ATP Innovations in Testing 2015
Palm Springs, CA

The Top 5 Mistakes in Performance Testing

(And how to avoid them)

ATP Innovations in Testing 2015 Palm Springs, CA

Agenda

1. Introductions
2. Level setting
3. The Top 5
4. Q &A
5. Interactive exercise

ATP Innovations in Testing 2015 Palm Springs, CA

Introductions

- EW Looney, Crucible Product Owner & CEO, BrightLink
- Scott Russell, Lead Test Development Professional, Alpine Testing Solutions
- Jack Terry, CEO, National Board of Examiners in Optometry

ATP Innovations in Testing 2015 Palm Springs, CA

Level setting

- What we mean by “performance testing:”
  - Scoring opportunities that:
    - Require psychomotor manipulation/interaction
    - Often require the candidate to “construct” a unique response
- What we’re focusing on in this session:
  - Human rated scoring

ATP Innovations in Testing 2015
Palm Springs, CA

The Top 5 Mistakes (and how to avoid them)

Mistake #1: Failure to take advantage of technology

You make a cure dent with sing!

You can't make an accurate judgment with missing data!

Symptoms may include

- Missing data
- Significant time spent managing data flow
- If you “touch” data more than once
- Too many “double checks”…

How many “double checks” are required?

How many “double checks” are required?

3?

How many “double checks” are required?

5?


8!

How to overcome this mistake

- Minimize data handling
- Maximize validation
- Use technology to do the “heavy lifting”

Mistake #2: Having vague scoring criteria

Mistake #2: Having vague scoring criteria

Symptoms may include:

- Contradictory/redundant performance level definitions

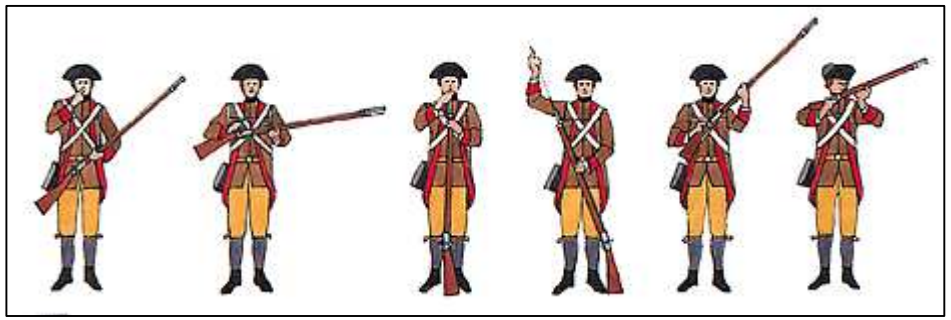
Problems usually stem from failure to account for the following….

Basic foundational information

1. What is the purpose of the test?
2. What are your expectations of the minimally qualified candidate?
3. What type of evidence would indicate a candidate is meeting expectations?
4. How can you gather that evidence from a candidate?

What do you intend to measure?

- End state
- Process
- Both



How will you rate it?

- *Empirical approach*
  - Appropriate when a successful candidate response is predictable
  - Quantifiable
  - More objective for raters
  - Less flexible for raters
  - Appropriate for both computer- or human-scored items
- *Non-empirical approach*
  - Appropriate when a successful candidate response is not predictable
  - More subjective for raters
  - More flexible for raters
  - Appropriate for human-scored items, only

How can raters classify candidate performance?

- Each performance level should be clearly defined and distinguishable from the others
  - Performance level descriptors are critical
  - Exemplars may be necessary

Performance level	Performance level descriptor	Exemplars
1	Navigates surface streets in a time efficient manner	- Times traffic lights - Chooses lightly trafficked routes
0	Navigates surface streets inefficiently	- Delayed at multiple traffic lights - Drives into avoidable traffic delays

A MANUAL FOR THE
ASSESSMENT OF
ENTRY-LEVEL CLINICAL SKILLS
IN OPTOMETRY

July 1987

International Association of Boards of Examiners in Optometry
Suite 805
5530 Wisconsin Avenue, N.W.
Washington, D.C. 20815

ATP Innovations in Testing 2015
Palm Springs, CA

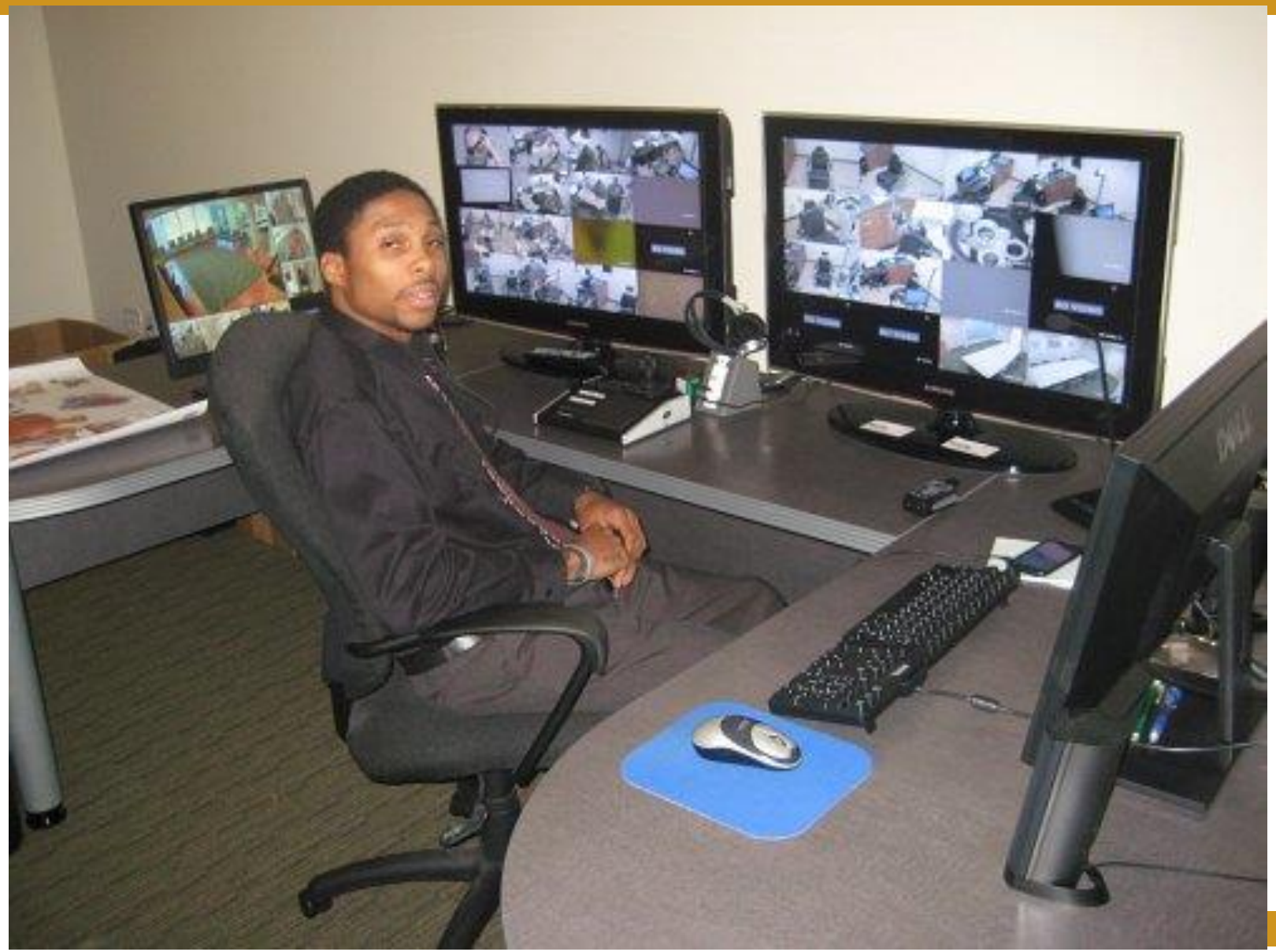
National Center of Clinical Testing in Optometry

ATP Innovations in Testing 2015
Palm Springs, CA

National Center of Clinical Testing in Optometry

ATP Innovations in Testing 2015
Palm Springs, CA

National Center of Clinical Testing in Optometry



ATP Innovations in Testing 2015 Palm Springs, CA

How to avoid vague scoring criteria

- Use Yes/No criteria when possible
  - Reduces ambiguity
- Provide clear performance level descriptors and exemplars for subjective subject matter
  - NBEO provides guidelines (Examiner comments on the exam form) on how to grade certain “subjective” items

How to avoid vague scoring criteria

- Ensure universally accepted/documented best practices inform scoring
  - Notify both raters and candidates
- Be clear with the prompts and scoring expectations of the candidate
  - NBEO provides Skills Summary Sheet in each room
  - NBEO posts makes evaluation forms available online
  - For NBEO exam, “Examine and Describe he macula and fovea” means that they MUST both examine and describe for both the macula and fovea
  - For NBEO exam, Examine and Describe the entire cornea
    - i.e., epithelium, stroma, endothelium

How to avoid vague scoring criteria

- Be clear with exam environment instructions to the candidate
  - NBEO provides Orientation Video before exam begins
  - For NBEO exam, the proctor is allowed to notify Candidate that they "Have a View" but the quality and clarity is not a given.
  - NBEO ensures that labels are strategically located
- Minimize variability of exam stimuli
  - For NBEO exam, Standard Patients (SP) with certain findings have tolerances:
    - E.g., If SP has a corneal opacity (scar), it is noted in their profile and the IHE/RE has SP data profile as a reference

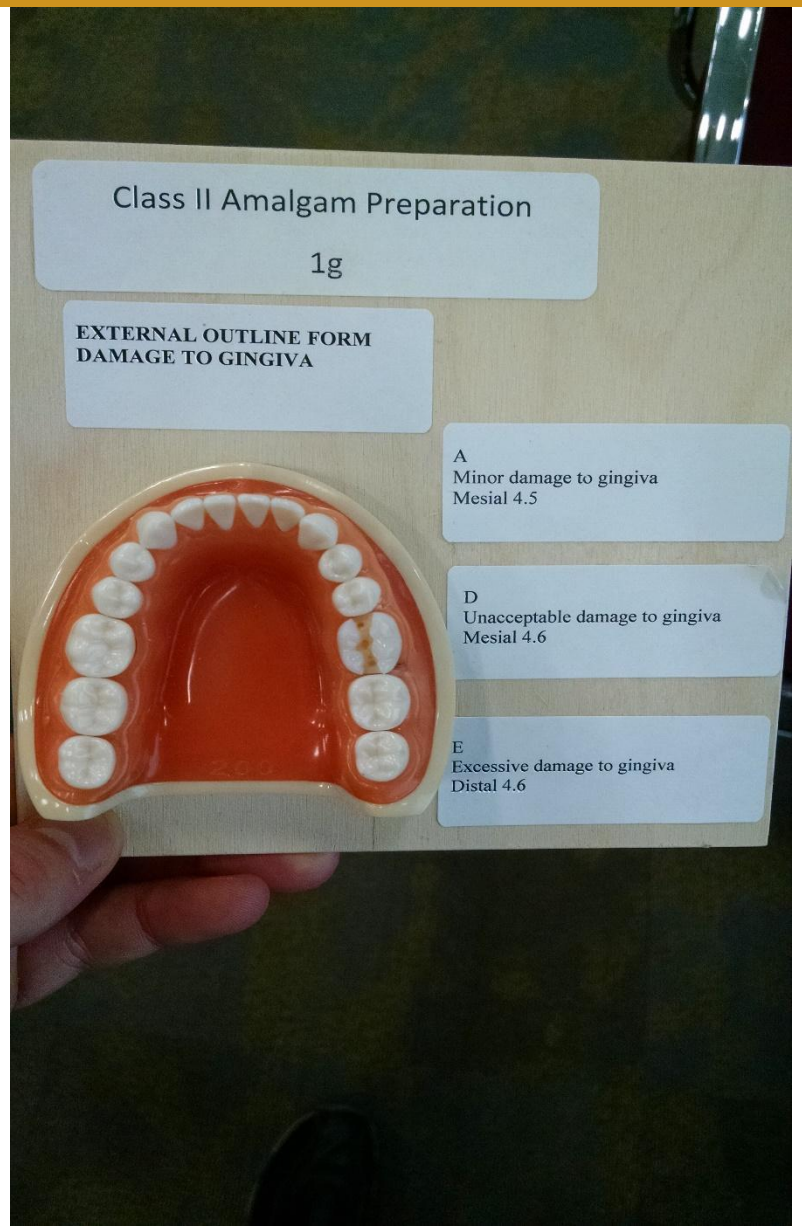
Mistake #3: Failure to calibrate evaluators

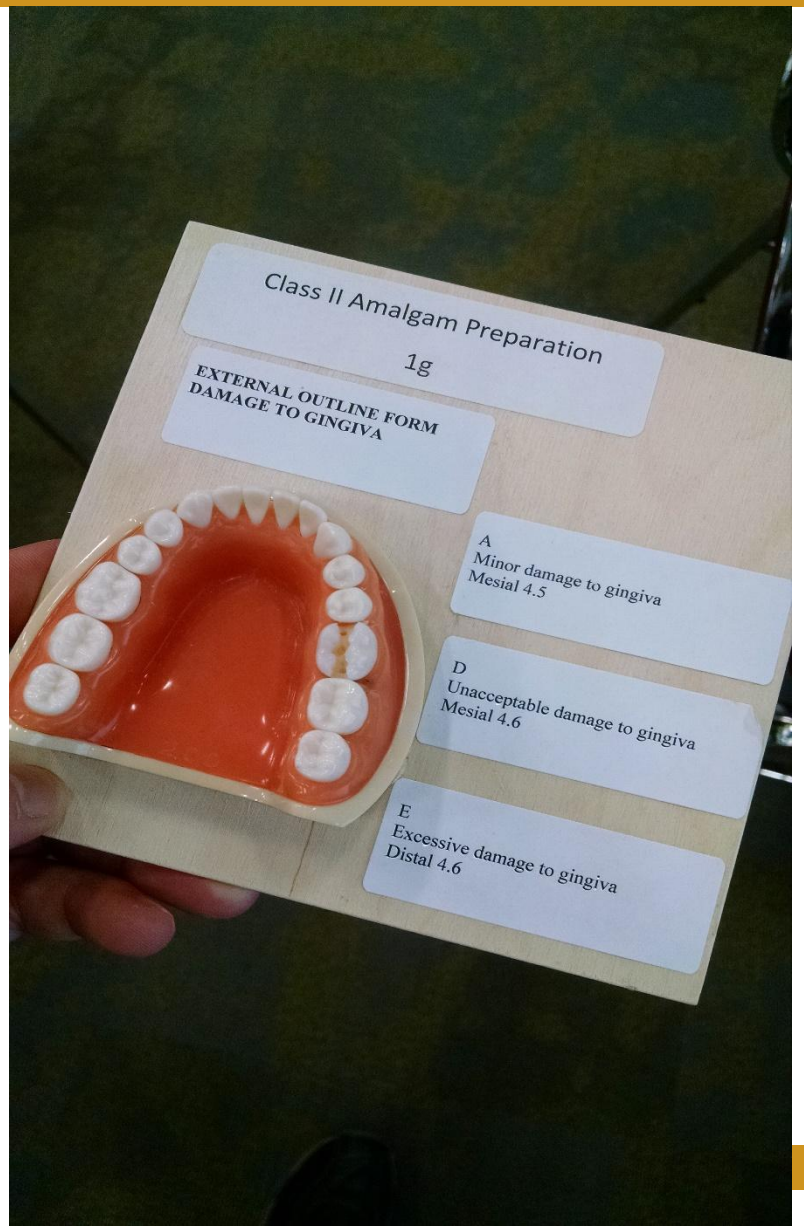
Symptoms may include:

- Poor inter-rater reliability
- Examiner disagreements regarding criteria
- Confusion around criteria
- Grandstanding on “important” aspects of the examination

How to avoid this mistake:

1. Make calibration real
2. Measure outcomes
3. Refine the process


Class II Amalgam Preparation
1g
EXTERNAL OUTLINE FORM
DAMAGE TO GINGIVA
A
Minor damage to gingiva
Mesial 4.5
D
Unacceptable damage to gingiva
Mesial 4.6
E
Excessive damage to gingiva
Distal 4.6


Class II Amalgam Preparation
1g
EXTERNAL OUTLINE FORM
DAMAGE TO GINGIVA
A
Minor damage to gingiva
Mesial 4.5
D
Unacceptable damage to gingiva
Mesial 4.6
E
Excessive damage to gingiva
Distal 4.6

[redacted]

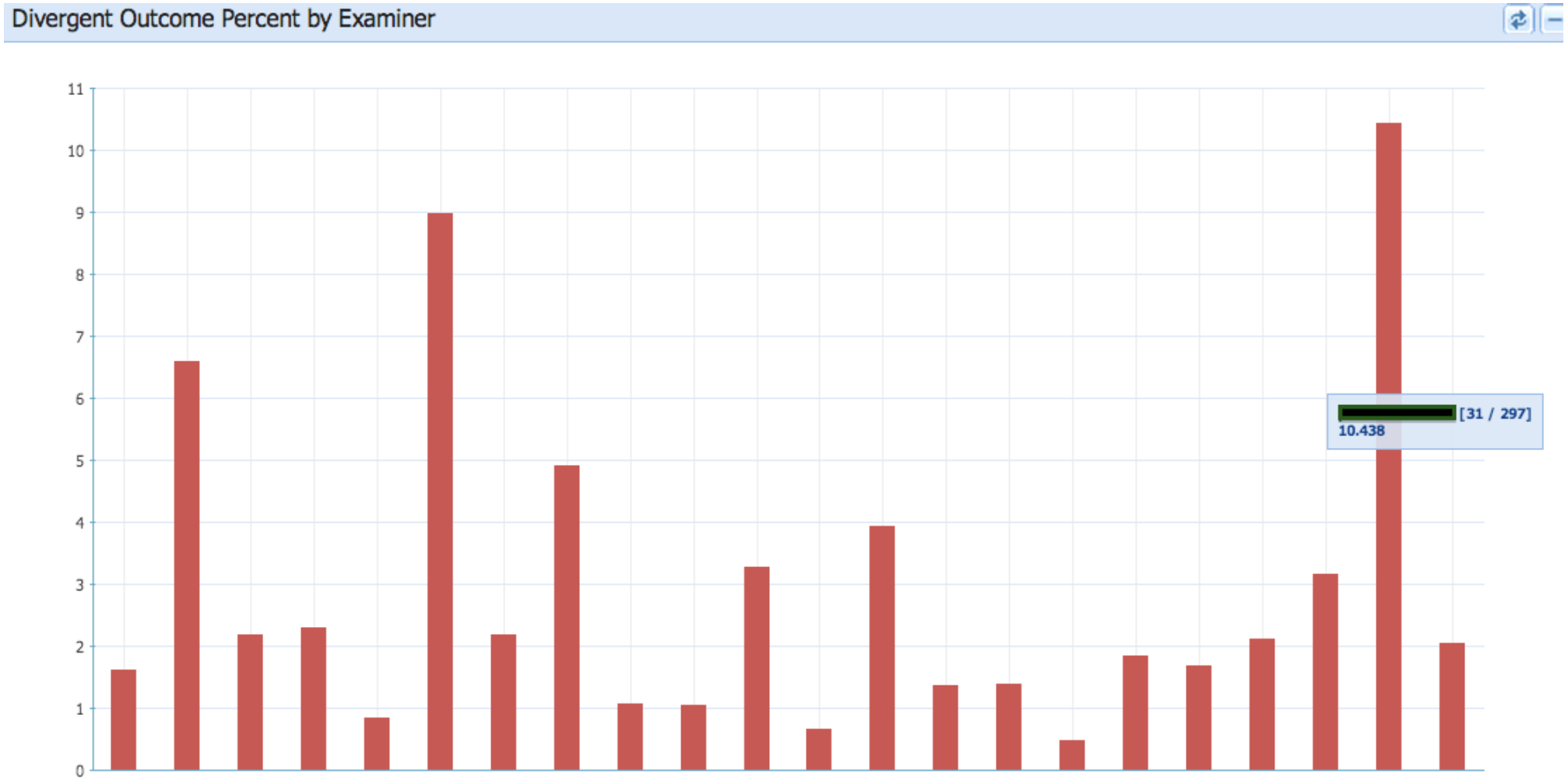
Your Marks	Correct Grades
External Outline Form: A+ (1)	External Outline Form: D (2)
No errors	Buccal-lingual width too wide Mesial occlusal Minor damage to adjacent tooth Mesial tooth
Internal Form: E (1)	Internal Form: E (1)
Axial wall too deep > 3.0mm Distal	Axial wall too deep > 3.0mm Distal
Finish: D (1)	Finish: D (1)
Presence of debris Mesial box	Presence of debris Mesial box

What measurements *can* you make?

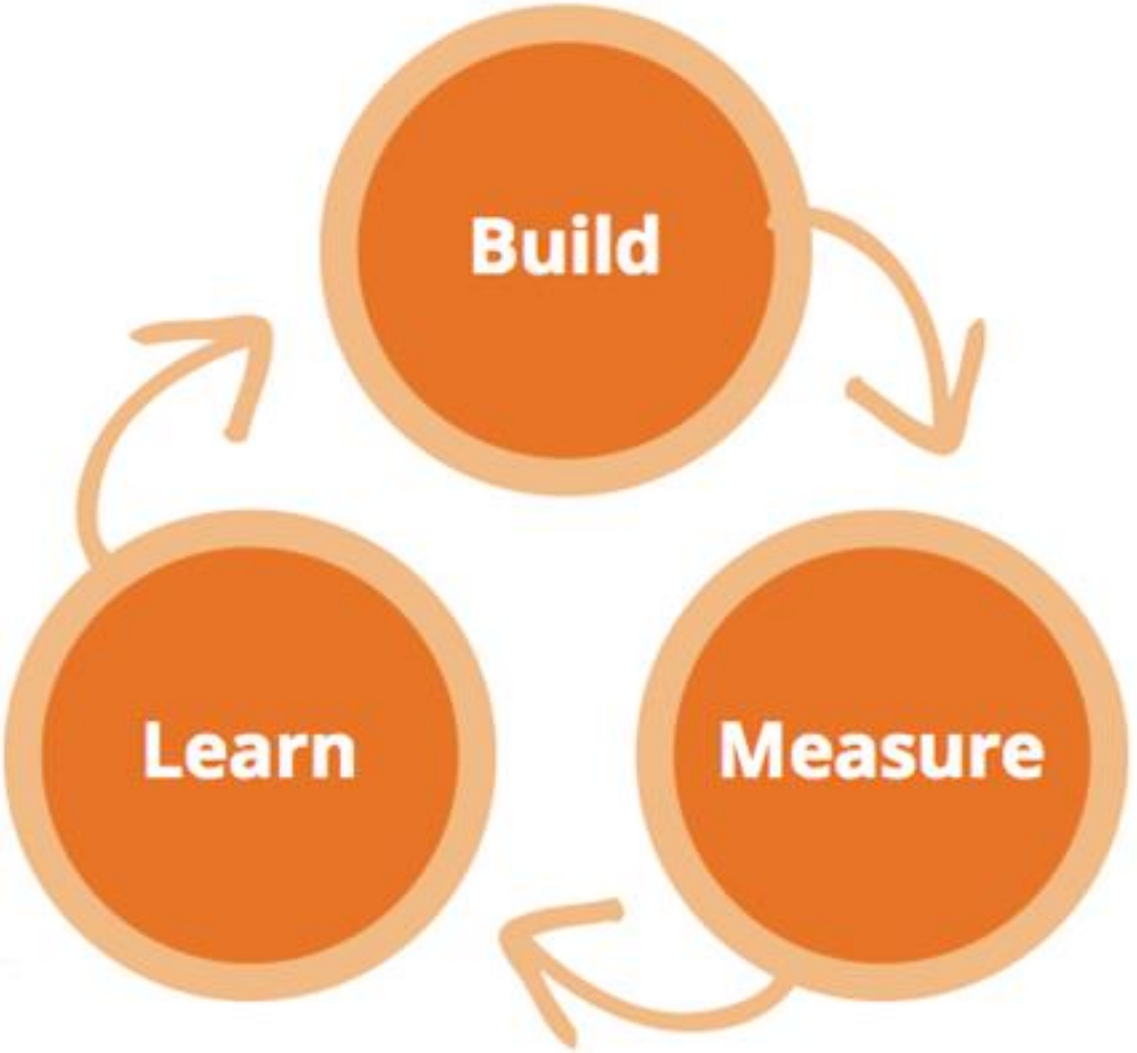
- Cronbach's Alpha
- Divergent Outcome %
- Mean score + Standard Deviation (normal distribution)
- Decision Consistency

ATP Innovations in Testing 2015
Palm Springs, CA

Measurement of Measurement



Refine the process



Mistake #4: Inadequately setting candidate expectations

- Symptoms include
  - Candidates may not "**prepare**" for the exam
  - Candidates may ignore or not fully review the online information including the Candidate Orientation video and Evaluation Forms
  - Cavalier attitude by more experiences Candidates
  - Need to take a systematic approach to the examination without "cutting corners"

- Problems usually stem from failure to account for the following…
  - The examination doesn't exactly match the steps used in actual practice

ATP Innovations in Testing 2015 Palm Springs, CA

How to avoid

- COMMUNICATION, COMMUNICATION, COMMUNICATION
  - Liaison with the schools and colleges of optometry emphasizing the importance of thorough preparation, especially by more seasoned student-clinicians
  - TestPoint newsletter to students with Part III CSE articles - communicate
  - Articles to the student association

Mistake #5: Allowing one (or a few) individuals too much control

ATP Innovations in Testing 2015 Palm Springs, CA

Symptoms include…

- One examiner can determine the outcome for an individual candidate
- Bottlenecks around one or a small group of examiners
- Process improvement slows, stalls, or is non-existent

Problems stem from failure to…

- Distribute evaluation responsibility
- Manage ego's
- Prevent "Over Specialization" in an area of the examination

How to avoid this mistake

- Exam process (See OSCE model)
- Distributed control
- Examination by team (multiple people with the same role)
- Manage the Tension: don't allow people to stick around too long in the same position

Questions?

Interactive exercise

ATP Innovations in Testing 2015 Palm Springs, CA

Exercise Instructions

1. Read the following information regarding the CECB program
2. Consider the discussion questions with your neighbors
3. Share feedback with the rest of the room

ATP Innovations in Testing 2015 Palm Springs, CA

Case Study: Certified Expert Cat Burglar (CECB)

Program information:

Assesses and recognizes the skill set of candidates who design and execute cat burglaries:

- At an expert level, and
- In adherence with the ethical and safety guidelines set forth by the Association of Professional Precious Art and Jewel Thieves

Format:

- Human-scored performance assessment
- 6-hour administration
- Candidates provided with prompts to carry out a simulated bank heist

Volume:

Approximately 150 candidates per administration

ATP Innovations in Testing 2015 Palm Springs, CA

CECB scoring and data handling procedures

- Each candidate followed from station to station by 3 raters.
- Each rater records his/her scores on a scoring sheet with a #2 pencil or black ink pen.
- Following administration, raters return scoring sheets to the Central Tabulator, who enters scores into a computer

Discussion:

1. What types of risks does the CECB program run using this methodology?
2. How might the CECB program mitigate those risks?

ATP Innovations in Testing 2015 Palm Springs, CA

CECB scoring procedures

- Raters rate candidates using non-empirical scoring criteria
- Raters are not given specific scoring instructions, as the rating sheet is self-explanatory
- Outlier ratings may be altered by the program's Chief Rater
- The following is an excerpt from the rating sheet:

Scoring area 1: Defeating Alarm Systems

Score	Points	Criteria
	3	Candidate creatively defeats alarm system
	2	Candidate defeats alarm system according to industry guidelines
	1	Candidate fails to comply with industry alarm defeating guidelines
	0	Candidate triggers alarm system

Discussion:

1. What types of risks does the CECB program run using this methodology?
2. How might the CECB program mitigate those risks?

CECB Candidate Preparation Information

- Following registration: Candidate are provided with a list of exam topics
- For security reasons, candidate informed of -- and trained on -- computer-simulated safe detonator interface the day of the examination

Discussion:

1. What types of risks does the CECB program run using this methodology?
2. How might the CECB program mitigate those risks?

Thank you!