2020 ICE EXCHANGE

ACCREDITATION

INNOVATION

CREDENTIALING

EDUCATION

2020 ICE EXCHANGE

Converting 4-Option Multiple Choice Questions to 3-Option and Integrating Both into Your Exam

Alpine Testing Solutions
Amanda A. Wolkowitz, Ph.D., Sr Psychometrician
Brett Foley, Ph.D., Sr Psychometrician, Dir. Professional Cred.
Corina Owens, Psychometrician

NCARB
Jared Zurn, AIA, NCARB, CAE, Vice President, Examination

ADOBE
Jim Mendes, Certification Developer

Overview

- Why make the switch from 4-option multiple choice (MC4) to 3-option (MC3)?
- How to make the switch?
- Final thoughts
- Questions

NCARB – Why make the switch?

- 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

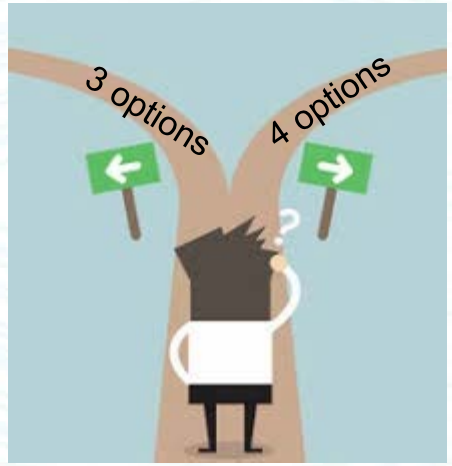
ADOBE – Why make the switch?

- Exam and item-level performance improvements
  - Validity (are we able to measure what we say we are measuring?)
  - Reliability (can we do it over and over and get the same results?)
  - Reduction of bias in our exams (Is it fair to all candidates?)
- Benefits to candidates
  - Less time to complete an item
  - Taking “better items” (see above)
- Benefits to Adobe Program
  - Easier (faster) to write
  - Face validity related benefits

Background on MC3 vs MC4

- MC3 vs. MC4 items is not a new topic and convincing one to switch is not the focus of our presentation
- Bottom line of research:
  - Quality of distractors more important than quantity
  - 3+ distractors can be used IF they are based on common errors
  - MC3 perform just as well or statistically better than MC4 items
- Quicker to answer → More items on exam →
  - Higher exam reliability
  - Higher content validity

Why are there hesitations to switch?



- Not many exams currently use MC3 items.
- Stakeholders may not believe that MC3 items are as good as (if not better than) other item types.
- Belief that it is easier to harvest MC3 items compared to MC4 items.
- Belief that guessing chances increases.

Making the Switch: Wholesale or Incremental Change?

1. Convert whole bank or only select items?

2. Use operationally or pilot test first?

Making the Switch: Which option to remove?

Pros and cons of different methods of reducing 4 options to 3 options:

1) Randomly select a distractor to delete

2) Select the worst performing distractor to delete based on expert judgement

3) Select the worst performing distractor to delete using empirical evidence

Making the Switch: Which option to remove?

Pros and cons of different methods of reducing 4 options to 3 options:

1) Randomly select a distractor to delete

2) Select the worst performing distractor to delete based on expert judgement

3) Select the worst performing distractor to delete using empirical evidence

Making the Switch: Which items to convert?

- Select the worst performing distractor to delete using empirical evidence
- Remove only items with a poorly performing distractor
- Now, what is the empirical method? Should we eliminate the option if the proportion selecting the distractor (p_D)
  A. <5%
  B. <1%
  C. <5%·q
  D. <5%·q *or* $r_D > 0$

  Note: q = 1-pvalue, r_D = item-total score correlation of distractor

Making the Switch: Which items to convert?

- Select the worst performing distractor to delete using empirical evidence
- Now, what is the empirical method? Should we eliminate the option if the proportion selecting the distractor (p_D)
  A. <5%
  B. <1%
  C. <5%·q
  D. <5%·q *or* $r_D > 0$

Note: q = 1-pvalue, r_D = item-total score correlation of distractor

Making the Switch: Which option to remove?

- Eliminate the worst performing distractor:

$$p_D < 5\% \cdot q \textit{ or } r_D > 0$$

- Tie breaker
  - Select distractor with less endorsement (i.e., smaller p_D)
  - If two distractors have equal endorsement, select distractor with higher correlation (i.e. higher r_D)
- Example 1:
  - p-value = 0.728
  - q-value = 1-0.728 = 0.272
  - 5%·q = 0.00136

Option	p	r	Is p_D < 5%q?	Is r > 0?
A	0.001	-0.130	Yes	No
B	0.100	-0.228	No	No
C *key*	0.728	0.406	---	---
D	0.172	-0.260	No	No

Making the Switch: Which option to remove?

- Eliminate the worst performing distractor:

$$p_D < 5\% \cdot q \textit{ or } r_D > 0$$

- Tie breaker
  - Select distractor with less endorsement (i.e., smaller p_D)
  - If two distractors have equal endorsement, select distractor with higher correlation (i.e. higher r_D)
- Example 1:
  - p-value = 0.728
  - q-value = 1-0.728 = 0.272
  - 5%·q = 0.00136

Option	p	r	Is p_D < 5%q?	Is r > 0?
A	**0.001**	-0.130	**Yes**	No
B	0.100	-0.228	No	No
C *key*	0.728	0.406	---	---
D	0.172	-0.260	No	No

Making the Switch: Which option to remove?

- Eliminate the worst performing distractor:

$$p_D < 5\% \cdot q \textit{ or } r_D > 0$$

- Tie breaker
  - Select distractor with less endorsement (i.e., smaller p_D)
  - If two distractors have equal endorsement, select distractor with higher correlation (i.e. higher r_D)
- Example 2:
  - p-value = 0.204
  - q-value = 1-0.204 = 0.796
  - 5%·q = 0.0398

Option	p	r	Is p_D < 5%q?	Is r > 0?
A *key*	0.204	0.167	---	---
B	0.394	-0.008	No	No
C	0.254	-0.158	No	No
D	0.136	0.035	No	**Yes**
Omit	0.011	-0.060	---	---

Making the Switch: Which option to remove?

- Eliminate the worst performing distractor:

$$p_D < 5\% \cdot q \textit{ or } r_D > 0$$

- Tie breaker
  - Select distractor with less endorsement (i.e., smaller p_D)
  - If two distractors have equal endorsement, select distractor with higher correlation (i.e. higher r_D)
- Example 2:
  - p-value = 0.204
  - q-value = 1-0.204 = 0.796
  - 5%·q = 0.0398

Option	p	r	Is p_D < 5%q?	Is r > 0?
A *key*	0.204	0.167	---	---
B	0.394	-0.008	No	No
C	0.254	-0.158	No	No
D	0.136	**0.035**	No	**Yes**
Omit	0.011	-0.060	---	---

Making the Switch: Which option to remove?

- Eliminate the worst performing distractor:

$$p_D < 5\% \cdot q \textit{ or } r_D > 0$$

- Tie breaker
  - Select distractor with less endorsement (i.e., smaller p_D)
  - If two distractors have equal endorsement, select distractor with higher correlation (i.e. higher r_D)
- Example 3:
  - p-value = 0.502
  - q-value = 1-0.502 = 0.498
  - 5%·q = 0.0249

Option	p	r	Is p_D < 5%q?	Is r > 0?
A *key*	0.502	0.161	---	---
B	0.390	0.309	No	**Yes**
C	0.085	-0.141	No	No
D	0.023	-0.172	**Yes**	No

Making the Switch: Which option to remove?

- Eliminate the worst performing distractor:

$$p_D < 5\% \cdot q \textit{ or } r_D > 0$$

- Tie breaker
  - Select distractor with less endorsement (i.e., smaller p_D)
  - If two distractors have equal endorsement, select distractor with higher correlation (i.e. higher r_D)
- Example 3:
  - p-value = 0.502
  - q-value = 1-0.502 = 0.498
  - 5%·q = 0.0249

Option	p	r	Is p_D < 5%q?	Is r > 0?
A *key*	0.502	0.161	---	---
B	0.390	**0.309**	No	**Yes**
C	0.085	-0.141	No	No
D	**0.023**	-0.172	**Yes**	No

Making the Switch: Which option to remove?

- Eliminate the worst performing distractor:

$$p_D < 5\% \cdot q \textit{ or } r_D > 0$$

- Tie breaker
  - Select distractor with less endorsement (i.e., smaller p_D)
  - If two distractors have equal endorsement, select distractor with higher correlation (i.e. higher r_D)
- Example 3:
  - p-value = 0.502
  - q-value = 1-0.502 = 0.498
  - 5%·q = 0.0249

Option	p	r	Is p_D < 5%q?	Is r > 0?
A *key*	0.502	0.161	---	---
B	0.390	**0.309**	No	**Yes**
C	0.085	-0.141	No	No
D	**0.023**	-0.172	**Yes**	No

Making the Switch: Which option to remove?

- Eliminate the worst performing distractor:

$$p_D < 5\% \cdot q \textit{ or } r_D > 0$$

- Tie breaker
  - Select distractor with less endorsement (i.e., smaller p_D)
  - If two distractors have equal endorsement, select distractor with higher correlation (i.e. higher r_D)
- Example 4:
  - p-value = 0.602
  - q-value = 1-0.602 = 0.398
  - 5%·q = 0.0199

Option	p	r	Is p_D < 5%q?	Is r > 0?
A *key*	0.602	0.349	---	---
B	0.367	-0.291	No	No
C	**0.015**	-0.099	Yes	No
D	**0.015**	-0.152	Yes	No

Making the Switch: Which option to remove?

- Eliminate the worst performing distractor:

$$p_D < 5\% \cdot q \textit{ or } r_D > 0$$

- Tie breaker
  - Select distractor with less endorsement (i.e., smaller p_D)
  - If two distractors have equal endorsement, select distractor with higher correlation (i.e. higher r_D)
- Example 4:
  - p-value = 0.602
  - q-value = 1-0.602 = 0.398
  - 5%·q = 0.0199

Option	p	r	Is p_D < 5%q?	Is r > 0?
A *key*	0.602	0.349	---	---
B	0.367	-0.291	No	No
C	**0.015**	-0.099	**Yes**	No
D	**0.015**	-0.152	**Yes**	No

Making the Switch: Which option to remove?

- Eliminate the worst performing distractor:

$$p_D < 5\% \cdot q \textit{ or } r_D > 0$$

- Tie breaker
  - Select distractor with less endorsement (i.e., smaller p_D)
  - If two distractors have equal endorsement, select distractor with higher correlation (i.e. higher r_D)
- Example 4:
  - p-value = 0.602
  - q-value = 1-0.602 = 0.398
  - 5%·q = 0.0199

Option	p	r	Is p_D < 5%q?	Is r > 0?
A *key*	0.602	0.349	---	---
B	0.367	-0.291	No	No
C	**0.015**	-0.099	**Yes**	No
D	**0.015**	-0.152	**Yes**	No

Making the Switch: Which option to remove?

- Eliminate the worst performing distractor:

$$p_D < 5\% \cdot q \textit{ or } r_D > 0$$

- Tie breaker
  - Select distractor with less endorsement (i.e., smaller p_D)
  - If two distractors have equal endorsement, select distractor with higher correlation (i.e. higher r_D)
- Example 4:
  - p-value = 0.602
  - q-value = 1-0.602 = 0.398
  - 5%·q = 0.0199

Option	p	r	Is p_D < 5%q?	Is r > 0?
A *key*	0.602	0.349	---	---
B	0.367	-0.291	No	No
C	**0.015**	-0.099	**Yes**	No
D	**0.015**	-0.152	**Yes**	No

Making the Switch: Which option to remove?

- Eliminate the worst performing distractor:

$$p_D < 5\% \cdot q \textit{ or } r_D > 0$$

- Tie breaker
  - Select distractor with less endorsement (i.e., smaller p_D)
  - If two distractors have equal endorsement, select distractor with higher correlation (i.e. higher r_D)
- Example 5:
  - p-value = 0.559
  - q-value = 1-0.559 = 0.441
  - 5%·q = 0.022

Option	p	r	Is p_D < 5%q?	Is r > 0?
A	0.197	-0.137	No	No
B *key*	0.559	0.334	---	---
C	0.108	-0.181	No	No
D	0.136	-0.160	No	No

Making the Switch: Which option to remove?

- Eliminate the worst performing distractor:

$$p_D < 5\% \cdot q \textit{ or } r_D > 0$$

- Tie breaker
  - Select distractor with less endorsement (i.e., smaller p_D)
  - If two distractors have equal endorsement, select distractor with higher correlation (i.e. higher r_D)
- Example 5:
  - p-value = 0.559
  - q-value = 1-0.559 = 0.441
  - 5%·q = 0.022

Option	p	r	Is p_D < 5%q?	Is r > 0?
A	0.197	-0.137	No	No
B *key*	0.559	0.334	---	---
C	0.108	-0.181	No	No
D	0.136	-0.160	No	No

Not a candidate for converting to MC3. Leave as MC4.

MC4 to MC3 in action: Case Study

- Step 1. Identify items with a NFD distractor

Division	# MC4 Items Analyzed	# Items with an NFD Identified	% of Items with an NFD Identified
CE	123	55	45%
PA	91	32	35%
PCM	78	27	35%
PDD	191	85	45%
PJM	109	56	51%
PPD	137	54	39%

MC4 to MC3 in action: Case Study

- Step 2. Calculate Rasch item measure bounds:
  - Fix Rasch measures for all items EXCEPT those being converted to MC3.
  - Fix all Person measures.
  - Freely calibrate Rasch item measure for converted items two ways:

Candidate ability does not change based on item type

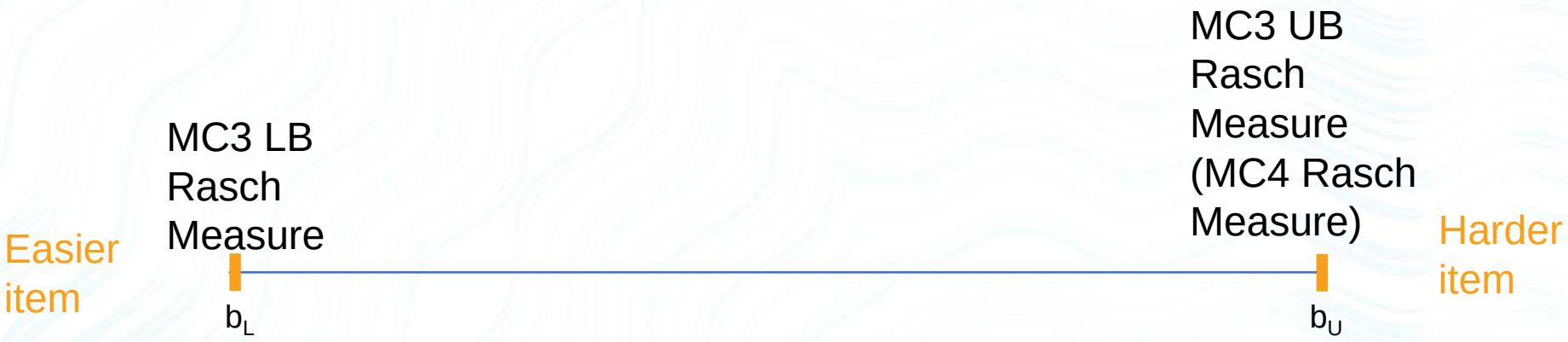
MC4 Rasch Measure

1) Assume all candidates who selected the NFD would answer the item correctly if it were an MC3 item, i.e. item is easier than MC4 item

2) Assume all candidates who selected the NFD would answer the item incorrectly if it were an MC3 item, i.e. item is harder than MC4 item

MC4 to MC3 in action: Case Study

- Step 3. Estimate the potential cut score change.



Difference in contribution to cut score = U – L, where

$$U = P(\theta) = \exp(\theta - b_U)/(1+\exp(\theta - b_U)) \text{ and}$$

$$L = P(\theta) = \exp(\theta - b_L)/(1+\exp(\theta - b_L))$$

MC4 to MC3 in action: Case Study

- Step 3. Estimate the potential cut score change.

Difference in Contribution to Cut Score	CE	PA	PCM	PDD	PJM	PPD
< 0.01	9	4	4	8	8	5
0.010 – 0.019	12	5	7	6	13	7
0.020 – 0.029	8	4	3	12	10	13
0.030 – 0.039	6	3	1	10	7	6
0.040 – 0.049	1	4	3	7	2	4
0.050 – 0.059	2	0	0	5	2	2
0.060 – 0.069	2	1	2	5	3	1
0.070 – 0.079	5	1	1	3	2	0
0.080 – 0.089	3	2	1	2	0	1
0.090 – 0.099	1	2	1	3	0	0
≥ 0.10	6	6	4	24	9	15
TOTAL	55	32	27	85	56	54

MC4 to MC3 in action: Case Study

- Step 4. Assemble new forms.
  - Goal for there to be no difference, within rounding, of the cut score when using MC3 items in place of MC4 items.
- Step 5. Administer forms.
  - Hold scores (we're pretty confident this method will work, but we want to be certain!)
  - Calibrate MC3 items.
  - Confirm cut score.
  - Release scores.
  - Have a backup plan – just in case.

Thoughts from going through process - NCARB

- Stakeholder acceptance was uneventful
- Allowed assurance to candidates that MC3 matter
  - Could affirmatively state that MC3s are operational
- From an item writer's perspective
  - Provided guidelines of when MC4 still makes sense
  - SMEs embraced the change quickly

Thoughts about potentially going through process

- Initial appeal is potential reduction in SME time to develop items
- Should we be concerned with a 33% guessing strategy vs. 25%
- Are there any candidate face-validity related concerns?
- Would we go back and apply to existing exams or implement with only new exams?
- Finally, will it produce a better performing item and exam?

Final Thoughts and Recommendations

- Only convert live MC4 to MC3 items during one administration period
  - This prevents candidates from knowing whether or not an MC3 item is pretest or scored when they are first introduced into testing.
  - Pretest MC3 items going forward to avoid the additional analyses and uncertainty.
- There is no rule that says you can only use MC4 items on an exam. Mix them up!
- MC3 items are easier to write and quicker to answer.
- MC4 items have their benefit, too, so you don’t need to eliminate them!

2020 ICE EXCHANGE

Thank you!

Questions?

Alpine

Amanda Wolkowitz, Ph.D.	Amanda.Wolkowitz@alpinetesting.com
Brett Foley, Ph.D.	Brett.Foley@alpinetesting.com
Corina Owens, Ph.D.	Corina.Owens@alpinetesting.com

NCARB

Jared Zurn, AIA, CAE, NCARB	jzurn@ncarb.org

ADOBE

Jim Mendes	jmendes@adobe.com