



Research in Examination Issues:  
What's Happening Now?



2013  
Annual  
Educational  
Conference  
Hyatt Regency at the Arch | St. Louis  
Oct. 3-5, 2013



Council on Licensure, Enforcement & Regulation  
www.clearhq.org  
© CLEAR 2013

---

---

---

---

---

---

---

---

How Much Is Enough? Statistics with Small  
Sample Sizes

Daniel H. Breidenbach, PhD, Applied Measurement Professionals, Inc.



2013  
Annual  
Educational  
Conference  
Hyatt Regency at the Arch | St. Louis  
Oct. 3-5, 2013



Council on Licensure, Enforcement & Regulation  
www.clearhq.org  
© CLEAR 2013

---

---

---

---

---

---


---

---

How Much is Enough?

How many candidates do we need for  
“reliable” item statistics?

- Small volume programs use classical statistics to evaluate item performance
- Goal of an examination is good measurement
- Want good level of difficulty –  $p$
- Want good discriminating power –  $r_{pb}$



Council on Licensure, Enforcement & Regulation  
www.clearhq.org  
© CLEAR 2013

---

---

---

---

---

---

---

---

### Why does it matter?

#### What do we do with $p$ and $r_{pb}$ ?

- Item pre-testing (or “tryout” or pilot testing)
- Scrutinizing items for flaws
- Setting cut scores (pre-equating or post-test equating)
- Validity of pass/fail decisions!

© CLEAR 2013

Council on Licensure, Enforcement & Regulation  
www.clearhq.org

---

---

---

---

---

---

---

---

### How much does it matter?

#### An example:

- After 12 candidates,  $p = .5$  – hard item?
- Next 4 candidates get it wrong,  $p = .38$
- Next 4 candidates correct,  $p = .63$
- Hard item, or quite reasonable?
- Effect on cut score?
- Same situation when  $n = 100$ ;  $.48 \leq p \leq .52$

© CLEAR 2013

Council on Licensure, Enforcement & Regulation  
www.clearhq.org

---

---

---

---

---

---

---

---

### How many candidates needed?

- Simulation study - looking at variation in values for different group sizes
- “True”  $p$  and  $r_{pb}$  based on  $n = 1525$  - actual candidate responses
- 140 items on examination
- Simulated group sizes of 10, 15, 20, 25, 50, 75, 100, 125, and 150 candidates
- Looked at difference between sample statistics and “true” statistics

© CLEAR 2013

Council on Licensure, Enforcement & Regulation  
www.clearhq.org

---

---

---

---

---

---

---

---

### What to do with all that data?

- For each group size
  - Select  $n$  sets of candidate responses
  - Calculate  $p$  and  $r_{pb}$  for those  $n$  candidates
  - Record difference between sample stat and "true" stat
  - Repeat 100 times
- Criterion: want average error of .05 across iterations

© CLEAR 2013

Council on Licensure, Enforcement & Regulation  
 www.clearhq.org

---

---

---

---

---

---

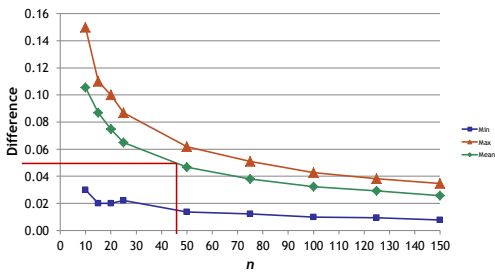
---

---

---

---

### Results? Difficulty



© CLEAR 2013

Council on Licensure, Enforcement & Regulation  
 www.clearhq.org

---

---

---

---

---

---

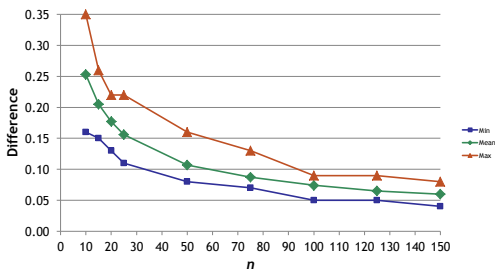
---

---

---

---

### Results? Discrimination



© CLEAR 2013

Council on Licensure, Enforcement & Regulation  
 www.clearhq.org

---

---

---

---

---

---

---

---

---

---

### What do we see?

- Smaller sample sizes have more variability
- More difficult items (low  $p$ ) have more variability than easier items
- Differences reach criterion with 50 candidates
- $r_{pb}$  values are less stable than  $p$  values
- Differences failed to reach criterion even with 150 candidates

© CLEAR 2013

Council on Licensure, Enforcement & Regulation  
www.clearhq.org

---

---

---

---

---

---

---

---

### What did we learn here?

- Difficulty estimates are not stable with  $n < 50$
- Discrimination indexes are particularly unstable with  $n < 150$
- Interpret statistics cautiously when sample size is low!

© CLEAR 2013

Council on Licensure, Enforcement & Regulation  
www.clearhq.org

---

---

---

---

---

---

---

---

### Non-Traditional Multiple Choice Items - Psychometric Issues

Steve Nettles, EdD, AMP Program Director



---

---

---

---

---

---

---

---

## History

- 1970's -problem-based approaches
  - patient management problems (PMPs) or simulations
  - used to assess decision-making skills.
- PMPs can have substantial evidence of content validity, and reliability.
  - Must be carefully developed and scored.

© CLEAR 2013

Council on Licensure, Enforcement & Regulation  
www.clearhq.org

---

---

---

---

---

---

---

---

## History

- PMPs are designed to test a specific content area in great depth.
  - Must carefully select group of problems to sample the content domain.
  - many PMPs must be included
    - To properly sample a complete content domain
    - To achieve sufficient reliability
  - Results in increased test administration time

© CLEAR 2013

Council on Licensure, Enforcement & Regulation  
www.clearhq.org

---

---

---

---

---

---

---

---

## Sample PMP

**PRACTICE SIMULATION PROBLEM**

**CARL RIVERS**

You are the respiratory therapist assigned to a 4-bed ICU on the midnight shift. On entering the unit, you find that 52-year-old Carl Rivers is being mechanically ventilated in the SIMV and pressure support mode, and appears to be resting comfortably. He is 182 cm (6 ft) tall and weighs 77 kg (170 lb).

During your initial ventilator check, which of the following would you evaluate?

**(SELECT AS MANY as you consider indicated in this Section, then click on the Go To Next Section button below to proceed to the next Section.)**

Exhaled tidal volume

Deep tendon reflexes

Respiratory flow rate

Ventilator mandatory rate

Color of urine

Maximum voluntary ventilation (MVV)

P102

Effective static (plateau) compliance

**Go To Next Section**

Candidate's Photo

**Current Section / Simulation History**

You are the respiratory therapist assigned to a 4-bed ICU on the midnight shift. On entering the unit, you find that 52-year-old Carl Rivers is being mechanically ventilated in the SIMV and pressure support mode, and appears to be resting comfortably. He is 182 cm (6 ft) tall and weighs 77 kg (170 lb).

During your initial ventilator check, which of the following would you evaluate?

**(SELECT AS MANY as you consider indicated in this Section, then click on the Go To Next Section button below to proceed to the next Section.)**

Ventilator mandatory rate  
8/min

**Go To Next Section**

03:50:49

© CLEAR 2013

Council on Licensure, Enforcement & Regulation  
www.clearhq.org

---

---

---

---

---

---

---

---

### PMP Issues

- Validity - Samples the content domain at great depth, but not breadth.
- Reliability - because of smaller number of items (problems) traditional reliability measures underestimate true reliability

© CLEAR 2013

Council on Licensure, Enforcement & Regulation  
www.clearhq.org

---

---

---

---

---

---

---

---

### Key Feature Problems (KFPs)

- Designed as an alternative to PMPs in the medical professions starting in the early 1990's.
- Designed to assess examinee ability to evaluate:
  - the critical features of a clinical situation, AKA
  - the unique, essential elements of the situation that result in resolution of the problem.
- Requires that decisions be made at specific decision points (*key features*) during the assessment and management of the case.

© CLEAR 2013

Council on Licensure, Enforcement & Regulation  
www.clearhq.org

---

---

---

---

---

---

---

---

### KFPs - Characteristics

- Requires candidates to:
  - apply knowledge to resolve a problem
  - Use knowledge to guide decisions
  - Formulate diagnostic/management impressions
  - Recommend investigative or follow-up procedures
  - Acquire data to monitor a course of action
  - Evaluate the probability of an outcome
  - Select a management course for the problem

© CLEAR 2013

Council on Licensure, Enforcement & Regulation  
www.clearhq.org

---

---

---

---

---

---

---

---

### KFPs - Benefits

- Problems are relatively short and focus on a few critical elements.
- Allows more problems to be included on an examination within a fixed period of time.
  - 10 key feature problems worth 90 points
  - 2 simulation problems worth 40 points
- More KFPs lead to broader sampling of the content domain
  - Potentially improving reliability and validity of test scores.

© CLEAR 2013

Council on Licensure, Enforcement & Regulation  
www.clearhq.org

---

---

---

---

---

---

---

---

### KFPs - Benefits

- Flexible on question formats and the number of responses to questions.
- Can use either selected response (MCQ) or constructed response.
  - (MCQs easier for computer based testing)
- Can have multiple correct answers for each question.

© CLEAR 2013

Council on Licensure, Enforcement & Regulation  
www.clearhq.org

---

---

---

---

---

---

---

---

A 10-month-old male, who was born at 28 weeks gestation with short bowel syndrome, is admitted to the hospital with lethargy, fever, and weight loss. The infant has a history of necrotizing enterocolitis and is status post-partial resection of the jejunum and majority of the ileum. The infant receives a combination of enteral, parenteral, and oral feedings at home.

Which of the following are MOST important to have upon starting initial assessment? Select FOUR of the following options.

- |  |   |
|--|---|
| A. blood pressure                        | H. nutrition history                    |
| B. blood omega-3 and omega-6 fatty acids | I. small bowel biopsy                   |
| C. bone age                              | J. stool output history                 |
| D. fecal elastase                        | K. tissue transglutaminase levels (TTG) |
| E. growth history                        | L. triene:tetraene                      |
| F. serum amylase and lipase              | M. stool pH and reducing substances     |

© CLEAR 2013

Council on Licensure, Enforcement & Regulation  
www.clearhq.org

---

---

---

---

---

---

---

---

## Standard Setting

- KFPs can have items with multiple correct answers.
- Traditional standard setting techniques (e.g., Modified - Angoff) need modifications to be applied to these alternative item types.
- Similar to Extended Angoff (Hambleton and Plake, 1995)

© CLEAR 2013

Council on Licensure, Enforcement & Regulation  
 www.clearhq.org

---

---

---

---

---

---

---

---

## Standard Setting Options

- Option 1 - Judges estimate the % of MCPs who will answer each correct option correctly
- Option 2 - Judges estimate the average score that 100 MCPs will obtain on each question (one or two decimals)

© CLEAR 2013

Council on Licensure, Enforcement & Regulation  
 www.clearhq.org

---

---

---

---

---

---

---

---

## Standard Setting - Option 1

Sequence	Item ID	Option 1					Mean	SD	Points
		Rater 1	Rater 2	Rater 3	Rater 4	Rater 5			
137	3B 4	80	85	80	85	80	82.0	2.7	2.5
*	*	70	80	75	80	80	77.0	4.5	
*	*	70	75	70	70	75	72.0	2.7	
138	3D 9	80	85	75	85	85	82.0	4.5	1.8
*	*	70	75	75	75	70	73.0	2.7	
139	1C 33	75	85	85	85	80	82.0	4.5	3.1
*	*	70	75	75	75	70	73.0	2.7	
*	*	70	75	75	80	75	75.0	3.5	
*	*	65	80	80	80	75	76.0	6.5	
Judge	M/SD	74	81	77	80	78	78.0	2.7	7.0

© CLEAR 2013

Council on Licensure, Enforcement & Regulation  
 www.clearhq.org

---

---

---

---

---

---

---

---



## Standard Setting - Option 2

Option 2								
Sequence	Item ID	Rater 1	Rater 2	Rater 3	Rater 4	Rater 5	Mean Points	SD
137	3B 4	2.2	2.1	2.3	2.2	2.0	2.2	0.1
138	3D 9	1.5	1.4	1.6	1.8	1.3	1.5	0.2
139	1C 33	2.9	3.2	2.8	3.1	3.3	3.1	0.2
Judge	Means/SD	2.2	2.2	2.2	2.4	2.2	6.8	0.1

© CLEAR 2013

Council on Licensure, Enforcement & Regulation  
 www.clearhq.org

---

---

---

---

---

---

---

---

---

---

## Psychometric Comparisons: PMPs vs. KFPs

- Validity - both based on Job/Practice Analyses
- Reliability
  - PMPs: vary from low 70's to low 90's
  - KFPs: vary from mid 70's to mid 80's
  - note: comparisons based on small number of candidates and limited KFP data

© CLEAR 2013

Council on Licensure, Enforcement & Regulation  
 www.clearhq.org

---

---

---

---

---

---

---

---

---

---

## Converting Paper Exams to CBT

Reed A. Castle, PhD  
 Professional Testing, Inc.

2013  
 Annual Educational Conference  
 Ryatt Regency at the Arch | St. Louis  
 Oct. 3-5, 2013

Council on Licensure, Enforcement & Regulation  
 www.clearhq.org  
 © CLEAR 2013

---

---

---

---

---

---

---

---

---

---

### General Thoughts

- Paper Advantage
  - Fixed time frame
  - Expense
    - Large groups
- CBT
  - Security
  - Convenience
  - Unique item types
  - Instant Scores

© CLEAR 2013 Council on Licensure, Enforcement & Regulation  
www.clearhq.org

---

---

---

---

---

---

---

---

### CBT Considerations

- Candidates
  - Access
    - More locations
    - Less travel costs
  - Scheduling
    - Take it close to home
    - Take it when you want
  - Scoring
    - Instant Scores

© CLEAR 2013 Council on Licensure, Enforcement & Regulation  
www.clearhq.org

---

---

---

---

---

---

---

---

### Organization

- CBT-Administrative Headaches relieved
  - Shipping
  - Facilities
  - Staff
  - Delayed scores
  - More Secure
- Some Providers will close certain months

© CLEAR 2013 Council on Licensure, Enforcement & Regulation  
www.clearhq.org

---

---

---

---

---

---

---

---

## Converting-Exam Type

- Linear Versus Adaptive
  - Bank requirements
  - Test Development requirements
  - Test Length
    - Adaptive shorter (less time and cost in the long run)
  - Content coverage

© CLEAR 2013

Council on Licensure, Enforcement & Regulation  
www.clearhq.org

---

---

---

---

---

---

---

---

## Studies adaptive transition

- Equivalency
- Assure same pass/fail decisions are made when going to adaptive
- Assure content bias does not creep in (e.g., content areas under representative)
- Exam will be challenging for both high and low ability candidates

© CLEAR 2013

Council on Licensure, Enforcement & Regulation  
www.clearhq.org

---

---

---

---

---

---

---

---

## Psychometric considerations

- Randomization of questions
  - Does this impact statistics or candidate anxiety
  - Traditional spiraling techniques
- Randomization of options
  - Literature has suggested that this can impact item difficulty
- Unique item types can create psychometric scoring challenges (e.g., item dependency, violations to test theory assumptions)
- Data-it can be very messy

© CLEAR 2013

Council on Licensure, Enforcement & Regulation  
www.clearhq.org

---

---

---

---

---

---

---

---

## Costs

- Paper considerations
  - Printing
  - Shipping
  - Staff resources
  - Rental costs
  - Administrator/Proctor Expenses
  - Exam day surprises (bands, medical, etc.)

© CLEAR 2013

Council on Licensure, Enforcement & Regulation  
www.clearhq.org

---

---

---

---

---

---

---

---

## CBT

- Get rid of the paper headaches, but have some unique considerations
- Publication fees (need for the QC)
- Seat time Fees
- Data handling
  
- International
  - Seat time
  - Availability of staff (time zones)

© CLEAR 2013

Council on Licensure, Enforcement & Regulation  
www.clearhq.org

---

---

---

---

---

---

---

---

## Cost Benefit analysis

- Identify all paper costs including potential threats
  - Administrative (staff time, printing, shipping, etc.)
  - Threats (cost of losing an exam book)
- Identify all CBT costs
  - Initial setup
  - Publication
  - Seat time
  - Initial psychometric investigations of transition

© CLEAR 2013

Council on Licensure, Enforcement & Regulation  
www.clearhq.org

---

---

---

---

---

---

---

---

### Decide

- Access and score reporting for candidates
- Amount of strain on the budget
- New programs or program with small numbers may not be able to afford
- Be realistic and base assumptions on the lowest realistic candidate volume

© CLEAR 2013

Council on Licensure, Enforcement & Regulation  
www.clearhq.org

---

---

---

---

---

---

---

---

### Speaker Contact Information

- Steve Nettles, EdD, Program Director, Applied Measurement Professionals, Inc. [snettl@goamp.com](mailto:snettl@goamp.com)
- Danny Breidenbach, PhD, Program Director, Applied Measurement Professionals, Inc. [dbreidenbach@goamp.com](mailto:dbreidenbach@goamp.com)
- Jodi Herold, PhD, Consultant, Toronto, Ontario, [jodi.herold@utoron.to.ca](mailto:jodi.herold@utoron.to.ca)
- Reed Castle, PhD, Executive Vice President-Partner, Professional testing, Inc., [rcastle@proftesting.com](mailto:rcastle@proftesting.com)



© CLEAR 2013

Council on Licensure, Enforcement & Regulation  
www.clearhq.org

---

---

---

---

---

---

---

---