Extending the Role of Psychometrics in Integrated Security Plans

Lisa S. O'Leary, Ph.D.

Alpine Testing Solutions

Context

- Rampant test fraud in information technology (IT) certification testing programs

- Widespread unauthorized exposure of exam forms and perpetual item exposure

- Continually administered exams

- Tight timeframes for piracy (days or weeks)

…leads to a need for realistic approaches to exam maintenance that can identify specific compromised content to reduce threats to the validity of score interpretation and use
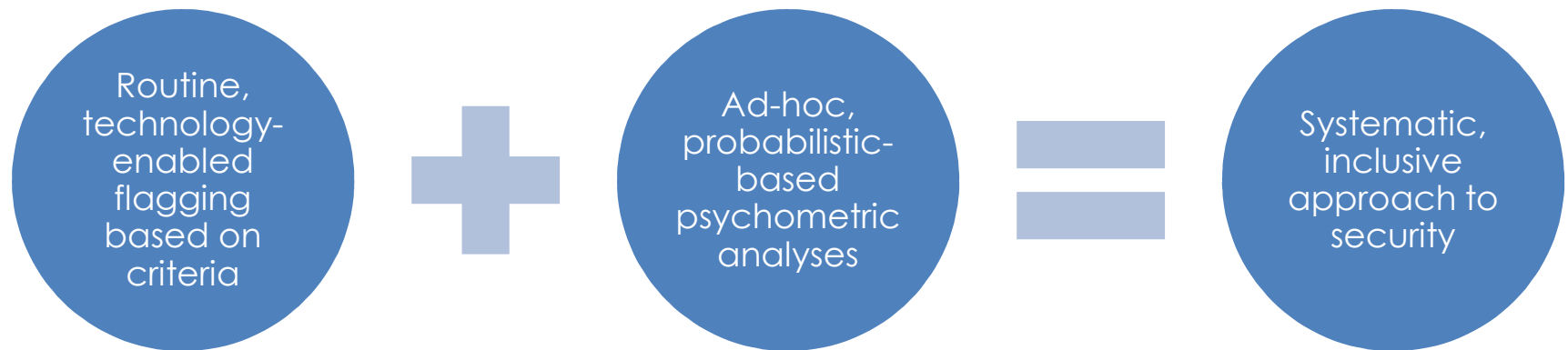
Goals

- Address rampant test fraud in IT certification
- Minimize the costs of piracy
- Maximize the measurement integrity and validity
- Deter future test fraud

Well-articulated data forensic procedures are "*a very effective way to* **communicate** *to candidates* **that cheaters leave behind irregular patterns of responses,** *and that even if they are sufficiently clever to successfully cheat on the exam,* **they will be unearthed by sophisticated statistical procedures** *being run in the background*" (Wollack & Fremer, 2013, p. 11)

Goals

- Proactive consideration of security throughout the test development process
  - Collecting continual security evidence supports the intended use/interpretation of test scores and integrity of credentialing decisions
  - Continually addressing program design, legal considerations, content development, and psychometric analyses protects the security of exam content

Routine, technology-enabled flagging based on criteria **+** Ad-hoc, probabilistic-based psychometric analyses **=** Systematic, inclusive approach to security

Credential Program 1

- Test available to everyone
- On-demand administration
- Immediate score report
  - Autom...
- Six mont...
- Annual U...

Reactionary Posture

- ...
- S...
- L...
- ☐...

Feedback

- Stakeholders like convenience
- Greater respect for earlier awards
- Award viewed with skepticism
- Confused why value is decreasing

Alpine
Testing Solutions

Credential Program 2

- Test available to everyone
- On-demand administration
- Immediate score report
  – Automa
- Monthly
- Six mont
- Periodic

Feedback
- Stakeholders like convenience
- Greater respect for earlier awards
- Award viewed with skepticism
- Confused why value is decreasing

Credential Program 3

- Test available to eligible candidates
- Windowed administration
- Delayed score report (6 weeks)
- Security
- Pilot item
- Item sel
  window
- Periodic

Proactive Posture
- Identify problems prior to
  av
- G
- Li
- Ti
  pr

Feedback
- Stakeholders complain about limited windows and lag in reporting
- No credit for maintaining value

Alpine Testing Solutions

Credential Program 4

- Test available to eligible candidates
- On-demand administration
- Near real-time score report (48-72 hours)
  - Automatic award
- Pilot item
- Data rep
  - Daily se
  - Daily he
- Periodic

Proactive Posture
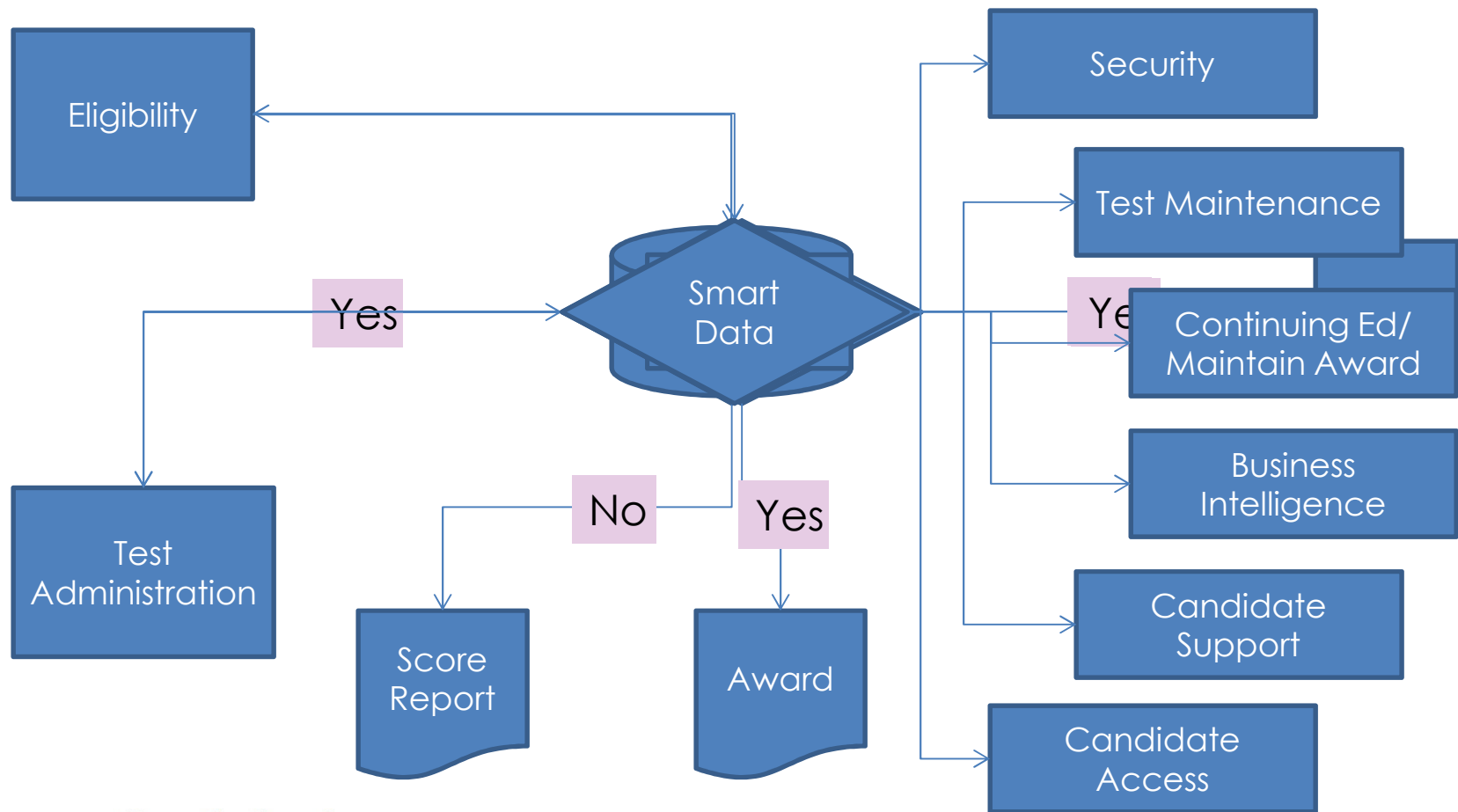- Id
  a
- G
- F
- Ti

Feedback
- Stakeholders like convenience
- No credit for maintaining value

Data Flow

Eligibility

Security

Test Maintenance

Yes → Smart Data

Continuing Ed/
Maintain Award

Business
Intelligence

Test
Administration

No

Yes

Candidate
Support

Score
Report

Award

Candidate
Access

Alpine
Testing Solutions

Security Baker's ½ Dozen

1) Candidate eligibility
2) Protecting intellectual property
3) Candidate flagging criteria
4) Differential performance flagging criteria
5) Alignment of exam maintenance plan
6) Alignment of forms maintenance and retake policy
7) Candidate education

Candidate Eligibility

Initial requirement

- ⏴ **Required: All**
  - ⏴ **Required: 1**
    - ☐ [680-411] IT Education Core Competencies
  - ⏴ **Required: 1**
    - ☐ IT Education Program Agreement
  - ⏴ **Required: 25**
    - ❏ IT Education Program Base Maintenance Fee - $25
    - ❏ IT Education Program CE Maintenance Surcharge - $100
    - ❏ IT Education Program Fee - CE Upgrade - $75

Eligibility Status

Exam Name	Status	Eligibility ID Number	Earliest test date	Testing ends on	Register to test
IT Education Core Competencies	Active	002819810	02-25-2011	03-28-2016	Schedule exam
Robust TCP/IP Network Systems Architecture	Active	002819814	02-25-2011	03-28-2016	Schedule exam
Scholastic Learning Systems Architecture	Active	002742379	01-19-2011	03-28-2016	Schedule exam
Instructional Development	Used				
Marketing 101	Used	038395475	03-24-2011	06-01-2079	Schedule exam
Functional Languages	Active	002819813	02-25-2011	03-28-2016	Schedule exam
Object Oriented Programming	Used				

Protecting IP

- Candidate agreements
- SME/Committee member agreements
- Document security
- Data security

Protecting IP

- Centralized Scoring
  - Keys stored in one, central location
  - Security checking prior to awarding grade/credential
 - Customized score reports

Candidate Flagging Criteria

- Set default flags to identify candidates with potentially suspect exam behavior or performance

| Security Flag	Suspect Action
Exam Retakes	Candidate takes same exam x times within y period
Rapid Exam Completion	x% percent of items completed in less than y seconds
Retake After Pass	Candidate takes the same exam after already passing
Large Score Differential	Score increase by more than x% within 2 attempts within y days
High score/Low time	Exam score above x%, time spent on exam less than y min
Too Little Exam Time	Candidate spent less than x minutes taking an exam
Possible Collusion	Candidates at the same test center on same date and scored within y% of each other on same exam
Security Items	Candidate correctly answered x security items out of y total security items
Differential Item Performance	x% or above on 1st item type and y% or under on 2nd item type
Watch List	Candidate is on the watch list at the time of taking a test
Banned list	Candidate is on the ban list at the time of taking a test

Alpine
Testing Solutions

Candidate Flagging Criteria

- Leads to evidence of possible form exposure and/or candidate pre-knowledge during operational administration
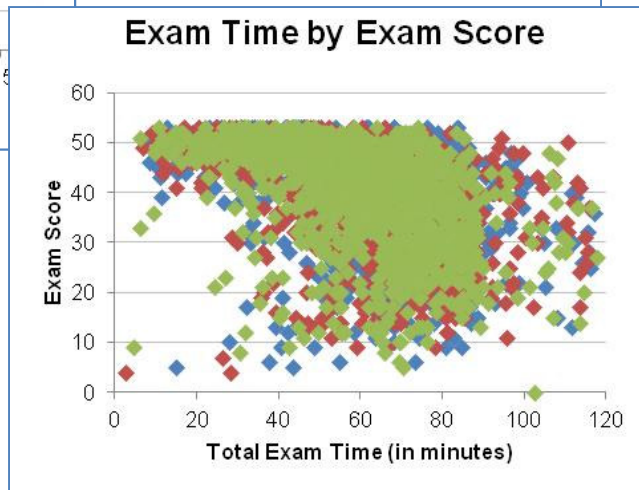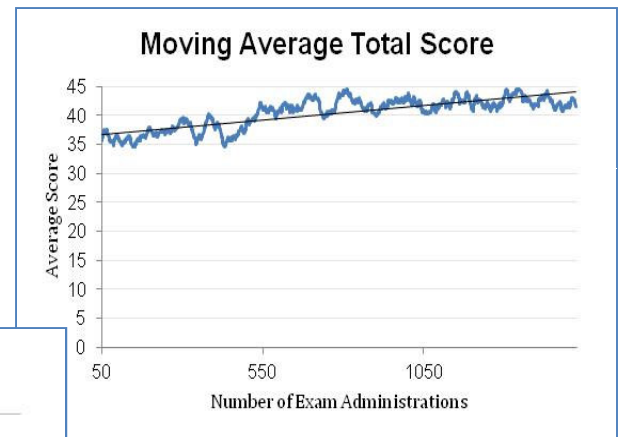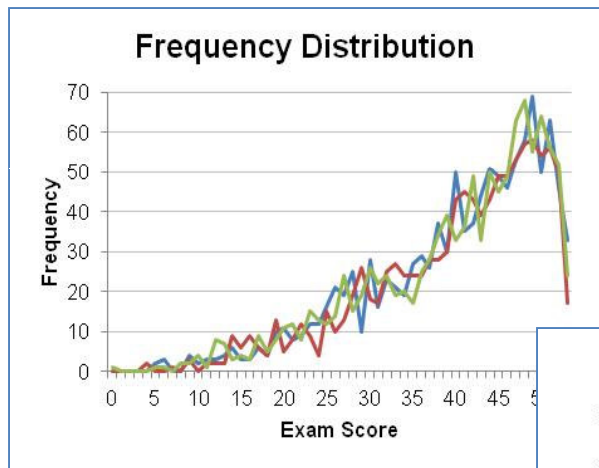
Differential Performance

- Differential person functioning (DPF)
  - Identify candidates likely to have had prior knowledge of exam content
  - Expect candidates with prior content knowledge to have high ability on scored items and low ability on unscored items; low probability of the two measures resulting from the same candidate
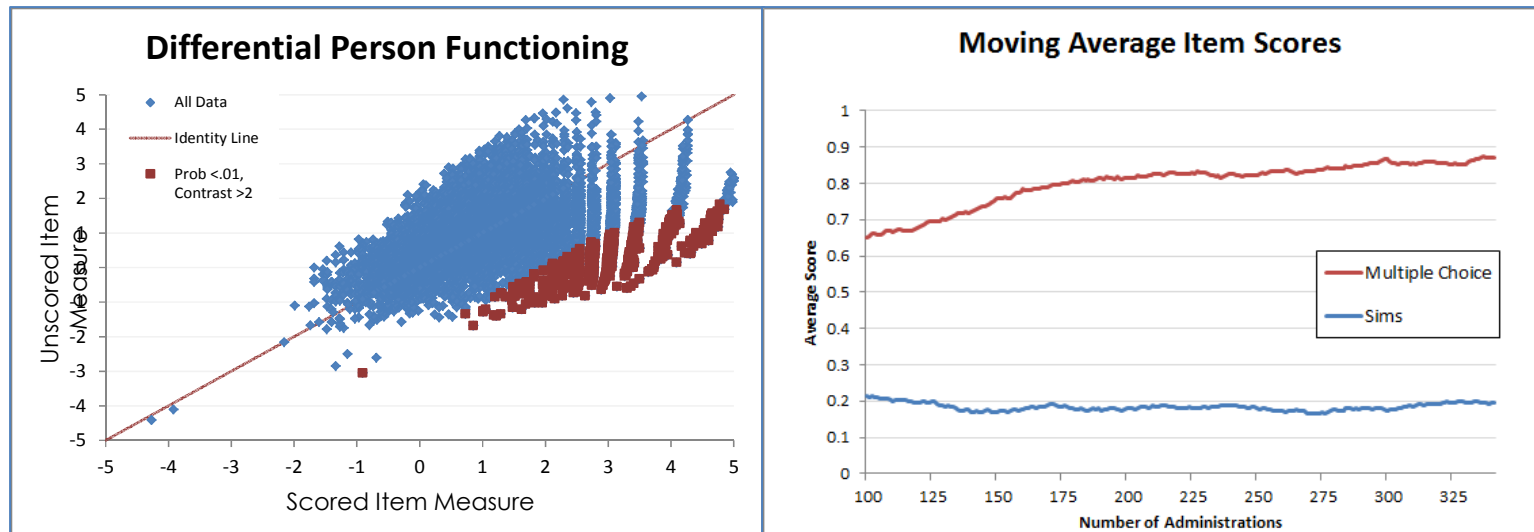
Differential Performance

- Use aggregate results to identify trends across test centers or geographic locations with potential issues or suspect patterns
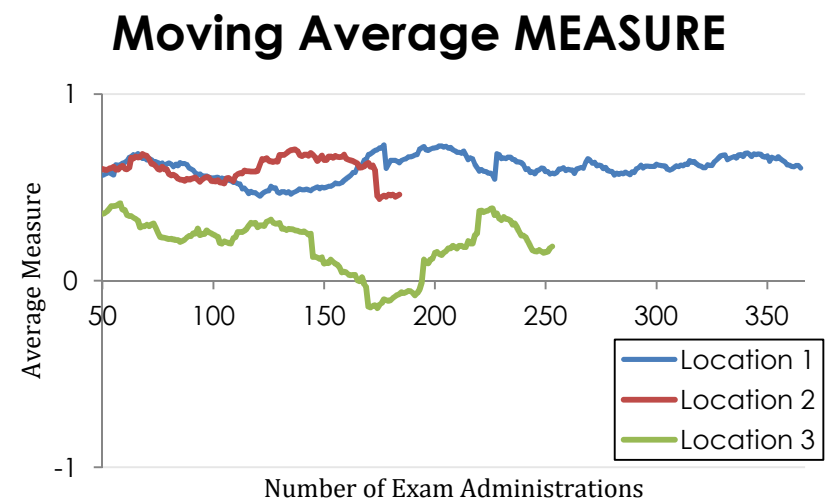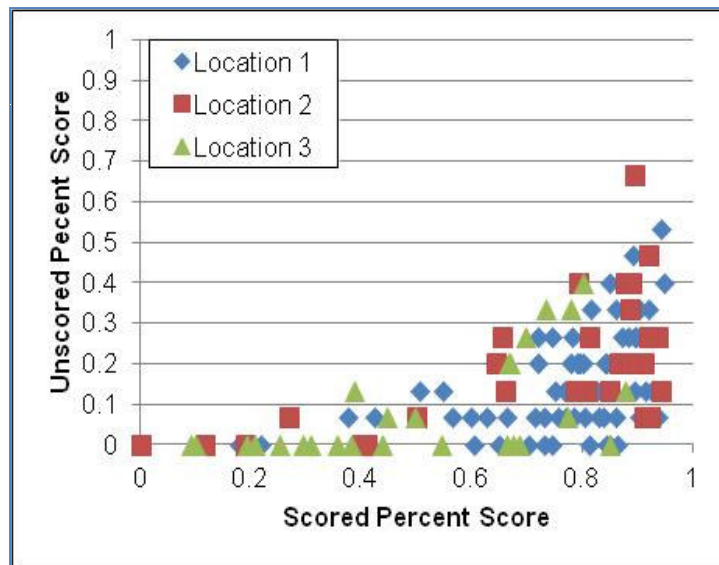
Differential Performance

- Differential item functioning (DIF)

  - Assess extent to which candidates' item pre-knowledge impacts item performance

  - Determine degree of item degradation and gather information to drive exam maintenance

  - Compare performance of DPF-flagged candidates to DPF non-flagged candidates

  - Expect compromised items to favor candidates with item pre-knowledge (DPF-flagged candidates); non-exposed items to be of equal difficulty to both candidate subgroups

Differential Performance

- Practical considerations
  - Item bank size
  - Bank exposure rate
  - Ratio of scored to unscored items
  - Differential performance of scored and unscored items
  - Impact of item degradation
  - Availability of new content
  - Capacity for follow-up action

Differential Performance

- DPF in conjunction with DIF can be used to:

  (1) detect when security breaches have occurred;

  (2) determine the overall extent of item exposure;

  (3) build cases against suspect candidates;

  (4) collaborate with other evidence to support the enforcement of sanctions against candidates;

  (5) highlight specific items with compromised content;

  (6) evaluate appropriate next steps for particular items and entire item banks

  …all while discussing the relevant psychometric and policy issues for each of these areas

Case Study

- 8,350 administrations of large-scale IT cert. exam
- 641 total items (227 scored, 414 unscored)
- Substantial item exposure issues

Total Exam Time by Exam Score
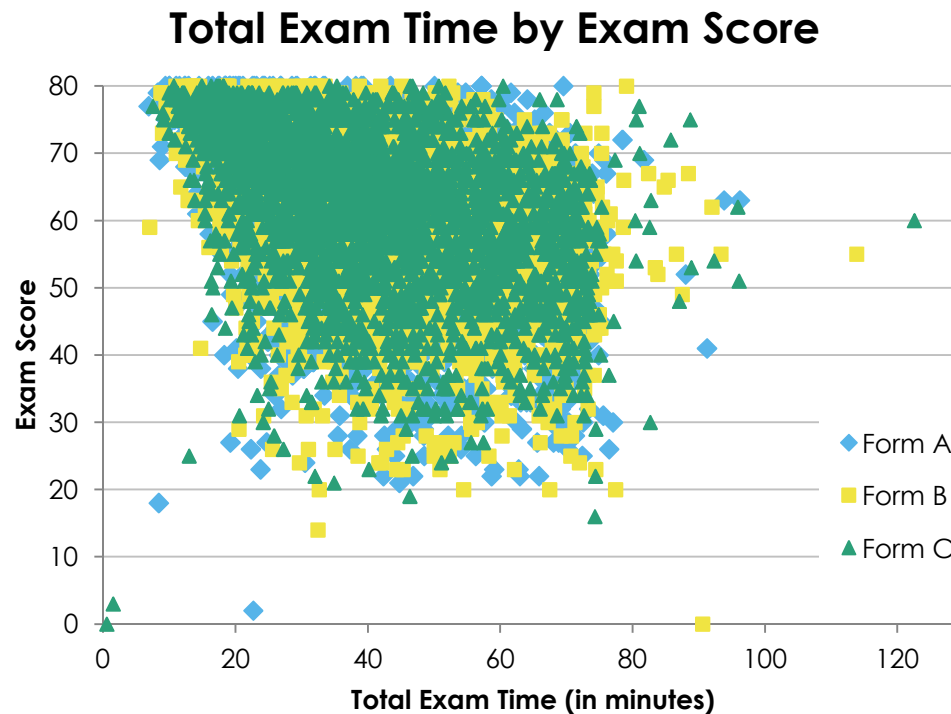
Case Study

- Compared candidates' performance on scored (80) and unscored (20) items
  - Assumed only scored items were exposed; unscored items were not yet compromised
- 531 candidates (6.4%) flagged for DPF

Differential Person Functioning

Legend:
- All Data
- Prob <.01, Contrast >2
- Identity Line

X-axis: **Scored Rasch Measure**
Y-axis: **Unscored Rasch Measure**
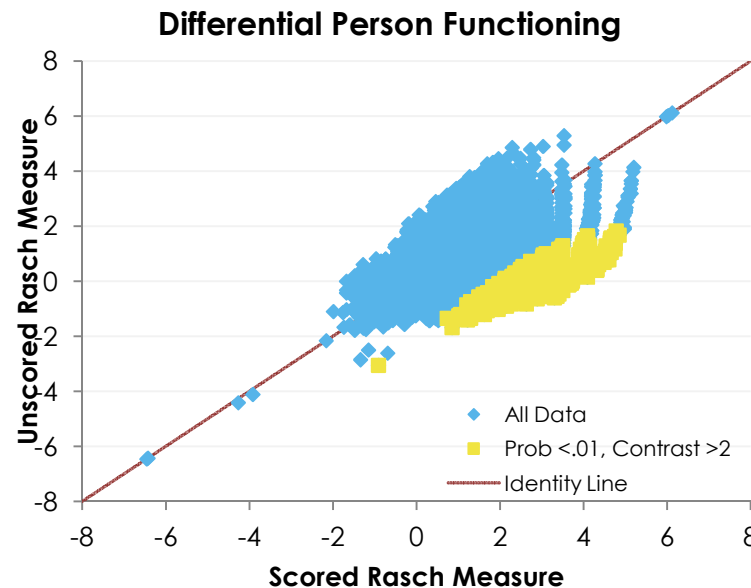
Case Study

- 138 items (20.2%) displayed DIF

Differential Item Functioning
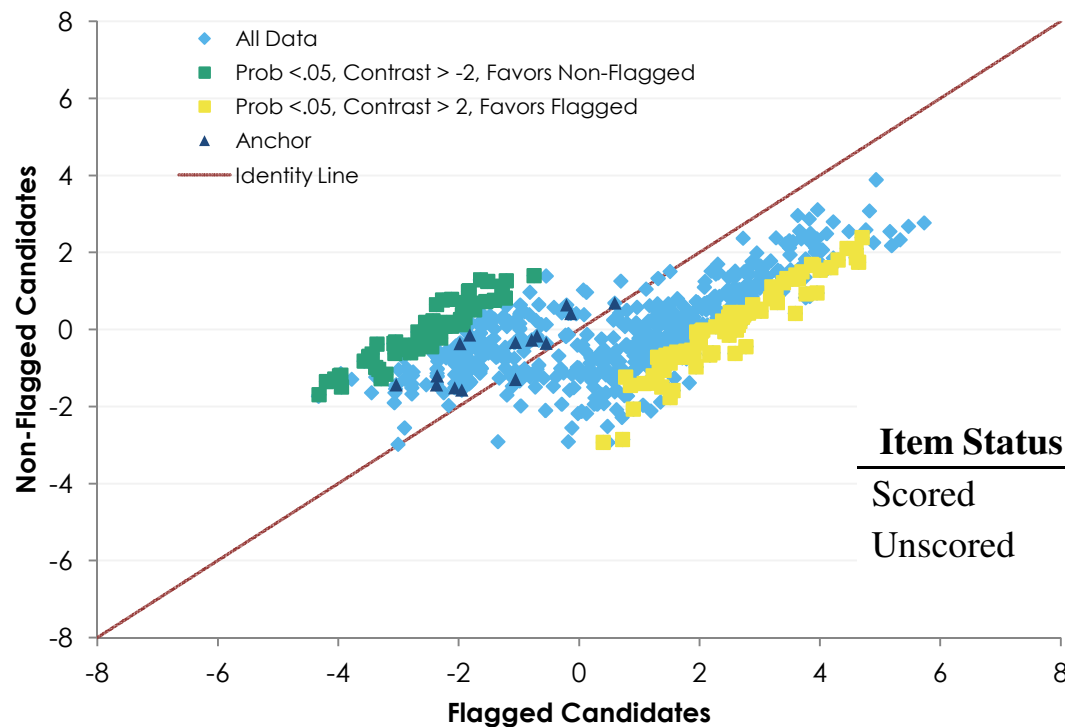
Item Status	Significant DIF Positive	Negative	No DIF
Scored	57	1	169
Unscored	0	80	334

Case Study

- 15 of the 169 scored items without evidence of DIF selected as anchor items for upgrade exam

  – Anchor items were proportionately representative of the blueprint; well-fitting to the model

Item ID	Section	Rasch Measure	P-value	Item-Score Correlation	Flagged Candidate		Non-Flagged Candidates		DIF Contrast	Prob.
					DIF Measure	DIF S.E.	DIF Measure	DIF S.E.		
6	1	0.61	0.60	0.53	-0.21	0.28	0.64	0.04	-0.85	0.003
20	1	0.42	0.64	0.49	-0.14	0.27	0.42	0.04	-0.56	0.045
32	1	-0.33	0.77	0.33	-1.06	0.39	-0.33	0.05	-0.73	0.063
47	1	-0.27	0.76	0.33	-0.79	0.29	-0.27	0.05	-0.52	0.085
57	1	-1.43	0.90	0.27	-2.37	0.7	-1.43	0.07	-0.94	0.185
66	2	0.69	0.58	0.34	0.59	0.16	0.69	0.04	-0.1	0.560
79	2	-1.57	0.91	0.28	-1.95	0.59	-1.57	0.07	-0.38	0.520
102	2	-1.43	0.90	0.25	-3.04	0.89	-1.43	0.07	-1.62	0.073
130	3	-1.29	0.89	0.25	-1.06	0.39	-1.29	0.06	0.23	0.565
151	3	-0.18	0.74	0.53	-1.82	0.42	-0.14	0.05	-1.68	0.000
160	4	-0.16	0.74	0.44	-0.7	0.28	-0.16	0.05	-0.54	0.063
175	4	-0.4	0.78	0.52	-1.98	0.42	-0.36	0.05	-1.62	0.000
182	4	-0.35	0.77	0.39	-0.55	0.27	-0.35	0.05	-0.2	0.473
200	5	-1.2	0.88	0.31	-2.36	0.69	-1.2	0.06	-1.16	0.093
212	6	-1.51	0.91	0.31	-2.07	0.51	-1.51	0.07	-0.55	0.283

Case Study

- Anchor items' item difficulty estimates stable within upgrade item bank

Item ID	Section	Number of Responses	P-value	Rasch Measure	Displacement
6	1	404	0.59	0.61	0.08
20	1	411	0.62	0.42	0.13
32	1	388	0.73	-0.33	0.29
47	1	399	0.75	-0.27	0.14
57	1	424	0.92	-1.43	-0.3
66	2	414	0.58	0.69	0.02
79	2	397	0.91	-1.57	0.09
102	2	444	0.91	-1.43	-0.01
130	3	444	0.90	-1.29	-0.02
151	3	436	0.79	-0.18	-0.24
160	4	416	0.78	-0.16	-0.14
175	4	412	0.77	-0.4	0.13
182	4	401	0.79	-0.35	-0.06
200	5	418	0.87	-1.2	0.19
212	6	395	0.89	-1.51	0.24

Alpine
Testing Solutions

Exam Maintenance Plan

- Administer Operational Forms that were built from Beta Item Selection Analysis

Conduct Health Check

Forms Re-Assembly

- Review exam and form-level statistics
- Delete items that are not performing well
- Set aside items viable after revision with SMEs
- Include well performing items on newly proposed forms
- Seed unscored items to pilot and obtain statistics

- Administer newly proposed forms, equated to prior version
- Conduct health check
- Update item banking with revised statistics and item decisions (as needed)
- Seed unscored items to pilot and obtain statistics

Forms Re-Assembly OR Content Refresh

Alpine Testing Solutions
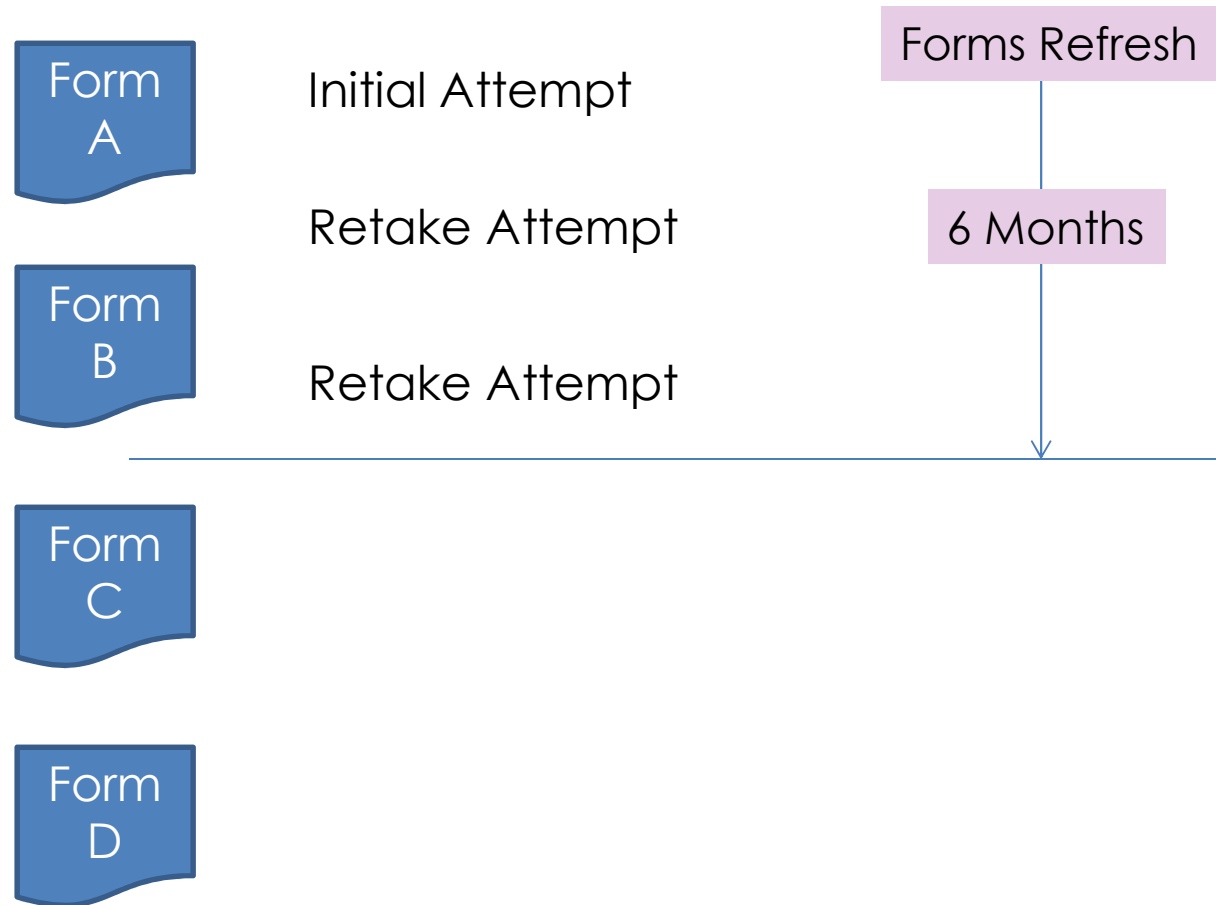
Exam Maintenance Plan

- Piloting unscored items on operational forms allows for:
  - Systematic content refreshing to enable retirement of items with poor statistical performance, suspected exposure issues, or expected bias
  - Proactive approach to enable addition of new items representative of content changes, upgrades, or updates

Pros

Cons

Refresh exam content & expand available item pool for future forms

Obtain item statistics on new items within an operational exam

Candidate effort is high as all items appear to be scored

Compare candidate performance on scored versus unscored items

Need a sufficiently large item pool to have unscored items to pilot

Risk exposure of unscored pilot items

Can add additional time burden to existing forms

Alpine
Testing Solutions

Exam Maintenance Plan

- Rapidity of content shifts impacts the frequency of necessary analysis and maintenance

  – Exams in dynamic and quickly changing domains require more frequent maintenance than those in more static domain areas

- Likelihood or suspicion of suspect candidates or exam behavior impacts the frequency of necessary analysis and maintenance

  – Compromised exams require more immediate maintenance to gauge the impact of the security breach

Forms Maintenance & Retakes

Form
A

Initial Attempt

Forms Refresh

Retake Attempt

6 Months

Form
B

Retake Attempt

Form
C

Form
D

Candidate Education

- Candidates are your friends!
- Value proposition of program
- Self policing

Conclusions

- Proactive consideration of security throughout the test development process increases validity of candidate decisions and the testing program

- Data repository enables diversified approach to exam security
  - Timely candidate detection and enforcement
  - Routine tracking and exam maintenance over time
  - In-depth analyses to address specific concerns

- Probabilistic-based methods for detection of suspect candidates enhance defensibility of actions and enforcement

Contact Information

- Lisa S. O'Leary, Ph.D.
Psychometrician
lisa.oleary@alpinetesting.com

www.alpinetesting.com