

Rater Training for Committees

Promoting Readiness and Defensible Contributions

Presented by: **Walter Tavares**, PhD
Co-Chair, Best Practice in Evaluation and Assessment (BPEA)
MD Education, Temerty Faculty of Medicine
Wilson Centre for Health Professions Education Research
Department of Medicine
Department of Health and Society
University of Toronto
Email: walter.tavares@utoronto.ca

BPAS

You have heard about BPAS

Promoting Readiness and Defensible Contributions

Promoting Readiness and Defensible Contributions

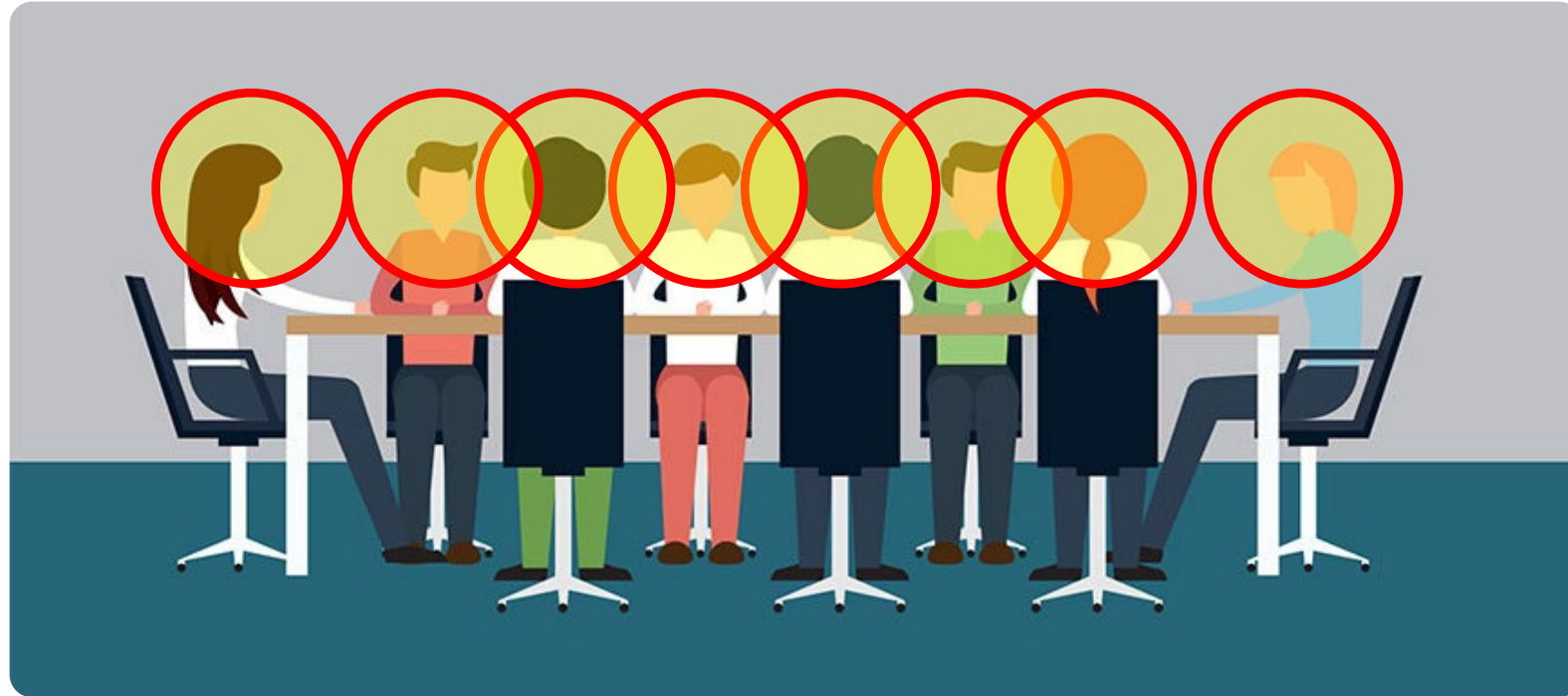


Promoting Readiness and Defensible Contributions

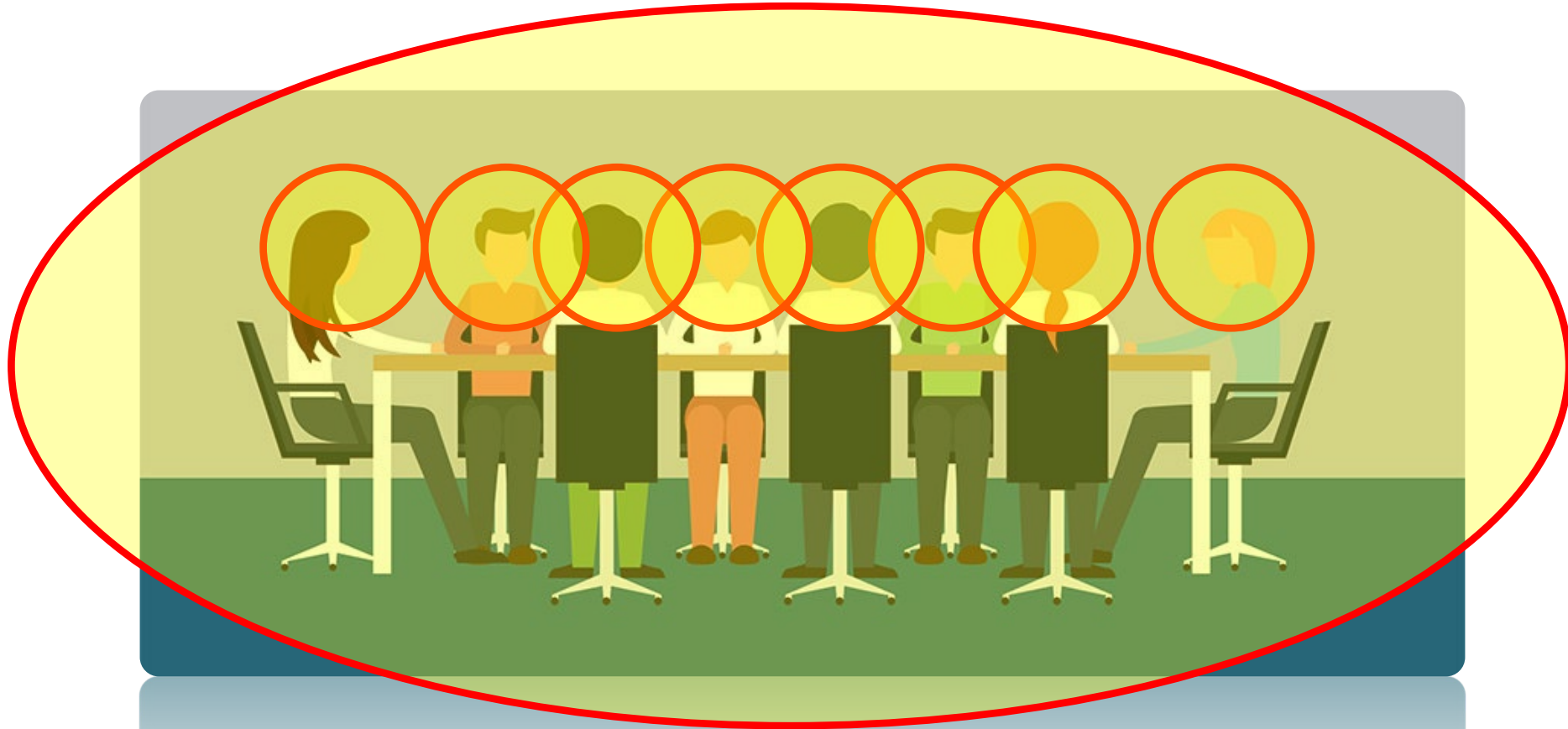


Human cognitive process.....

Promoting Readiness and Defensible Contributions



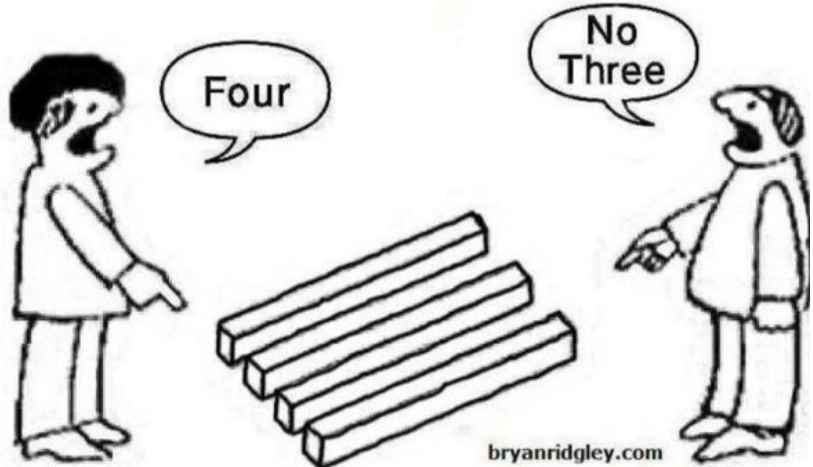
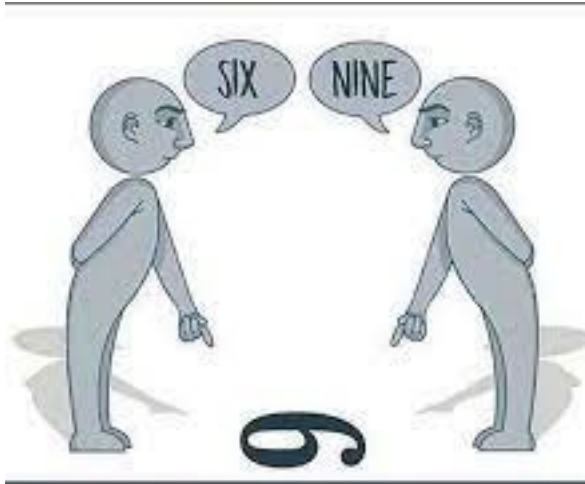
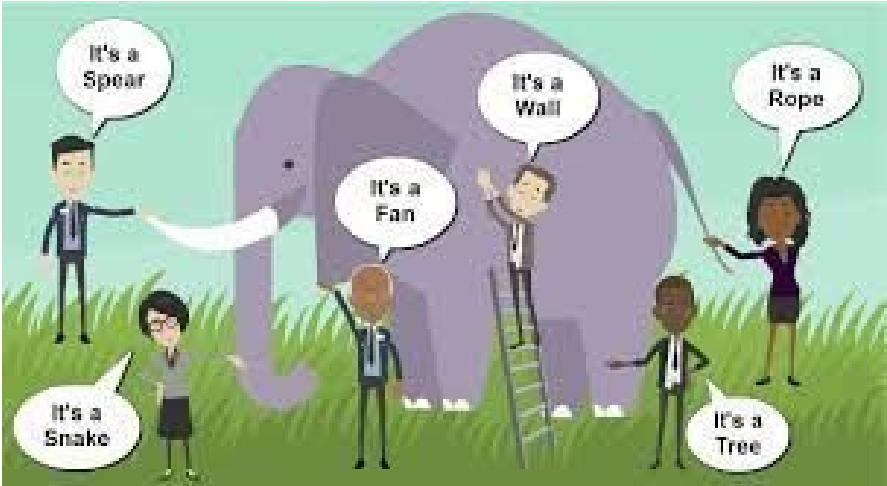
Human cognitive process.....in a social context



Promoting Readiness and Defensible Contributions



“Seeing the Same Thing Differently”



Advances in Health Sciences Education
<https://doi.org/10.1007/s10459-019-09939-9>

REFLECTIONS



The compatibility principle: on philosophies in the assessment of clinical competence

Walter Tavares^{1,2} · Ayelet Kuper^{1,3,4} · Kulamakan Kulasegaram^{1,5,6} · Cynthia Whitehead^{1,5}

Received: 22 March 2019 / Accepted: 25 October 2019
© Springer Nature B.V. 2019

Abstract

The array of different philosophical positions underlying contemporary views on competence, assessment strategies and justification have led to advances in assessment science. Challenges may arise when these philosophical positions are not considered in assessment design. These can include (a) a logical incompatibility leading to varied or difficult interpretations of assessment results, (b) an “anything goes” approach, and (c) uncertainty regarding when and in what context various philosophical positions are appropriate. We propose a compatibility principle that recognizes that different philosophical positions commit assessors/assessment researchers to particular ideas, assumptions and commitments, and applies a logic of philosophically-informed, assessment-based inquiry. Assessment is optimized when its underlying philosophical position produces congruent, aligned and coherent views on constructs, assessment strategies, justification and their interpretations. As a way forward we argue that (a) there can and should be variability in the philosophical positions used in assessment, and these should be clearly articulated to promote understanding of assumptions and make sense of justifications; (b) we focus on developing the merits, boundaries and relationships within and/or between philosophical positions in assessment; (c) we examine a core set of principles related to the role and relevance of philosophical positions; (d) we elaborate strategies and criteria to delineate compatible from incompatible; and (f) we articulate a need to broaden knowledge/competencies related to these issues. The broadened use of philosophical positions in assessment in the health professions affect the “state of play” and can undermine assessment programs. This may be overcome with attention to the alignment between underlying assumptions/commitments.



Common rater “errors”

Poor differentiation | Halo Error | Hawks and Doves | Stringency and Leniency | Contrast Effects | Central Tendency | etc.

“Rater Training”

Taking into consideration purpose, context, and process, optimizing rater cognitive behaviors and contributions in a social context that results in more defensible or trustworthy decisions about candidates.

Common “rater training”

Rater Error Training | Performance Dimension Training | Frame of Reference Training | Behavior Observation Training

1. Raters use **different** dimensions, schemas
2. **Idiosyncrasy** is substantial and unrelated to expertise
3. More **differentiated performance schemas** in experienced raters
4. Use **personal performance theories** and constructs
5. Do **not map neatly** onto rating tools



ASSESSOR

1. Domains idiosyncratically take on **variable degrees of importance**
2. Construct **irrelevant** features present
3. Standardization may **constrain** raters
4. Strong **subjective** influences





1. As trainable – they vary because they do not apply criteria correctly, used varied frames of reference
2. As fallible – limitation in human cognition
3. As meaningfully idiosyncratic – use context and contextual sensitivity



Recommendations

1. Define the “**CONSTRUCT**”

Use faculty/assessors, identify relevant dimensions, align or share with candidate expectations, ensure construct is shared

Recommendations

2. Share a “**PHILOSOPHICAL POSITION**”

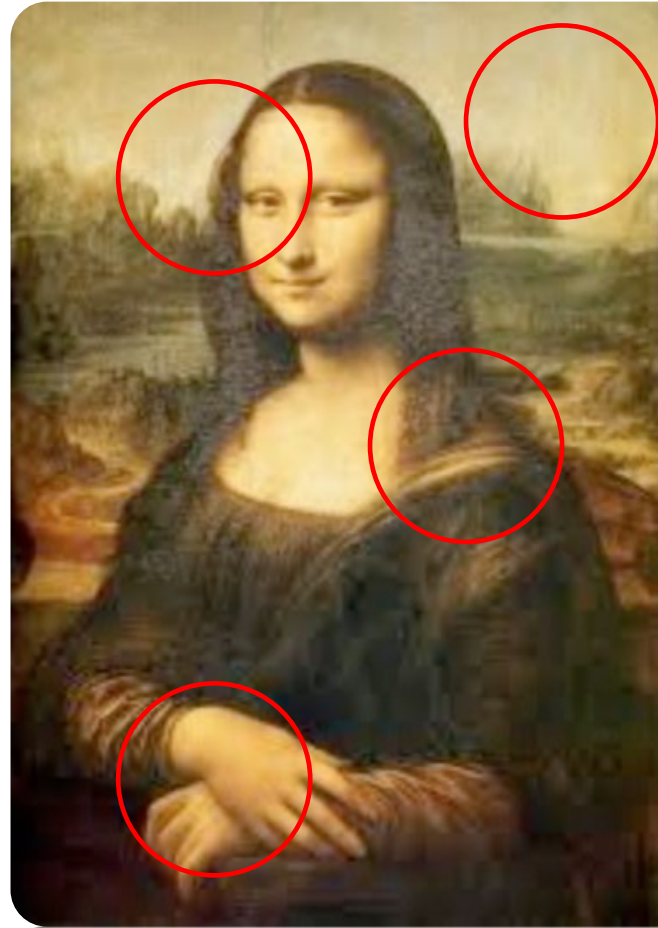
Views about assessment, role of “bias”, role or accuracy, contributions, objectivity and subjectivity

Recommendations

3. Discuss the “**RULES OF THE GAME**”

Assessment purpose, procedure, their role, your expectations of them, tools, overview of entire assessment process, scoring and standard setting rules etc.





Recommendations

4. Provide “ASSESSMENT GOALS”

Outline good assessment practice (e.g., differentiate between candidates and dimensions, use range of scale, weighing of sources)

Recommendations

5. Discuss "SOCIAL INFLUENCES AND BARRIERS"

Eliminate / minimize problematic issues (e.g., gatekeeper), embrace others (e.g., social responsibility), discuss the diffusing of responsibility without giving up accountability

Recommendations

6. Provide opportunities for "REFLEXIVITY"

Discuss and surface views, positionality, meaning of beliefs and values
in this context

Recommendations

7. WORK ON “**INTERPRETIVE SCHEMES**” for **DIMENSION** and **SAMPLES**

Understand what and why

Recommendations

8. Work on “**INTERPRETIVE SCHEMES**” IN “**FRAME OF REFERENCE**” TRAINING

“Shared conceptualizations of what constitutes quality – can be specific performance expectations or generic.” Rather than impose a FOR, derive it from faculty!

Recommendations

9. Prioritize **RATERS, NOT TOOLS**, as the “measurement” device.

Overly prescriptive tools and raters may not align. Makes translation challenging, may promote “objectivity” but constrain judgment or processing. Promote dimension relevant global ratings.

Recommendations

10. Re-conceptualize from an “intervention” to
a “**PROGRAM**”

Less about cause and effect or accuracy – more
about promoting defensibility and quality. Long
term, ongoing, integrated with metrics / feedback.

Recommendations

BONUS: ADOPT A **FACULTY / RATER FIRST** APPROACH

Engage faculty / raters in development of dimensions, expectations, frame of referencing, tools etc.

Recommendations

BONUS: USE RATERS WITH **RELEVANT EXPERTISE**, AND
WHO **ASSESS WELL**.

Examine the performance and idiosyncrasies

Yeates P1,2, O'Neill P2,3, Mann K4, Eva K5. Seeing the same thing differently: Mechanisms that contribute to assessor differences in directly-observed performance assessments, *Advances in health Sciences Education*, 18 (3): 325-41

Rater Training for Committees

Promoting Readiness and Defensible Contributions

Presented by: **Walter Tavares**, PhD

Assistant Professor | Scientist

Health Professions and Practice, Department of Health and Society

Wilson Centre for Health Professions Education Research

MD Education, Temerty Faculty of Medicine

Institute of Health Policy Management and Evaluation

University of Toronto

Office: 416-287-7478

Email: walter.tavares@utoronto.ca



UNIVERSITY OF
TORONTO