# What are the Optimum Test Development Decisions for Making Situational Judgment Tests (SJTs)?

# Evidence from SJTs of emotional intelligence and personality

**Carolyn MacCann**
School of Psychology
University of Sydney, Sydney, Australia

Email: carolyn.maccann@sydney.edu.au

# Summary

1. **Brief Background—Emotional Intelligence**

2. **Assessment Development Summary**
   - Situational Test of Emotion Management (STEM)
     - Multi-media adaptation (multi-media emotion management assessment; MEMA)
     - Downward extension for children
   - Personality Assessment (Dependability)

3. **Test development decisions and validity evidence:**
   a. Item formats (MCQ vs ratings);
   b. Instruction types ("would do" versus "should do");
   c. Mode (text-only vs video);
   d. Scoring procedures

# Part 1

# Background on Emotional Intelligence

# Brief Background: EI

## Mixed Models of EI

- EI is a mix of constructs
  - character traits, emotion-related abilities, motivation, beliefs, etc.
  - Usually measured with self or other-rating scale items

  Example: I know when to speak about my emotional problems to others
  - 1 – Strongly disagree
  - 2 – Disagree
  - 3 – Neither agree nor disagree
  - 4 – Agree
  - 5 – Strongly agree

**Ability EI predicts:**
1. Academic achievement ($k$ = 60, $N$ = 8472, **ρ = .30**)
2. Workplace performance ($k$ = 10, $N$ = 887, **ρ = .17**)
3. Medical school grades ($N$ = 367)
   - **.23** (interpersonal) vs **.08** (academic)

## Ability Model of EI

- EI is a set of abilities involved in processing and manipulating emotional information
  - Measured by ability items – require test taker to process/evaluate emotional information

**Emotional Intelligence**

- Emotion Management
- Emotion Understanding
- Emotion Facilitation
- Emotion Perception

**????**

1. MacCann et al., in prep;  2. Joseph & Newman, 2010, *Journal of Applied Psychology*;  3. Libbrecht, Lievens, Carette, & Côté, 2014, *Emotion*

# Part 2

# Assessment Development Summary

# Test Development:
## Situational Test of Emotion Management (STEM)

- Situational Judgment Test (SJT) developed in three steps:
    1. **Situations generated**: Semi-structured interviews to elicit everyday emotional events ($N$ = 51, ½ students)
    2. **Responses generated**: Free response "**best**" and **worst** answers to 138 situations (n = 30 for 3 x 46 items each)
    3. **Scoring keys generated**:
        1. 12 experts (1 excluded) *pick best response* – score = *proportion of experts rating that option*
        2. 6 experts *rate the effectiveness of each response* – score = *distance from average expert rating (& variants)*
        - **NOTE:** Can also generate scoring key from non-expert samples (*consensus scoring*)
- **44 items (later short form – 18 items)**

Mario is showing Min, a new employee, how the system works. Mario's boss walks by and announces that Mario is wrong about several points, as changes have been made. Mario gets on well with his boss, although they do not normally have much to do with each other.

A. Make a joke to Min, explaining he didn't know about the changes
B. Not worry about it, just ignore the interruption
C. Learn the new changes
D. Tell the boss that such criticism was inappropriate.

MacCann, C., & Roberts, R. D. (2008). New paradigms for assessing emotional intelligence: Theory and data. *Emotion, 8*, 540-551.

Allen, V., Rahman, N., Weissman, A., MacCann, C., Lewis, C., & Roberts, R. D. (2015). The Situational Test of Emotional Management–Brief (STEM-B): Development and validation using item response theory and latent class analysis. *Personality and Individual Differences, 81*, 195-200.

# Multi-media Adaptation
## Multimedia Emotion Management Assessment (MEMA)

| Task | Question Number | | Testing Tools |
|---|---|---|---|
| Situational Judgment Section 1 | 1 of 2 | | ◄ Back  ► Next |

Brad Alcott is a college student who has come to see Professor Carson during office hours.

Use the PLAY button to view their dialogue on Brad's academic progress.  Then click "Next" in order to see four potential responses from Brad.

MacCann, C., Lievens, F., Libbrecht, N., & Roberts, R. D. (2016). Differences between multimedia and text-based assessments of emotion management: An exploration with the multimedia emotion management assessment (MEMA). *Cognition and Emotion, 30,* 1317-1331.

16 scenarios
Each with 4 possible responses
$\alpha = .61$

Use each PLAY button to view the four potential responses from Brad. Then indicate which of these responses would be BEST.

**Click on the BEST choice.**

**You can replay any of the videos by clicking on the Play buttons.**
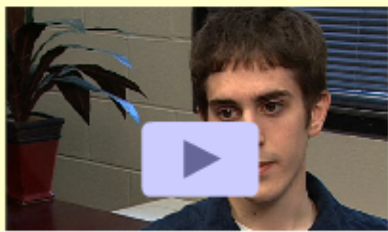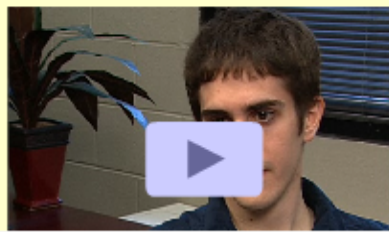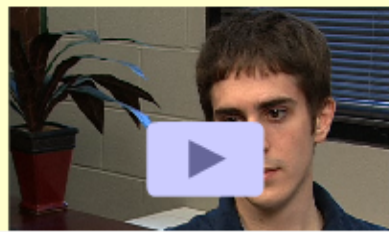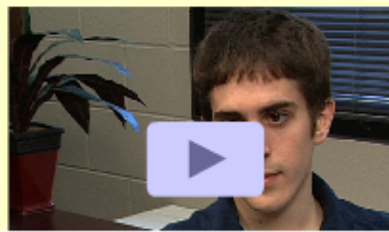
A

B

C

D

**Table 1.** Illustration of how the original text-based STEM item was transformed into a script for filming of the MEMA.

| STEM item | Script for MEMA item |
|---|---|
| *Scenario* | *Scenario (video: 45sec)* |
| Mario is showing Min, a new employee, how the system works. Mario's boss walks by and announces Mario is wrong about several points, as changes have been made. Mario gets on well with his boss, although they do not normally have much to do with each other | *Maria (a Hispanic female, age unimportant) is standing with James (an Asian-American male, age unimportant) in front of a filing cabinet*<br>   Maria [to James]: So it's really pretty straightforward. As soon as the invoices are sent down to Laura, just create a new red folder and file them by invoice number. Make sure you also create a green file and put it under the customer's name at the same time, because if you forget about it, that can cause lots of problems down the road<br>   JAMES [a little overwhelmed]: So, wait—the RED file is for the new invoices? Aren't some of the invoice files yellow?<br>*Angela (a white woman, mid-forties to mid-fifties) overhears the conversation as she walks by. She stops, listens as James finishes his question, and interrupts before Maria can respond* |
| What action would be the most effective for Mario? | ANGELA: Maria, you're explaining it all wrong—weren't you at the meeting last week? We've changed the color-coding system. New invoices are yellow now, not red. You're just going to confuse him! |
| *Response options* | *Response options (video: 15–30second per response)* |
| A. Make a joke to Min, explaining he didn't know about the changes<br>B. Not worry about it, just ignore the interruption<br>C. Learn the new changes*<br>D. Tell the boss that such criticism was inappropriate | A. MARIA (humorously): Well, you certainly seem to know your stuff, Angela. You're a virtual color wheel!<br>B. MARIA (angrily): Angela, I'm trying to explain the system to James, and you're just confusing him. Why don't you go work with your own intern and leave us to this?<br>C. MARIA (calmly): Actually, I did miss that meeting last week. James, let's go find Lana and make sure we're doing this right*<br>D. MARIA (apologetically): I'm really sorry, James; I think I messed up. Just let Angela tell you how to do the files; she seems to know more about them than I do |
| | *Response justifications (text).* Why did you choose this as the best response? Select all that apply<br>Maria directly expresses concern to James for her mistake. (Justifies C or D)<br>Maria shows Angela that her interruption is rude and distracting. (Justifies B)<br>Maria admits that she might have made an error. (Justifies C or D)<br>Maria attempts to find correct information for James. (Justifies C) |

Note: The best item is marked with an asterisk for both STEM and MEMA.

- Two item writers review interview transcripts and items, and write scripts for video-based items
- Revised by expert panel
  - <u>Stem</u> = 2 people interacting
  - <u>Response option</u> = head-and-shoulders of protagonist only

# Test Development: STEM-Youth

Jill is given an official warning for entering a restricted area. She was never informed that the area was restricted and will lose her job if she gets two more warnings, which she thinks is unfair.
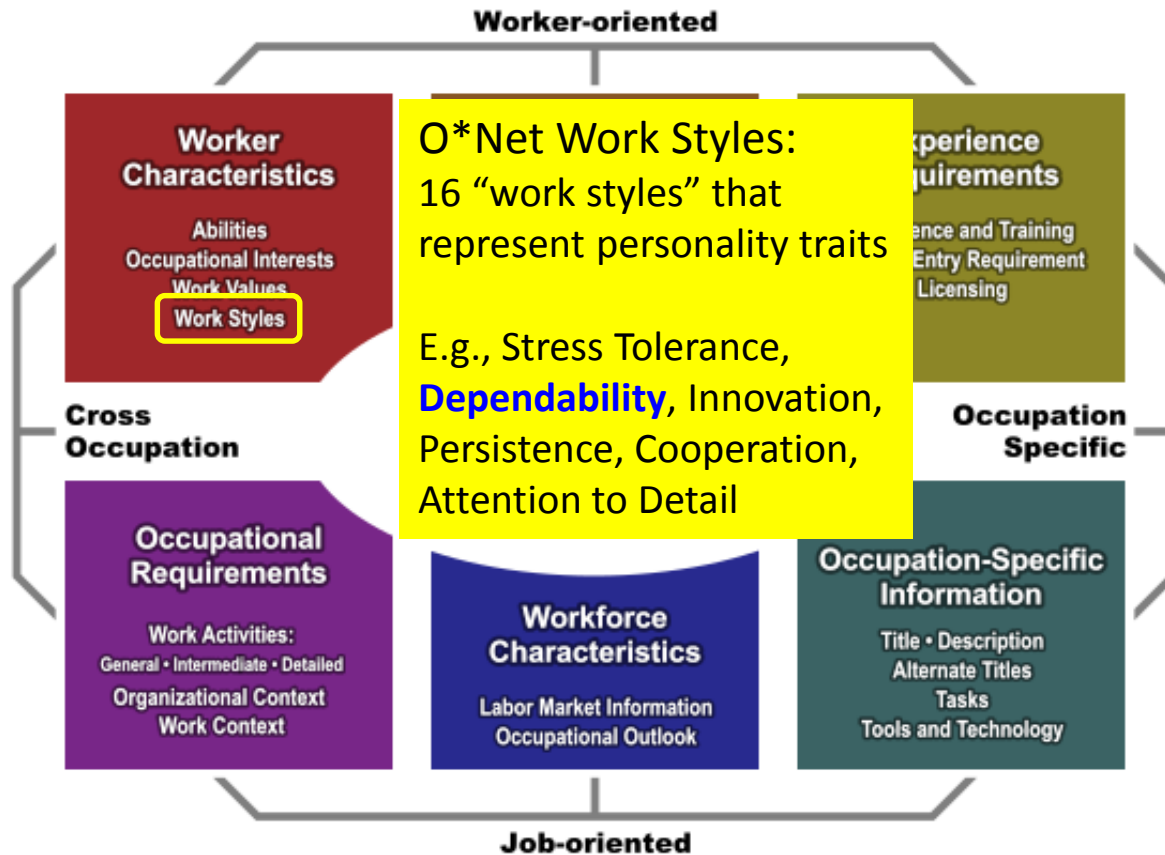
A. Think about the unfairness of the situation.
B. Accept the warning and be careful not to go in restricted areas from now on.
C. Explain that she didn't know it was restricted.
D. Take a few deep breaths and calm down about it.

You are given a warning by your teacher for having entered a restricted area. You were never informed that the area was restricted and will do detention if you get two more warnings, which you think is unfair.

A. Spend a lot of time complaining to your friends about this
B. Accept the warning and be careful not to go in the restricted area from now on
C. Explain that you didn't know the area was restricted
D. Take a few deep breaths and calm down about the situation

- 11 items
- Children <u>rate how likely each response</u> would be for them (*what would you do in this situation?*) ($\alpha$ = 71)
- Also administered in parent-report format (*what would you child do in this situation?*) ($\alpha$ = .69)

# SJTs of Personality



**O*Net Work Styles:**
16 "work styles" that represent personality traits

E.g., Stress Tolerance, **Dependability**, Innovation, Persistence, Cooperation, Attention to Detail

www.onetonline.org

Interviews:
"Tell me about a time when you or a colleague of yours has been <O*Net adjective phrase>.

What was the situation?

What happened?"

- reliable
- responsible
- dependable
- industrious/hard-working
- efficient
- punctual
- consistent
- well-prepared
- unreliable
- lazy
- frivolous, wasted time

# Example: Dependability SJT

Item choices represent different levels of the construct

"You are asked to deliver a critical report to your supervisor by close of business today. At your company, reports such as this one are supposed to be prepared according to specific procedures and guidelines. If you follow all the steps in the order suggested, however, you will not meet the deadline."

*How likely are you to respond with each of following actions?*

1. Ignore the procedures and guidelines and do only the most essential parts of the report to meet the deadline.
2. Keep working on the report, following all procedures and guidelines, and give your supervisor whatever you have completed by the end of the day.
3. Follow the procedures and guidelines and work into the night so you can deliver the completed report by start of business tomorrow.
4. Tell your supervisor that you cannot complete the report by close of business today.
5. Ignore the procedures and guidelines, but take as much time as you need to do the job.

# Part 3

# Test Development Decisions:
# Validity Evidence

# SJT Test Development Decisions

1. <u>Instruction type</u>: maximum vs typical performance
   - ***What would you do in this situation? (can "fake good")***
     - (Typical performance, "would do", behavioural tendency)
   - ***What is the most effective response? (can't "fake good")***
     - (Maximum-performance, "should do", knowledge)
2. <u>Response Type</u>:
   - Pick the best (**multiple-choice**)
   - **Rate** each response (and **how many scale points**?)
   - **Rank** the responses
   - Pick the best and worst
   - Allocate points
3. <u>Mode</u>:
   - <u>Video-based</u> versus <u>text</u>
4. <u>Scoring Decisions</u>:
   - ***MANY***
   - Expert judgment vs Norm Sample vs Theory-based
   - IRT vs CTT
   - Distance (multi-dimensional distance), proportional

Mario is showing Min, a new employee, how the system works. Mario's boss walks by and announces that Mario is wrong about several points, as changes have been made. Mario gets on well with his boss, although they do not normally have much to do with each other.
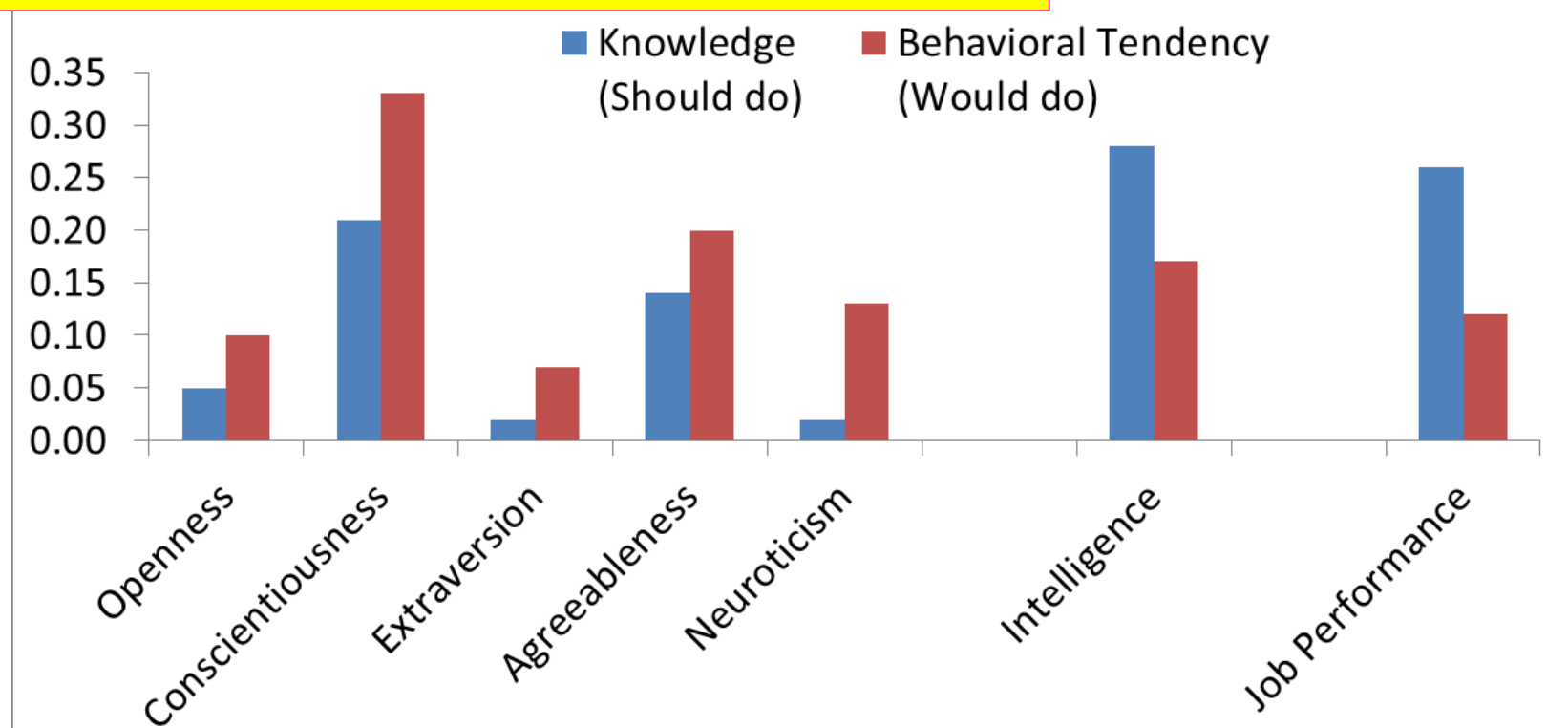
A. Make a joke to Min, explaining he didn't know about the changes
B. Not worry about it, just ignore the interruption
C. Learn the new changes
D. Tell the boss that such criticism was inappropriate.

# 1. Instruction Type: "Would do" vs "Should Do"

## Correlations with **personality**, **intelligence**, **job performance**

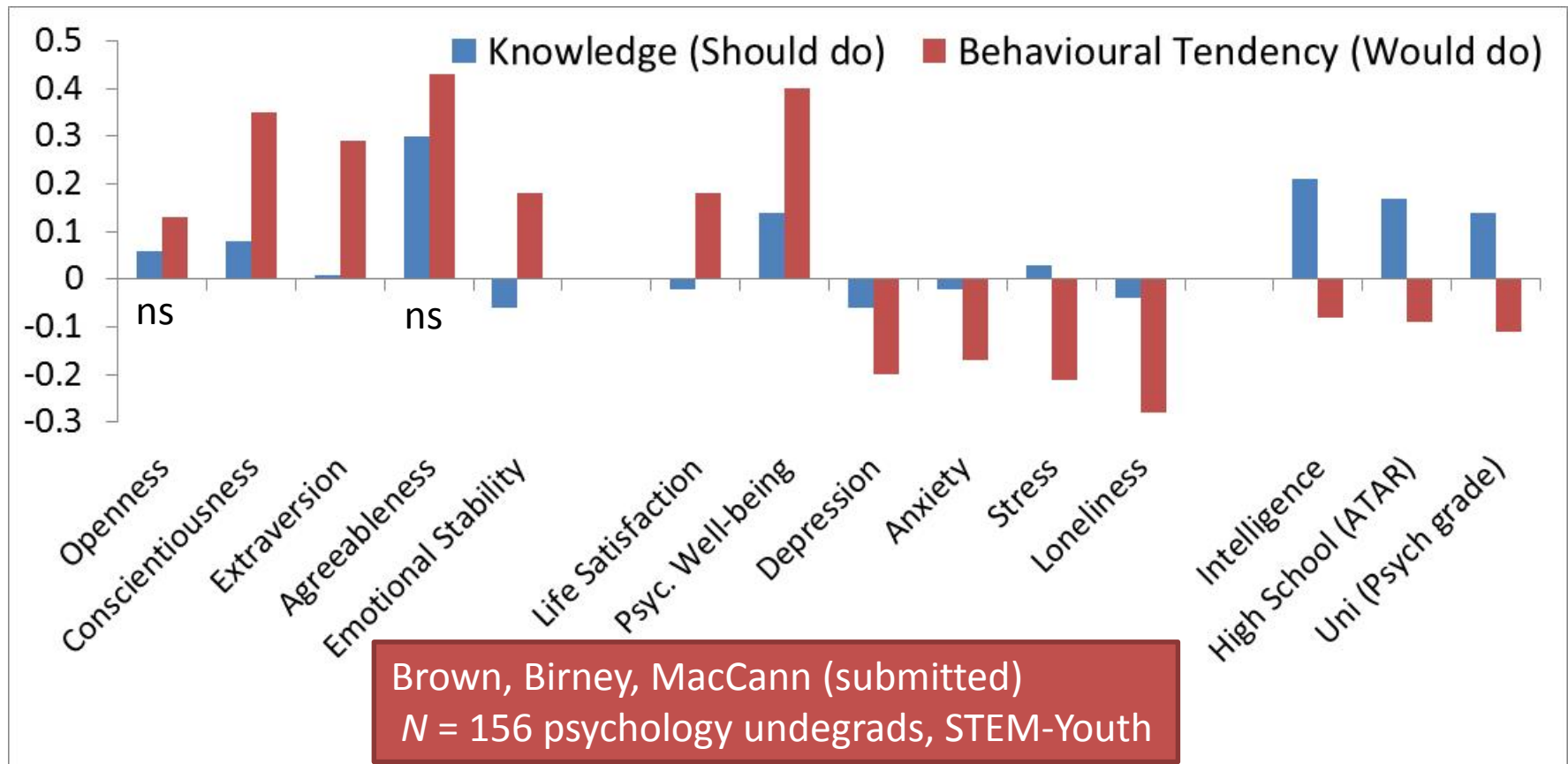**Behavioural Tendency** – higher with personality
**Knowledge** – higher with intelligence, job performance



$k$ = 3 to 8, $N$ = 290 to 763, (lowest for job performance)

McDaniel, M. A., Hartman, N. S., Whetzel, D. L., & Grubb, W. (2007). Situational judgment tests, response instructions, and validity: a meta-analysis. *Personnel Psychology*, *60*, 63-91.

# 1. Instruction Types: "Would do" vs "Should Do"

Alpha = .73 for "should do", .79 for "would do"



Brown, Birney, MacCann (submitted)
*N* = 156 psychology undegrads, STEM-Youth

**Personality:**
Higher corrs with behavioural tendency
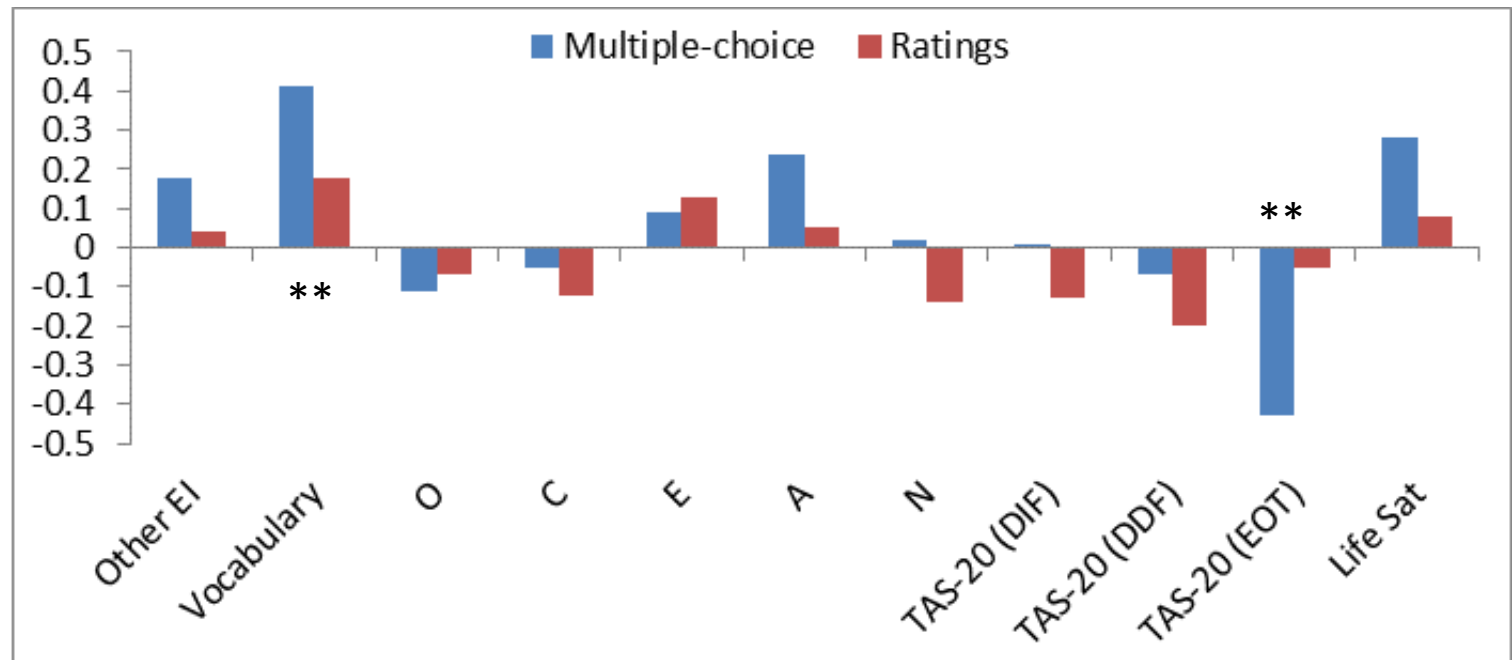
**Wellbeing (self-report):**
Higher corrs with behavioural tendency

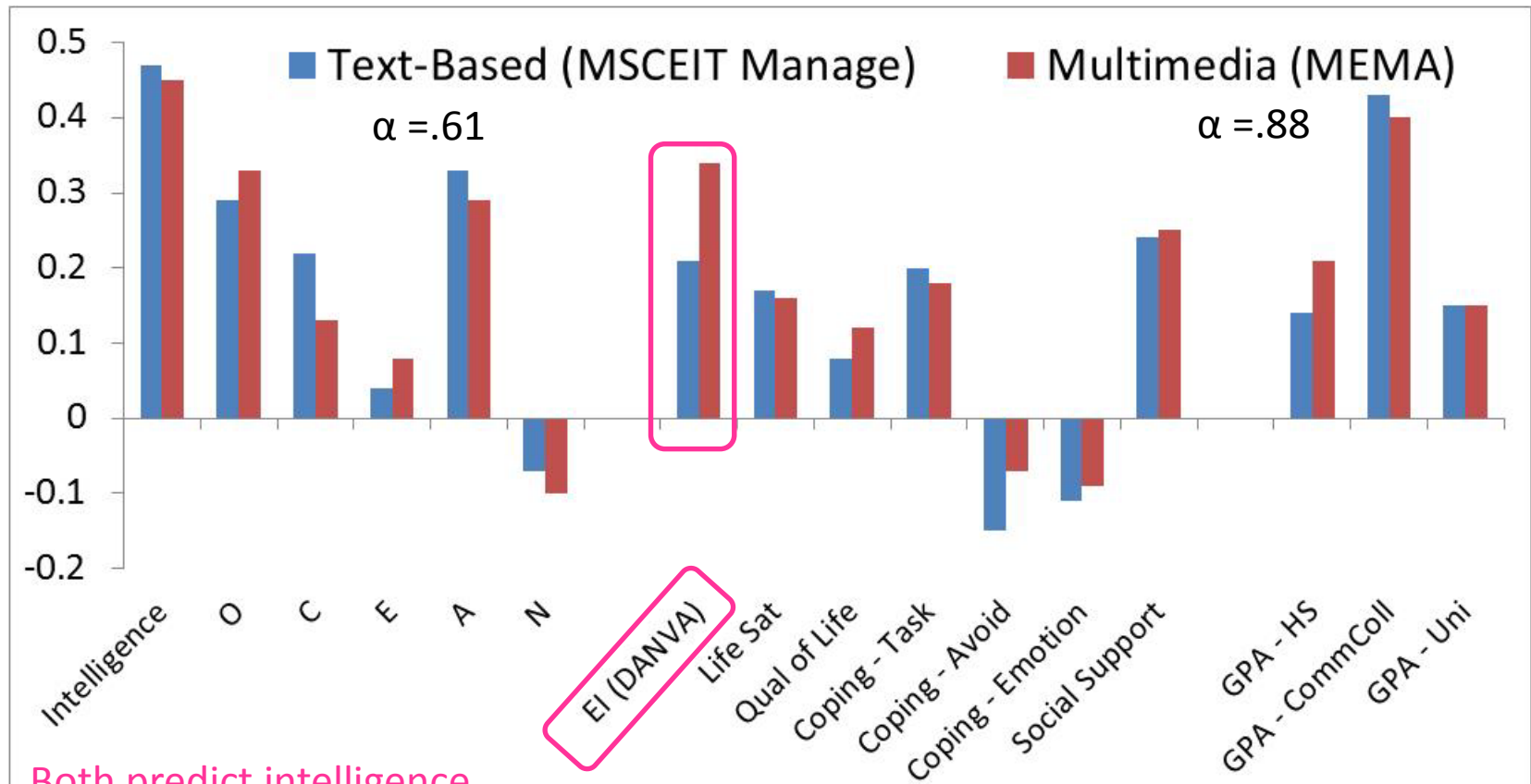**Intelligence + school achievement:**
Higher corrs with knowledge

# 2. Item Format

- Situational Test of Emotion Management (STEM)
  - 112 complete MCQ ($\alpha = .68$)
  - 91 rate each option ($\alpha = .92$)
  - ALL "Should do" (knowledge)

- Differences
  - Ratings-based **more reliable**
  - MCQ **more valid**



**MacCann, C., & Roberts, R. D. (2008). New paradigms for assessing emotional intelligence: theory and data.** *Emotion, 8(4),* **540.**

# 3. Multimedia vs Text



- Both predict intelligence Openness, agreeableness
- N = 427 USA students
- Both show *small* prediction of wellbeing/coping measures
  Only difference – multimedia more strongly related to emotion recognition
- Both predict academic achievement

MacCann, C., Lievens, F., Libbrecht, N., & Roberts, R. D. (2016). Differences between multimedia and text-based assessments of emotion management: An exploration with the multimedia emotion management assessment (MEMA). *Cognition and Emotion, 30,* 1317-1331.

# 4. Comparison of Scoring Techniques (Dependability SJT)

## One factor CFA model

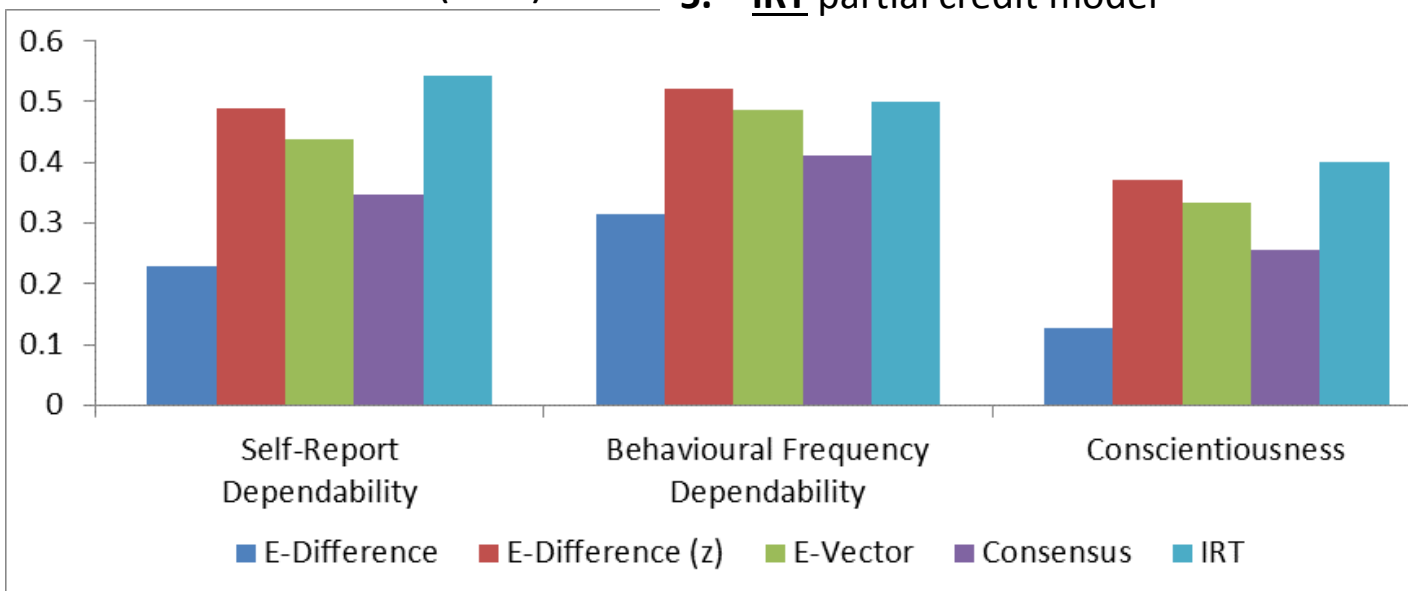| Scoring | CFI | RMSEA | Mean loading |
|---------|-----|-------|--------------|
| E-Diff | .85 | .05 | .42 |
| E-Diff (z) | **.97** | **.03** | **.44** |
| E-Vector | .93 | .04 | .42 |
| Consensus | .68 | .09 | .43 |
| IRT | .86 | .07 | .45 |

*N* = 546 Mturk workers (USA)

1. **Expert difference**: (E-Diff)
2. **Standardized expert difference:** (E-Diff z)
3. **Profile similarity** to expert ratings (E-vector)
4. **Consensus** proportions (e.g., if 45% of people rate option as "2", a "2" response = 0.45)
5. **IRT** partial credit model

**Correlations with Personality Traits**

N = 1,271 (3 studies)



Olaru, G., Zaromb, F., Burrus, J., Wilhelm, O., MacCann, C., Roberts, R. D. (2017). Situational Judgment Tests as a method for measuring personality: Development and validity evidence for a test of Dependability. In preparation.

# Summary/Conclusions

1. **Rate-the-response** more reliable than **MCQ**
   - More information (but also longer test time)
   - *But* ratings seem less valid?
     - Other research suggests ratings show *less adverse impact*, *higher r with personality (lower r with intelligence)* (Arthur et al., 2014, JAP)

2. Maximum vs typical perf. Instructions are **<u>very different</u>** for identical content
   - Typical = behavioural tendency (*fakeable, but often closer to what the instrument is trying to measure*)
   - Maximum = knowledge

3. Video vs text-based
   - At least in my study, they were very similar (***but Lievens & Sackett [2006] found otherwise***)
   - Is it worth the cost?

4. Scoring methods
   - The standard method for EI (consensus scoring) was the worst!
   - Standardized expert distance and profile scoring metrics were the most reliable
   - IRT-based, standardized distance, and profile scoring metrics were the most valid (construct-validity)

# THE END