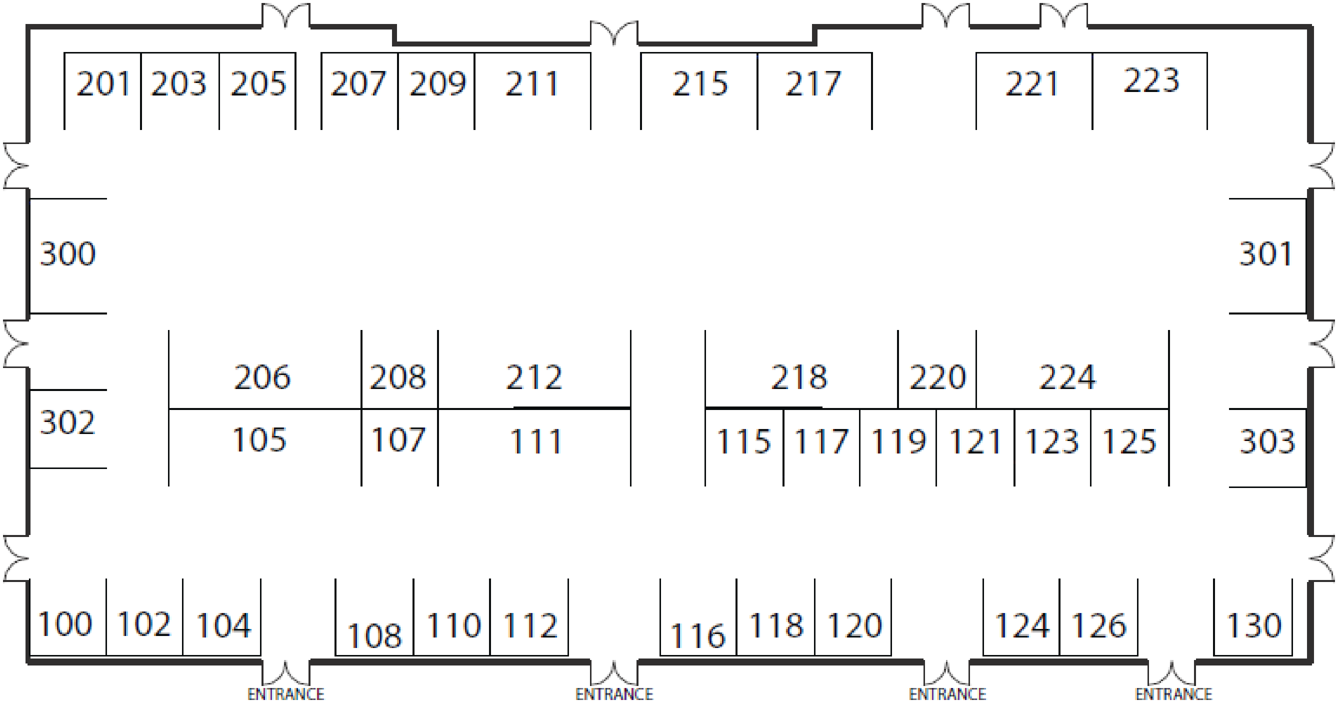
Join our LinkedIn group.
Simply search "Innovations in Testing"

Join us on Twitter.
#atpconf

View videos on YouTube.
Visit youtube.com/innovationsintesting

For more details, visit innovationsintesting.org/connect

Exhibit Hall Map and Exhibitors



100 – Institute for Credentialing Excellence
102 – Excelsoft Technologies
104 – Heuristic Solutions
105 – ETS
107 – Learnosity
108 – Copyright Clearance Center
110 – WIRIS MATH
111 – Pearson VUE
112 – Performance Assessment Network (PAN)
115 – Strasz Assessment Systems
116 – Applied Measurement Professionals
117 – Zoomorphix Systems
118 – Questionmark
119 – BrightLink
120 – Caveon Test Security
121 – Innovative Exams
123 – American National Standards Institute (ANSI)
124 – CertNT, LLC
125 – Comms Multilingual Ltd
126 – ACT, Inc
130 – Assessment Systems
201 – uCertify
203 – Comira
205 – EnlightKS
206 – Certification Management Services
207 – Software Secure, Inc
208 – Tao Testing
209 – Castle Worldwide Inc
211 – Professional Examination Service (ProExam)
212 – PSI Services LLC
215 – Internet Testing Systems
217 – The College Board
218 – Prometric
220 – Examity
221 – Alpine Testing Solutions
223 – Kryterion, Inc
224 – Yardstick
300 – CITO
301 – ATA
302 – ProctorU
303 – Cyveillance, Inc