Twenty-eight SJT scoring methods: influence on internal consistency reliability, adverse impact and correlation with personality

Wendy E. de Leng[1], Karen M. Stegers-Jager[1], Adrian Husbands[2], Johnathan S. Dowell[3], Marise Ph. Born[4] & Axel P.N. Themmen[1]

[1]Institute of Medical Education Research Rotterdam, Erasmus MC, the Netherlands

[2]Medical School, University of Buckingham, UK

[3]School of Medicine, University of Dundee, UK

[4]Department of Psychology, Erasmus University Rotterdam, the Netherlands

Scenario

Michael questions Sarah, a fellow medical student about extreme and provocative comments about individuals' sexual preferences on her Facebook page. Sarah argues she should be free to express her personal views. She also insists that her personal views have no bearing on her performance as a medical student or patient care.

Response options

1. Advise Sarah to remove all controversial comments from her Facebook page.
2. Alert Facebook that Sarah's page contains potentially inappropriate content as they could remove it.
3. Ask Sarah to ensure her privacy settings are restricted so her page is inaccessible to patients or the general public.
4. Inform a member of staff about Sarah's Facebook comments.
5. Withhold advice to Sarah as her views do not affect patient care or performance as a medical student.

SJTs have no objectively correct answers

What is the capital of France?

a. Paris	√	→	1
b. Versailles	✗	→	0
c. Bordeaux	✗	→	0
d. Disney Land	✗	→	0

Erasmus MC

Subject matter experts – Rational scoring key

Response instructions

Pick one

Please rank the following toppings on a scale of 1 to 5, with 1 being your favorite.

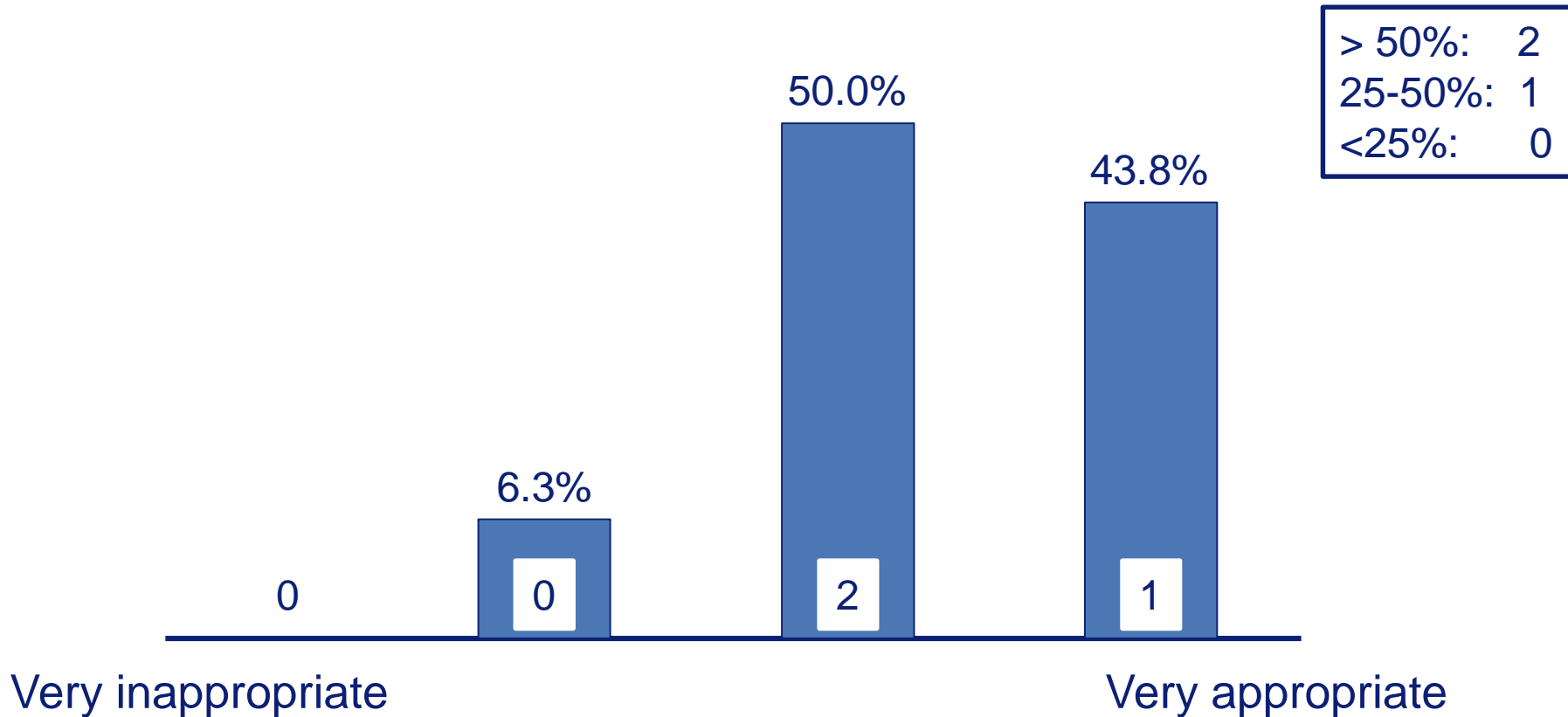| 3 | Pepperoni |
5	Anchovies
2	Mushrooms
4	Olives
1	Sausage

Ranking

Rating

Erasmus MC

The present study

Integrity-based SJT:

- 10 scenarios

- 5 response options

- 4-point rating scale (*very inappropriate – very appropriate*)

Sample:

- 931 medical school applicants – 2 cohorts

- 16 Subject matter experts

Original scoring method – endorsement rate

> 50%:   2
25-50%: 1
<25%:   0

50.0%

43.8%

6.3%

0

0

2

1

Very inappropriate

Very appropriate

Advise Sarah to remove all controversial comments from her Facebook page.

Erasmus MC

Toward an Understanding of Situational Judgment Item Validity and Group Differences

Michael A. McDaniel
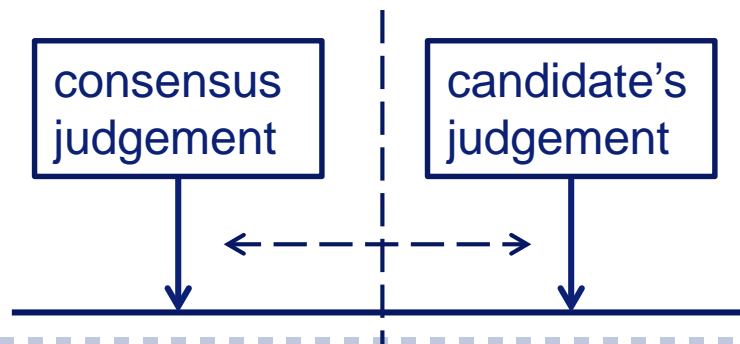Virginia Commonwealth University

Joseph Psotka and Peter J. Legree
U.S. Army Research Institute for the Behavioral and Social Sciences
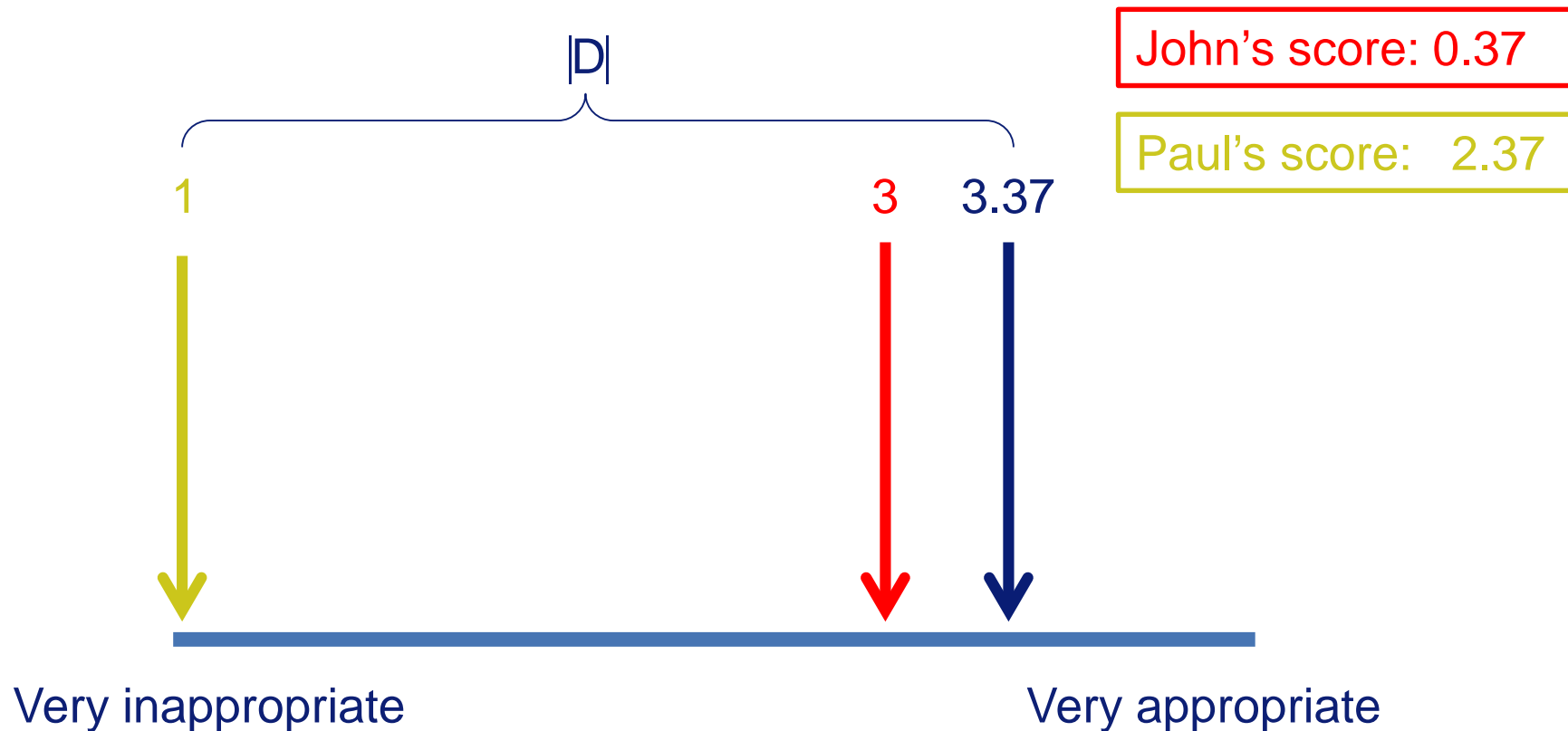
Amy Powell Yost
Capital One

Jeff A. Weekley
Kenexa

Three different scoring methods:

- Raw consensus

- Standardized consensus (within-person z transformation)
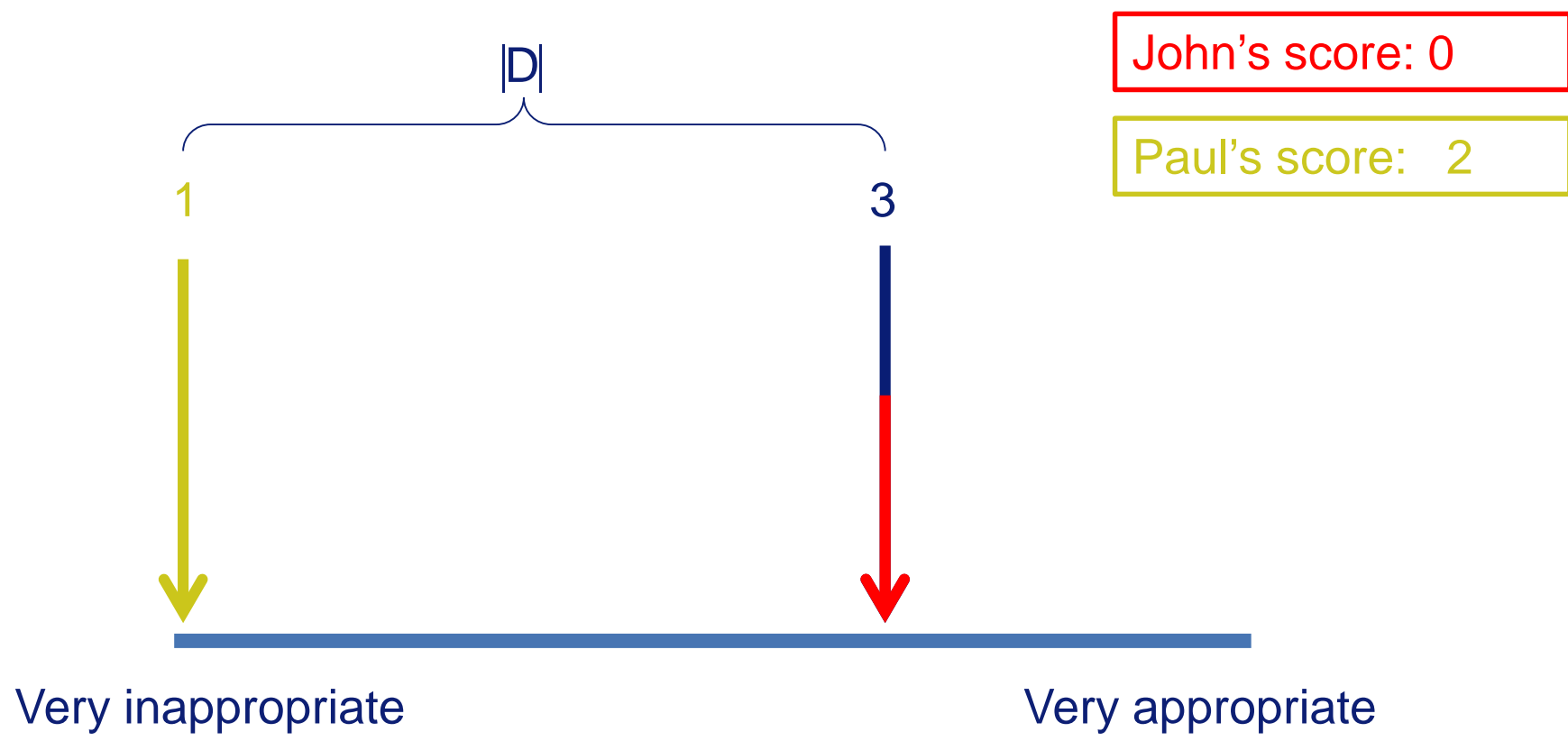
- Dichotomous consensus

Scoring Methods – Absolute distance from the SME mean

|D|

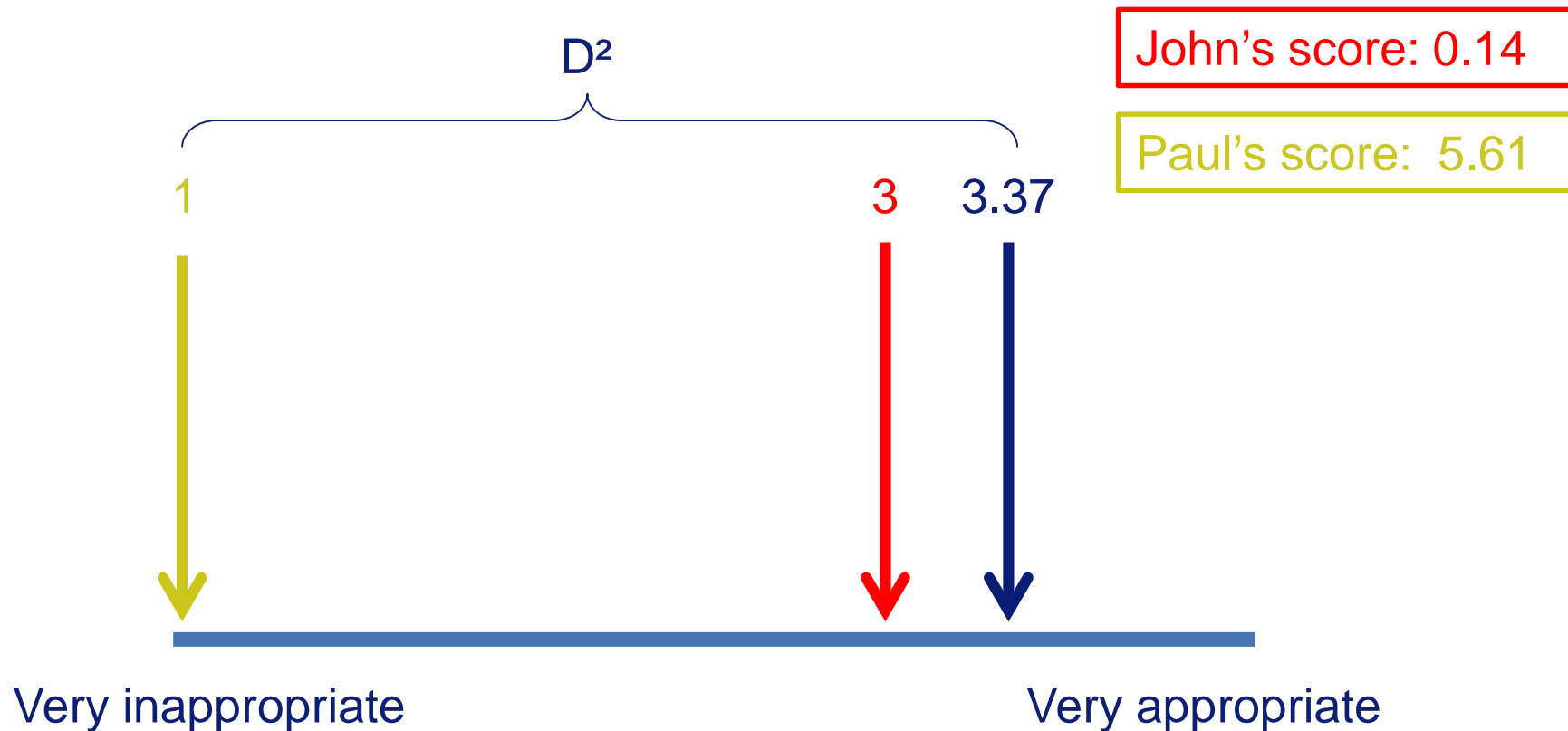John's score: 0.37

Paul's score: 2.37

1   3   3.37

Very inappropriate   Very appropriate

Advise Sarah to remove all controversial comments from her Facebook page.

Erasmus MC

Scoring Methods – Absolute distance from the SME **median/mode**

|D|

John's score: 0

Paul's score: 2
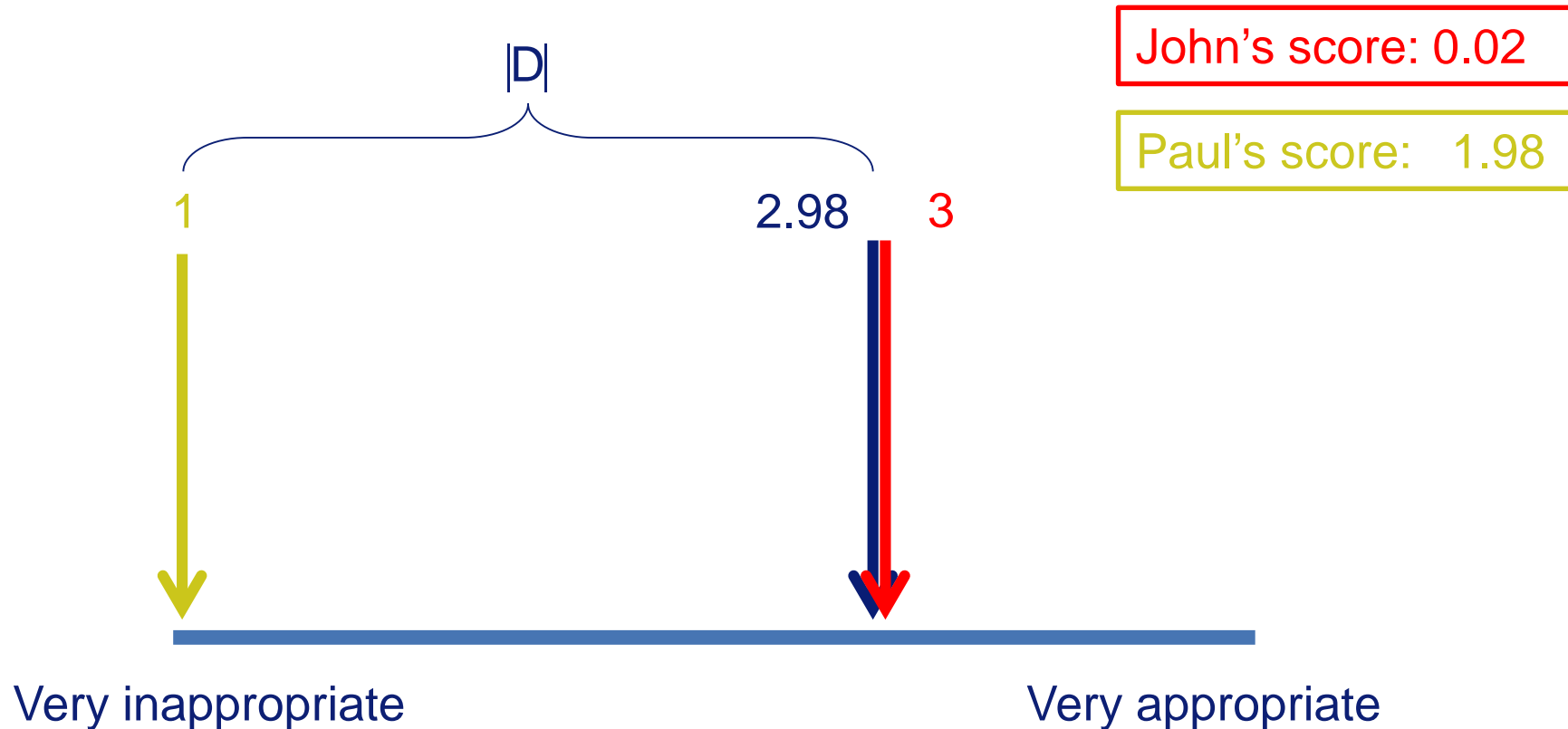
1

3

Very inappropriate

Very appropriate

Advise Sarah to remove all controversial comments from her Facebook page.

Erasmus MC

Scoring Methods – <u>Squared</u> distance from the SME mean

D²

John's score: 0.14

Paul's score: 5.61

1   3   3.37

Very inappropriate   Very appropriate

Advise Sarah to remove all controversial comments from her Facebook page.

Erasmus MC

Scoring Methods – Absolute distance from the <u>group</u> mean

|D|

John's score: 0.02

Paul's score: 1.98

1   2.98   3

Very inappropriate   Very appropriate

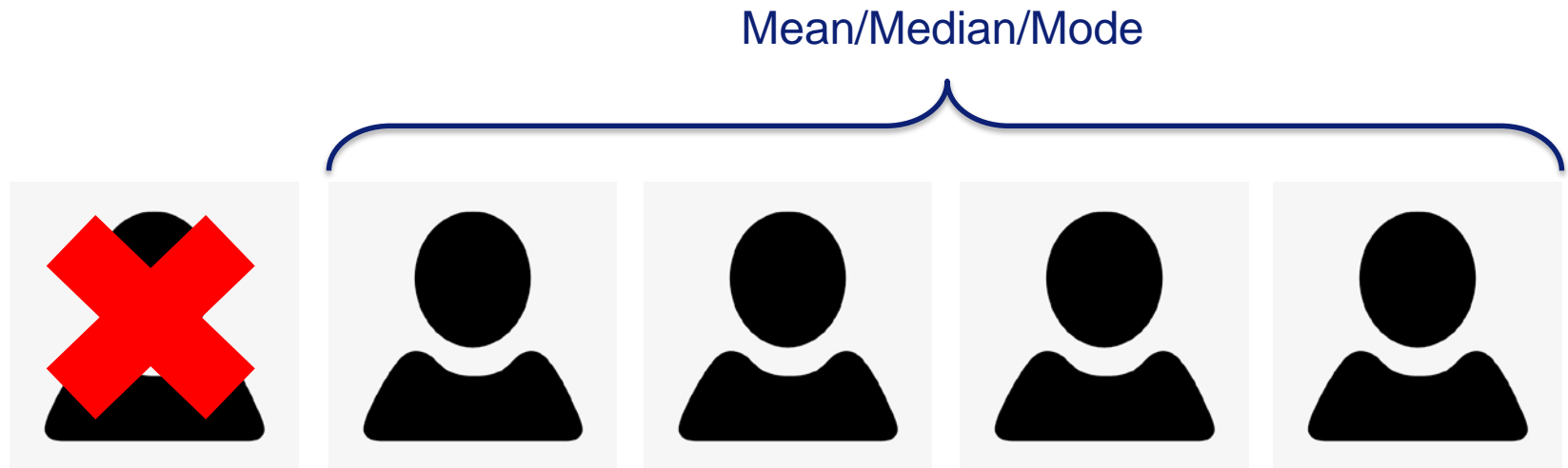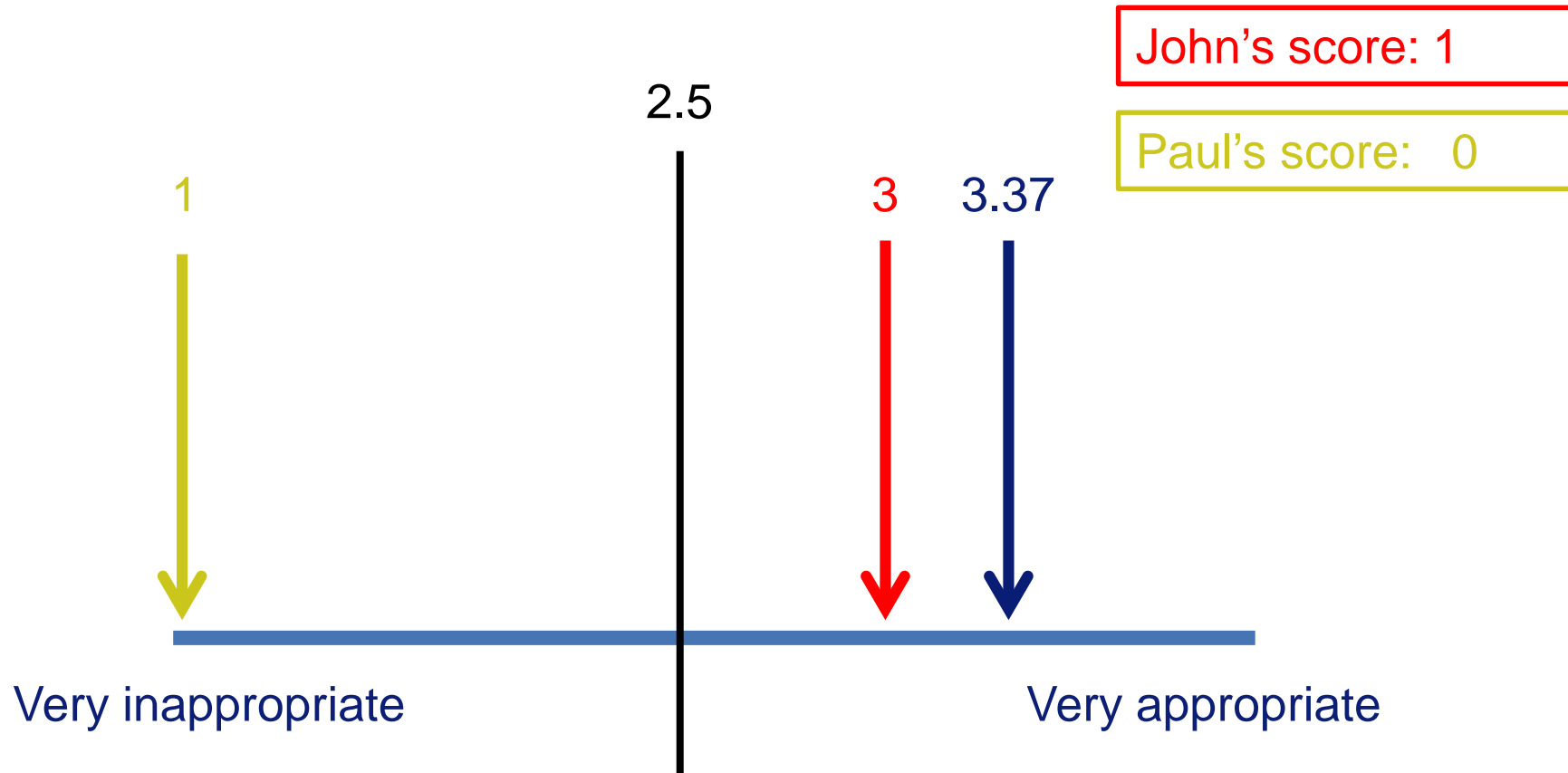Advise Sarah to remove all controversial comments from her Facebook page.

Erasmus MC

Scoring Methods – Absolute distance from the <u>group</u> mean

Mean/Median/Mode

Leave-one-out method

Erasmus MC

Scoring Methods – Dichotomous

2.5

1

3   3.37

John's score: 1

Paul's score: 0

Very inappropriate

Very appropriate

Advise Sarah to remove all controversial comments from her Facebook page.

Erasmus MC

Scoring Methods – Dichotomous with negative correction
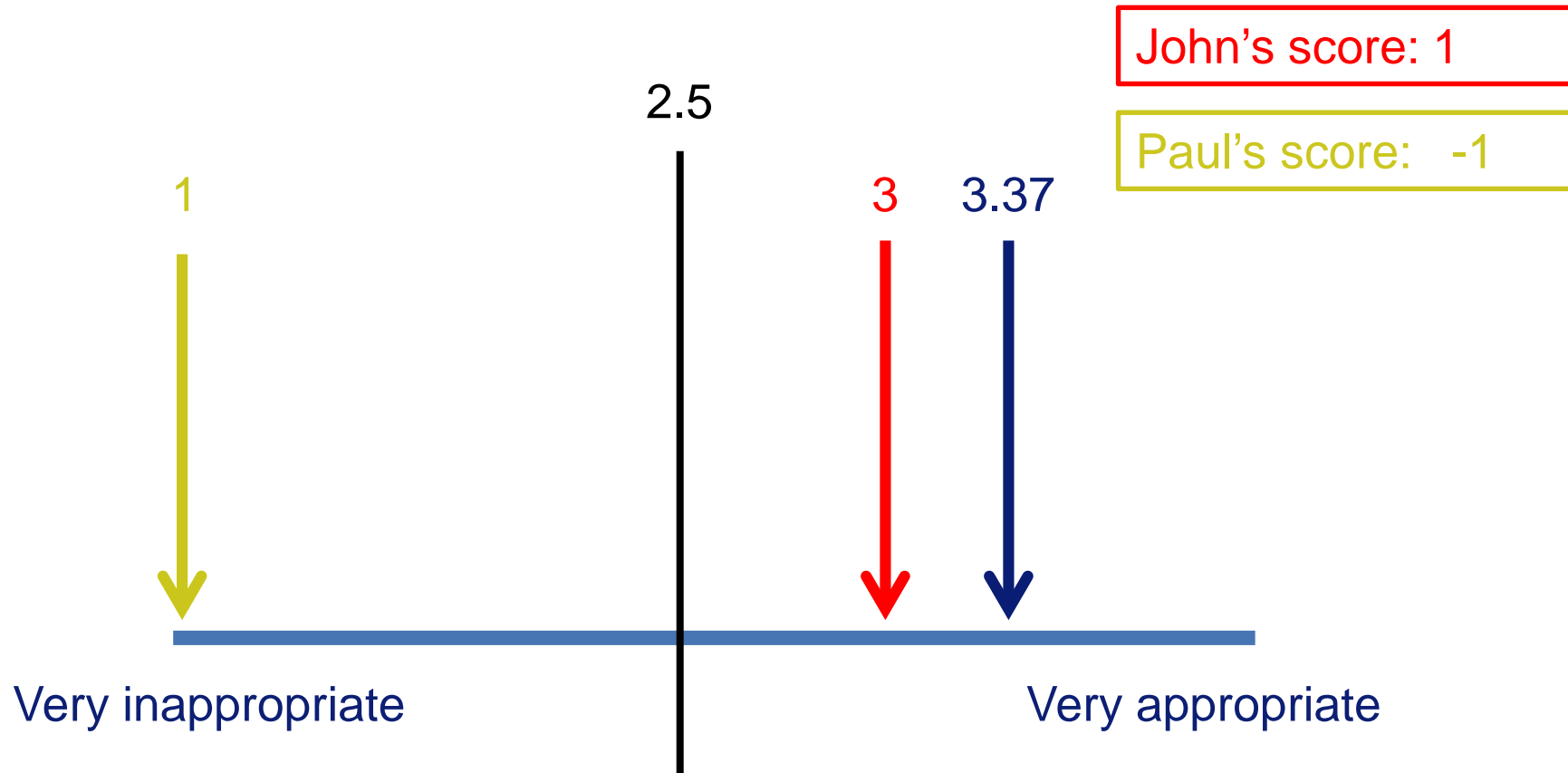
John's score: 1

Paul's score: -1

2.5

1

3   3.37

Very inappropriate

Very appropriate

Advise Sarah to remove all controversial comments from her Facebook page.

Erasmus MC

28 scoring methods – 4 aspects

- Control for systematic error:

1) raw consensus

2) standardized consensus

3) dichotomized consensus

- Reference group:

1) subject matter experts

2) candidates themselves

- Distance:

1) absolute

2) squared

- Consensus judgement:

1) mean

2) median

3) mode

Erasmus MC

28 scoring methods were compared on:

- Internal consistency reliability (coefficient alpha)

- Ethnic subgroup difference Dutch vs. non-Western minorities
(Cohen's *d*)

- Correlation with Conscientiousness, Agreeableness and Neuroticism
(Pearson's correlation)

Results & Discussion: Internal consistency reliability

- Coefficient alpha: .33 - .73

    ↑ raw consensus   ↓ dichotomous consensus

- Interpretation coefficient alpha: unacceptable - acceptable

- Score variance

- Critique on coefficient alpha

    "I doubt whether coefficient alpha is the best way of judging the reliability of the instrument to which it is applied."

    Cronbach & Shavelson, 2004

Erasmus MC

Results & Discussion: Ethnic subgroup difference

- Effect size ethnic subgroup difference: 0.48 – 0.66

    ↓ dichotomous consensus

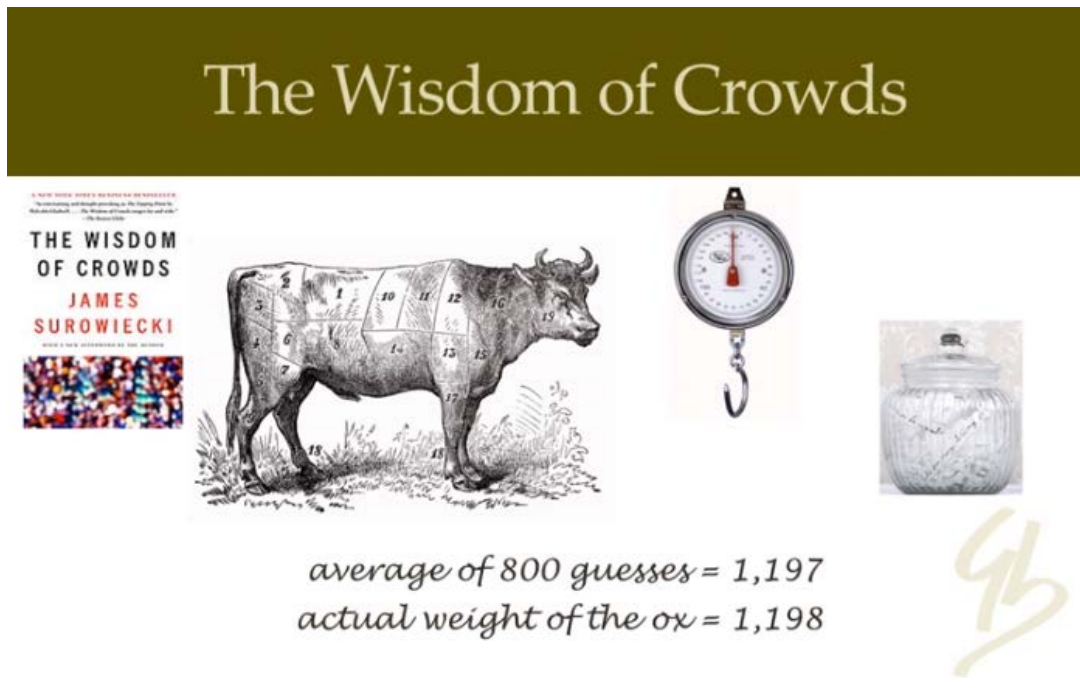
- raw consensus = standardized consensus

- All scoring methods led to a significant difference in favour of Dutch candidates

Results & Discussion: correlation with personality

- Correlation with personality: all below .15

- Non-cognitive traits vs. non-cognitive behaviours

Erasmus MC

Results & Discussion: Scoring methods in general

- Most influential aspect: control for systematic error (raw, standardized, dichotomous consensus)

- Expert reference group = candidate reference group

The Wisdom of Crowds

THE WISDOM OF CROWDS
JAMES SUROWIECKI

average of 800 guesses = 1,197
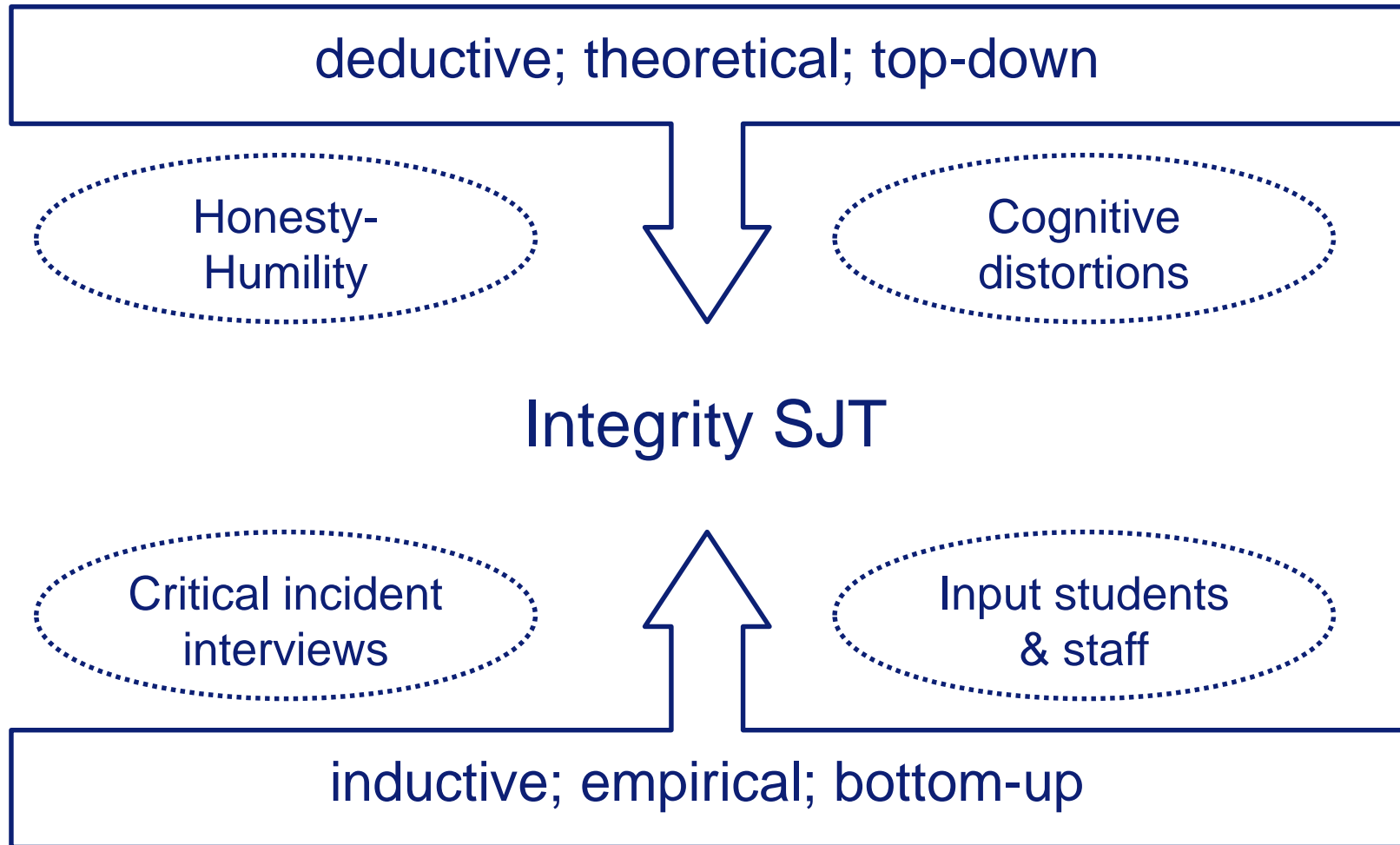*actual weight of the ox = 1,198*

Erasmus MC

Critique on the cross-cultural applicability of the integrity-based SJT

Development of a new Integrity SJT

Inductive and deductive development

deductive; theoretical; top-down

Honesty-Humility

Cognitive distortions

Integrity SJT

Critical incident interviews

Input students & staff

inductive; empirical; bottom-up

Erasmus MC

What to do vs. What <u>not</u> to do

Two different skills:

- Correctly recognizing what one should do in a challenging situation

- Correctly recognizing what one should <u>not</u> do in a challenging situation

What <u>not</u> to do: ↑ convergent validity

More consensus on what is inappropriate than appropriate in a challenging situation

Take-home messages

Scoring methods matter

Group of candidates itself as a reference group

Combine inductive and deductive development

SJT score on what one should <u>not</u> do

Contact: w.deleng@erasmusmc.nl