**EPSY8225: Operational Measurement February 17, 2020**

**Class Notes**

**The QC Concerns of Performance Assessment**

The characteristics of PA present significant challenges regarding QC – conceptualization of the PA, measurement design, scoring, analysis, and reporting.

1. Scoring subjectivity – human scoring
2. Clarity and functionality of the task requirements and scoring guidelines
3. PA Measures the “can do” part of KSAs – the production of KSAs
4. PA includes more authentic tasks – what a person would do in the field – on the job
5. Based on job-task analyses or professional standards of what a person should be able to do
6. Proximity to the construct – more direct observation?

**Sources of Score Variability**

1. Candidates
2. Cheating
	1. Appropriate/inappropriate preparation
	2. …
3. Motivation
	1. performance orientation, outcome orientations
	2. intrinsic, extrinsic
4. effort (may be connected to motivation), engagement, attention
5. person behavior is different – different engagement
6. mental and physical condition
	1. fatigue
	2. anxiety
	3. nervousness
	4. stereo-type threat
7. repeated test taking – familiarity
8. instructional sensitivity (quasi-candidate characteristic)
9. Knowledge Skills Abilities
	1. Practice, instructional quality, experience, studying
	2. Opportunity to learn
10. Tasks
11. Level of difficulty
	1. Language complexity
	2. Assessment time
	3. Task material complexity – and familiarity
12. Cognitive complexity (intended or unintended)
	1. Does not necessarily correlate with task difficulty
13. Task quality
	1. Clarity – task specificity
	2. explicitness of instructions
	3. task relevance – construct representation
14. (task format)
15. Represent learning outcomes or expectations
	1. learning progression, grade relevant
	2. assume opportunity to learn (courts: instructional/curriculum validity)
16. Raters
17. Personal bias
	1. Personal experience of scoring
	2. Training on scoring
	3. Explicit training on bias (includes both the quality of training and the rater’s understanding of bias from the training)
	4. Cultural identity, gender identity, professional identity
	5. Subject-matter expertise
18. Subject-matter expertise
19. Mental and physical condition
	1. Energy level, fatigue
	2. Boredom
	3. Environment
	4. Sobriety
		1. Caffeine, nicotine, sugar
		2. Alcohol and other drugs
	5. Motivation
20. Job conditions
	1. Compensation
	2. Environment
	3. Coworkers (and all they bring to the scoring table) – morale
	4. Institutional values and respect for employees
	5. Friendly environment
	6. Flexibility of scoring routines (breaks, availability of water/snacks)
	7. Merit-based bonuses (positives & negatives)

Wolfe & Gitomer

Improving the PA tasks through principled design

1. Reduce guessing – added a section to the instructions: “How my response will be scored.”
2. Optimize decision – added a section: “Making good decisions.”
3. Provide more structure to the questions (the prompts)
4. Improve scoring
	1. More training samples
	2. More bias training
	3. Bridge questions for raters to focus on the variety of evidence available in a responses
5. Improve the rubric – add more focus and structure

**How to cheat?**

Where cheating is… the intent to deceive to obtain an unfair advantage

* + 1. Test taking behavior
	1. Copying from neighbor (may change seats to ensure high-ability neighbor)
	2. Memorizing content, questions, answers – passing them on
	3. Sneaking notes into test session, cheat sheets
	4. Confederate test takers, ghost writers
	5. Asking for help or assistance from others – nonverbal/verbal
	6. Mobile devices, other technology to obtain outside assistance
	7. Using support materials – unauthorized materials
	8. Falsify medical conditions to secure accommodations (test taker or parent)
	9. Talking during exam in any language
	10. Drugs - nonprescribed
		1. Administrator behavior
1. Confirming test taker identification
2. Test content on the walls
3. Inappropriate test administration
4. Coaching during the exam
5. Inappropriate room arrangements – other administration procedures are not followed
6. Changing answers
7. Deciding who takes the exam (sending kids home for the day)
8. Failing to report securing violations
9. Not following test security, storing materials securely
10. Providing improper accommodations (giving extra time, breaks)
11. Not maintaining secure test taking environment
12. Inappropriate test prep including test questions or question models (best preparation for achievement tests is instruction focused on learning objectives or state standards)
	* 1. Testing program characteristics (school, institution, test publisher)
13. Lack of or limited security protocol
14. Failure to adequately train test administrators
15. Failing to secure test content, testing materials
16. Failure to report or investigate violations
17. Failure to design test forms or administration settings that eliminate opportunities for test-takers to cheat
18. Decision making regarding test design that may secondarily benefit test security

**Detection**

* + 1. Statistical (probabilistic models – results could be due to chance)
1. Erasure patterns
2. Item response time analysis
3. Person fit statistics
4. Similarities in response patters – particularly incorrect answers
5. Similar response patterns by school
6. Unusual score gains/losses upon retake
	* 1. Pre-test administration observations
7. Search online sources of test prep (item banks)
8. Examine test prep materials – test prep companies – no secure test items
9. Security audit – confirm the security of test materials (untampered)
10. Revise security protocols based on prior administration experiences
	* 1. During test administration
11. Follow test security protocol for check-in (no devices); confirm identification
12. Randomize student seating and test booklet assignment
13. E-signal blocks, no access to wifi
14. Video/audio recording, drones
15. Maintain security procedures for allowing breaks
	* 1. Post-test administration
16. Maintaining material security – all materials are accounted for
17. Security audit
18. Search online for released test items (language varies)
19. Monitor test taker volumes and score trends (volume and performance)
20. Do a good job of designing tests with test security in mind (these are not always compatible)