Introductions

Dr. Brett Foley

- **Director of Professional Credentialing & Senior Psychometrician, Alpine Testing Solutions**

Joan Paros

- **Assistant Vice President of Examination, National Council of Architectural Registration Boards (NCARB)**

Dr. Amanda Wolkowitz

- **Senior Psychometrician, Alpine Testing Solutions**

Overview

- **Background**
- **Why make the switch?**
- **How to make the switch?**
- **NCARB Case Study Results**
- **Lessons Learned**
- **Questions**

Background

- **Quality vs. quantity of distractors**
- **MC3 items**
  - Perform equal or better to MC4 items
  - Answered quicker than MC4 items
  - Little impact on guessing behavior

- **Better questions and more time →**
  - Increase in exam reliability
  - Increase in content validity

Why did NCARB make the switch?

ASSESSMENT for a Better World

Architect Registration Examination® (ARE®)
- Consists of 6 separate divisions

Pre-pandemic challenges with the exam
- Item development & SME time on task
- Testing time is often fully used

Additional challenges due to the pandemic
- Venturing into online proctoring

How to make the switch

1. **Establish rule for identifying items with a non-functioning distractor (NFD)**
   - Example: Distractor selected by fewer than 5% of candidates or positive item-score correlation

Option	p	r
A	0.001	-0.130
B	0.100	-0.228
C (key)	0.728	0.406
D	0.172	-0.260

Option	p	r
A (key)	0.502	0.161
B	0.390	0.309
C	0.085	-0.141
D	0.023	-0.172

How to make the switch

1. **Establish rule for identifying items with a non-functioning distractor (NFD)**

   - Example: Distractor selected by fewer than 5% of candidates or positive item-score correlation

Option	p	r
A - NFD	0.001	-0.130
B	0.100	-0.228
C (key)	0.728	0.406
D	0.172	-0.260

Option	p	r
A (key)	0.502	0.161
B	0.390	0.309
C	0.055	-0.141
D	0.053	-0.172

How to make the switch

1. **Establish rule for identifying items with a non-functioning distractor (NFD)**
   - Example: Distractor selected by fewer than 5% of candidates or positive item-score correlation

Option	p	r
A - NFD	0.001	-0.130
B	0.100	-0.228
C (key)	0.728	0.406
D	0.172	-0.260

Option	p	r
A (key)	0.502	0.161
B - NFD	0.390	0.309
C	0.055	-0.141
D	0.053	-0.172

2. Fix Rasch item and person measures not impacted by MC3.

Fix "Green" items. They did not have a NFD, i.e., not MC3.

Fix all persons. Their ability did not change.

Person	Item 1	Item 2	Item 3 - NFD	Item 4	Item 5	Item 6 – NFD	Item 7	Item 8
1	●	●	●	●	●	●	●	●
2	●	●	NFD	●	●	NFD	●	●
3	●	●	●	●	●	●	●	●
4	●	●	NFD	●	●	●	●	●
5	●	●	●	●	●	●	●	●
6	●	●	●	●	●	●	●	●

How to make the switch

2. Freely calibrate Rasch measures impacted by MC3 in TWO WAYS:

*Assume those selecting the NFD respond to MC3 item **correctly***

Person	Item 1	Item 2	Item 3 - NFD	Item 4	Item 5	Item 6 – NFD	Item 7	Item 8
1	•	•	•	•	•	•	•	•
2	•	•	1	•	•	1	•	•
3	•	•	•	•	•	•	•	•
4	•	•	1	•	•	•	•	•
5	•	•	•	•	•	•	•	•
6	•	•	•	•	•	•	•	•

*Assume those selecting the NFD respond to MC3 item **incorrectly***

Person	Item 1	Item 2	Item 3 - NFD	Item 4	Item 5	Item 6 – NFD	Item 7	Item 8
1	•	•	•	•	•	•	•	•
2	•	•	0	•	•	0	•	•
3	•	•	•	•	•	•	•	•
4	•	•	0	•	•	•	•	•
5	•	•	•	•	•	•	•	•
6	•	•	•	•	•	•	•	•

EASIER ITEM
(Lower Rasch value)

ACTUAL RASCH VALUE LIKELY SOMEWHERE IN BETWEEN EXTREMES

MORE DIFFICULT ITEM
(Higher Rasch value)

How to make the switch

3. **Estimate cut score TWO WAYS:**
   a. *Using lower range Rasch estimates (NFD = 1)*
   b. *Using upper range Rasch estimates (NFD = 0)*

4. **Assemble pre-equated forms so the two cut scores are as similar as possible**

   Ideally, both round to the same value

Note: Limit the number of converted items on a form.

NCARB Case Study Results

- **NCARB had 4 pre-equated forms assembled for each of 6 divisions of the Architect Registration Examination (ARE®)**

- **2 forms in each division released in Dec. 2020 with converted MC3 items. Scores were held to verify cut scores.**

- **Final 2 forms in each division released in Feb. 2021. Scores were not held, but were verified.**

ASSESSMENT for a Better World

NCARB Case Study Results

Converted Items

- Small % of MC items converted (10-20%)

- Rasch item difference between NFD = 0 and NFD = 1 small (typically < 0.10)

- Difference between the two estimated cut scores were all within 0.15 and near an integer cut score (natural rounding rules used)

- MC3 items were not anchored between the two sets of released forms. Of the 49 MC3 items appearing on both sets, the difference between the calibrated MC3 parameters were:
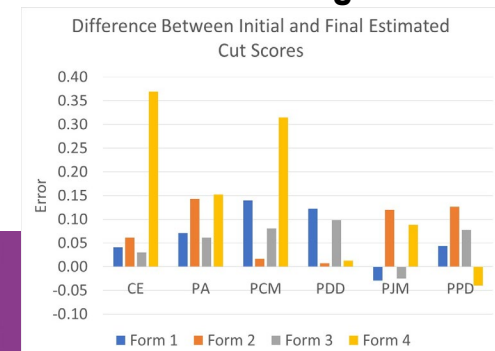
	Within 0.10	Within 0.25	Within 0.40	Within 0.50	Within 0.90
% of Common MC3 Items	24%	71%	84%	92%	100%

Evidence of success

- Of the 176 MC3 items, the difference between the estimated and calibrated item measures were:

	Within 0.10	Within 0.25	Within 0.40	Within 0.50	Within 0.90
% of MC3 items	25%	65%	79%	88%	93%

- 100% accuracy within rounding of initial and final estimated cut scores for ALL 24 forms

- Greatest difference between initial and final cut scores was 0.37 points.

- 93% (n = 22) of final cut scores were within ≈0.15 points of initial cut score range

Difference Between Initial and Final Estimated Cut Scores

Lessons Learned

NCARB perspective:

- Candidates have had no problems adapting to MC3 items
- Candidates are not confused by some MC4 and some MC3
- Stakeholders are comfortable with this change
- We probably could have done *more* MC4 to MC3 conversions

Psychometric perspective:

- Careful planning and small number of operational conversions was beneficial
- On average, MC3 items were just *slightly* easier than MC4.
- On average, MC3 items were answered more quickly
- Preliminary results show no negative impact on item-score correlation

Recommendations

- **Carefully define the NFD**
- **Transition a modest number of items**
- **Hold scores to confirm accuracy of cut score estimates, if possible**
- **Only do it once – pretest MC3 items moving forward**
- **It is worth it! Easier and less expensive to develop and candidates don't complain about having fewer options!**

Questions?

Brett.Foley@alpintesting.com

jparos@ncarb.org

Amanda.Wolkowitz@alpinetesting.com

Thank you!