Criteria, Practices and Ethical Pitfalls when Selecting Assessment Instruments for your Assessment Centre

Gert Roodt (University of Johannesburg, South Africa)
Francois de Kock (University of Cape Town, South Africa)
and **Sandra Schlebusch** (LEMASA, South Africa)

Speakers' Corner
ACSG 14 – 15 March 2013
ICACM, Spier, Stellenbosch

UNIVERSITY OF JOHANNESBURG

University of Cape Town · iYunivesithi yaseKapa · Universiteit van Kaapstad

LEMASA
Unlocking potential.

Discussion Agenda

- Background
- Claims/practices
- Criteria to assess AC simulations
- Ethical pitfalls

Perspective Taken

- User of Assessment Centre instruments
- Decision to be taken:

What instruments to include in Assessment Centre to deliver an Assessment Centre that adhere to all requirements

Background

Ethics broadly refers to being good to self and to others (Rossouw & Van Vuuren, 2010)

Ethics in Assessment Centres is about an assessment procedure and process that is good for all stakeholders (Muleya et al., 2013)

Background (continued)

- Users bombarded with choices
- Usually under pressure to decide on instrument(s)
- Instrument distributors' claims speak to identified need
- High-stakes decisions to be made based on results delivered by instruments
- Appears as though all instruments will deliver required results

Claims made by Distributors

- …"the centre / instrument is valid and reliable for this population"
- …"the AC showed a correlation of .41"
- …"the centre / instrument determines potential"
- …"any psychologist can administer, score and interpret results"
- …"these AC exercises can be tailored to need; their flexibility make them ideal to…"

Claims made by Distributors

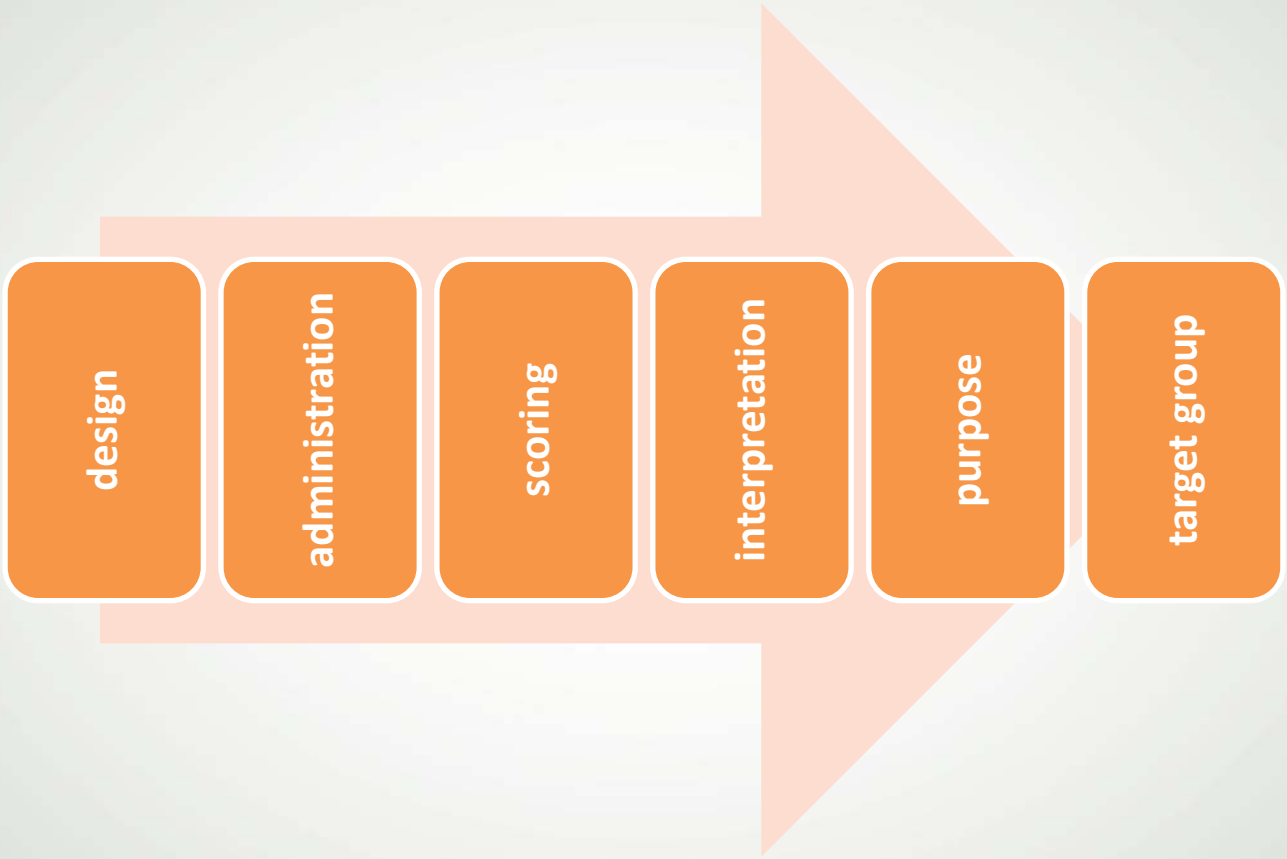
- "The X is the most widely used measure in the UK."
- "Many of our clients have a long history of purchasing our products."
- "The nature of ACs and the ways in which they can be used is very different to the highly standardised way in which psychometrics are used."
- "Whilst psychometrics lend themselves to normative comparisons, the same is not true for ACs."
- "We know that the format of our ACs work effectively as they have been used successfully for many years"

Criteria to assess ACs

We know the criteria, but when are they sufficiently addressed?

Typical criteria of good tests

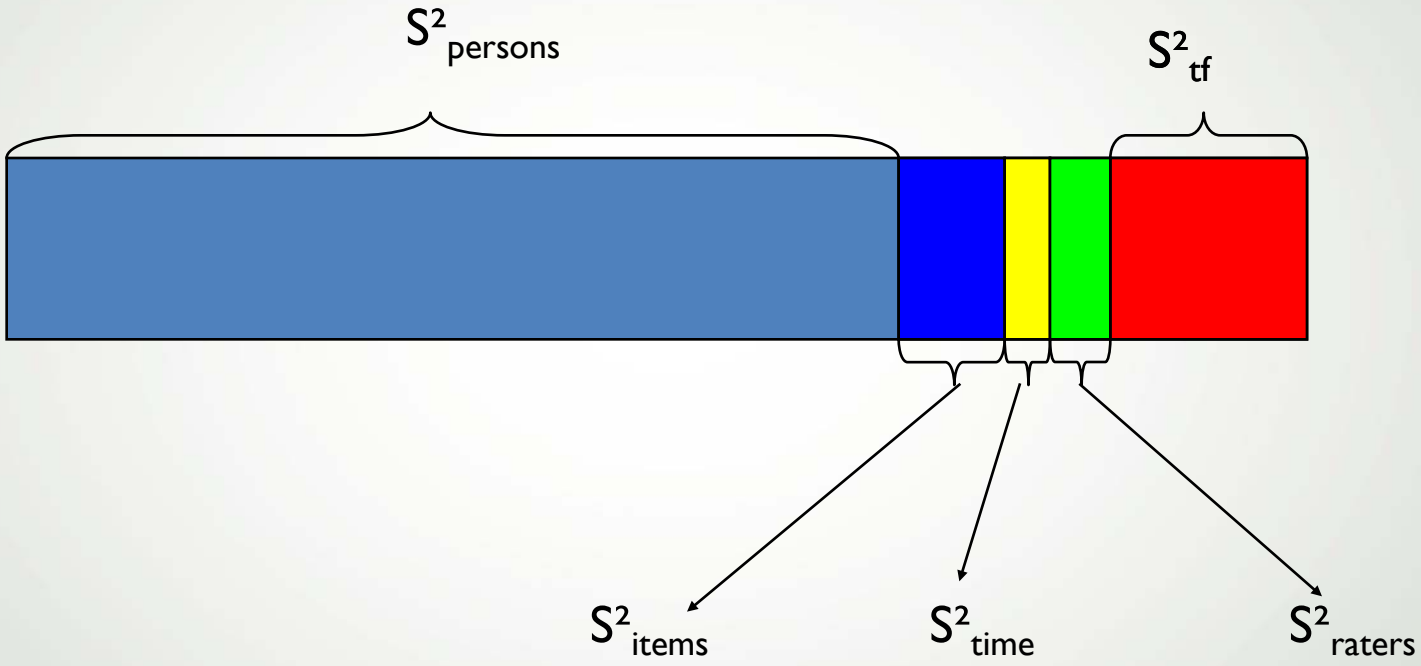
Considerations spectrum



Reliability

- AC manual and administrator's manual?
- Does AC manual contain very clear instructions
  - Assessment conditions
  - Materials
  - Instructions to candidates
  - Time limits
  - Use of roleplayers and prompts
  - Administration procedures and scoring
- Assessors are key in score production
    - Evidence of assessor consistency?
    - Inter-rater reliability vs Agreement?
    - Assessor training, calibration and certification
    - Evidence of generalisability across assessors?

Reliability (generalisability…)



Reliability (generalisability…)

From Murphy (2011) 'Validating ACs'	
Dimension	**Question**
Candidates	Overall differences in candidate performance
Dimensions	Does the pool of candidates show more strength in some competency areas than others?
Assessors	Are assessors calibrated?
C x D	Do candidates show different strengths and weaknesses?
C x A	Do assessors agree about candidates?
A X D	Do assessors agree about dimensions (competencies)
C x D x A	Do assessors agree in their evaluations of the patterns of strength and weakness of different candidates?

Validity evidence

- Valid…for what?
- Evidence to support use of specific simulations to aid decisions in terms of specific criteria - alignment
- Inordinate focus on construct validity issues to detriment of other pieces of validity puzzle
- Because ACs have strong work-sample flavour, content validation should also be important
  - Evidence of content validation in development
  - Use of techniques to create more ‘objectivity’?
  - Evidence that dimensions sufficiently activated?

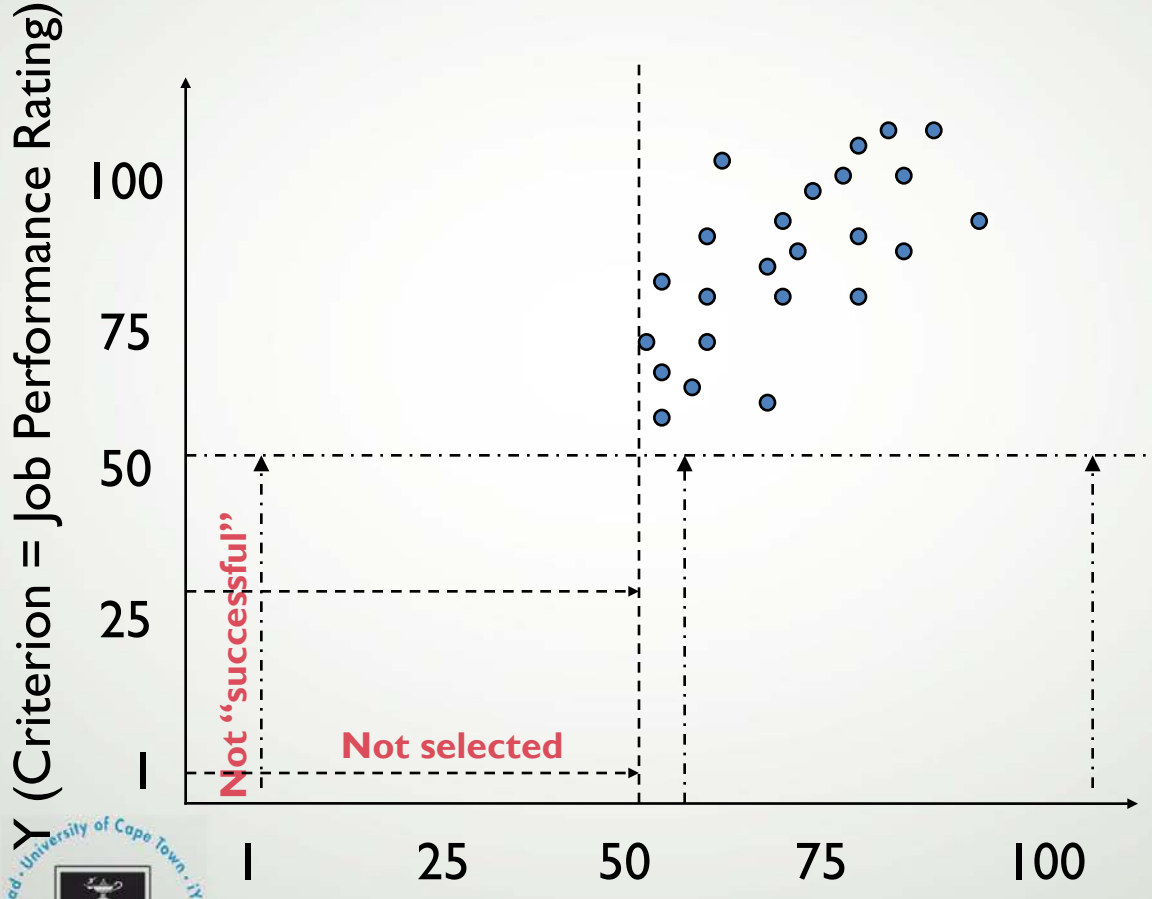
Criterion-related validity

- When is *r* or *R* 'big enough'?
  - Range from $-1 < r < 1$
  - Gaugler et al. (1987): (corrected) mean $r = .37$
  - Hermelin, Lievens, Robertson (2007): $.24 < \rho < .32$ ($\rho = .28$)
  - 'Validity ceiling'

- Statistical artefacts taken into account in manual?
  - Restriction of range in scores
  - Unreliability in measurement
  - Adequate sample size for statistical power

Statistical artefacts

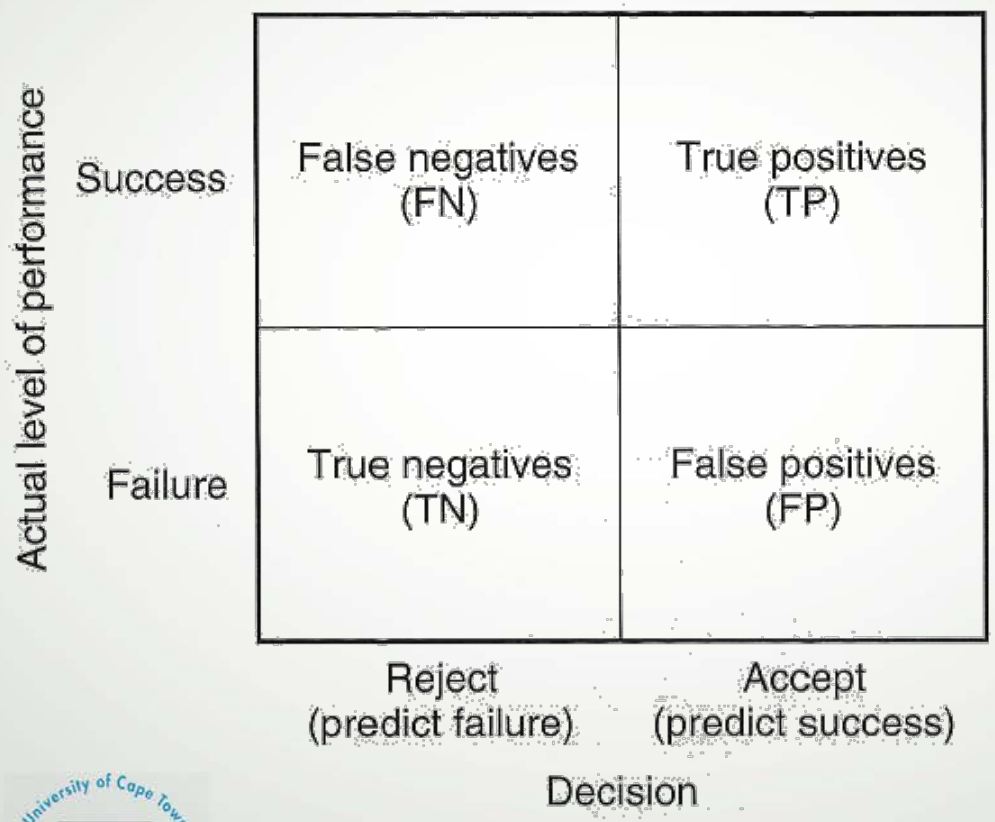
Range restriction – both predictor and criterion: r_{xy}= .09



Validity (…continued)

- Incremental validity evidence over *g*?
- Shared variance between exercises may make them superfluous
- Evidence that they 'work' for specific purposes development, e.g. DAC interventions?

Decision errors considered?



(Source: Murphy & Davidshofer, 2005)

Validity coefficients - are we overcritical?

Random decisions	
FN .30	TP .20
TN .30	FP .20

Using a test with a validity of .40	
FN .21	TP .29
TN .39	FP .11

Using a test with a validity of .70	
FN .19	TP .31
TN .41	FP .09

Outcomes of Random Decisions and Decisions Based on Tests With a Base Rate of .50 Selection Ratio of .40

(Source: Murphy & Davidshofer, 2005)

Even low validities can improve decisions

Table 9-4 TAYLOR-RUSSELL TABLE SHOWING THE EXPECTED PROPORTION OF SUCCESSES WITH A BASE RATE OF .50

	Selection ratio									
Validity	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90
.00	.50	.50	.50	.50	.50	.50	.50	.50	.50	.50
.10	.54	.54	.53	.52	.52	.51	.51	.51	.51	.50
.20	.67	.64	.61	.59	.58	.56	.55	.53	.53	.52
.30	.74	.71	.67	.64	.62	.60	.58	.56	.54	.52
.40	.82	.78	.73	.69	.66	.63	.61	.58	.56	.53
.50	.88	.84	.78	.74	.70	.67	.63	.60	.57	.54
.60	.94	.90	.84	.79	.75	.70	.66	.62	.59	.45
.70	.98	.95	.90	.85	.80	.75	.70	.65	.60	.55
.80	1.00	.99	.95	.90	.85	.80	.73	.67	.61	.55
.90	1.00	1.00	.99	.97	.92	.86	.78	.70	.62	.56
1.00	1.00	1.00	1.00	1.00	1.00	1.00	.83	.71	.63	.56

(Source: Murphy & Davidshofer, 2005)

Predictive validity of ACs by criterion

Table 1: Weighted Validities Corrected for Artifacts

	Coefficient
Total	.37
Performance Criteria	
Ratings of General Performance	.36
Ratings of General Potential	.53
Ratings on Dimension	.33
Performance in Training	.35
Career Advancement	.36
Purpose of Assessment Center	
Promotion	.30
Early Identification	.46
Selection	.29
Research	.48

Source: http://www.wip.ddiworld.com/pdf/ddi_thevalidityofddiassessmentcenters_wp.pdf

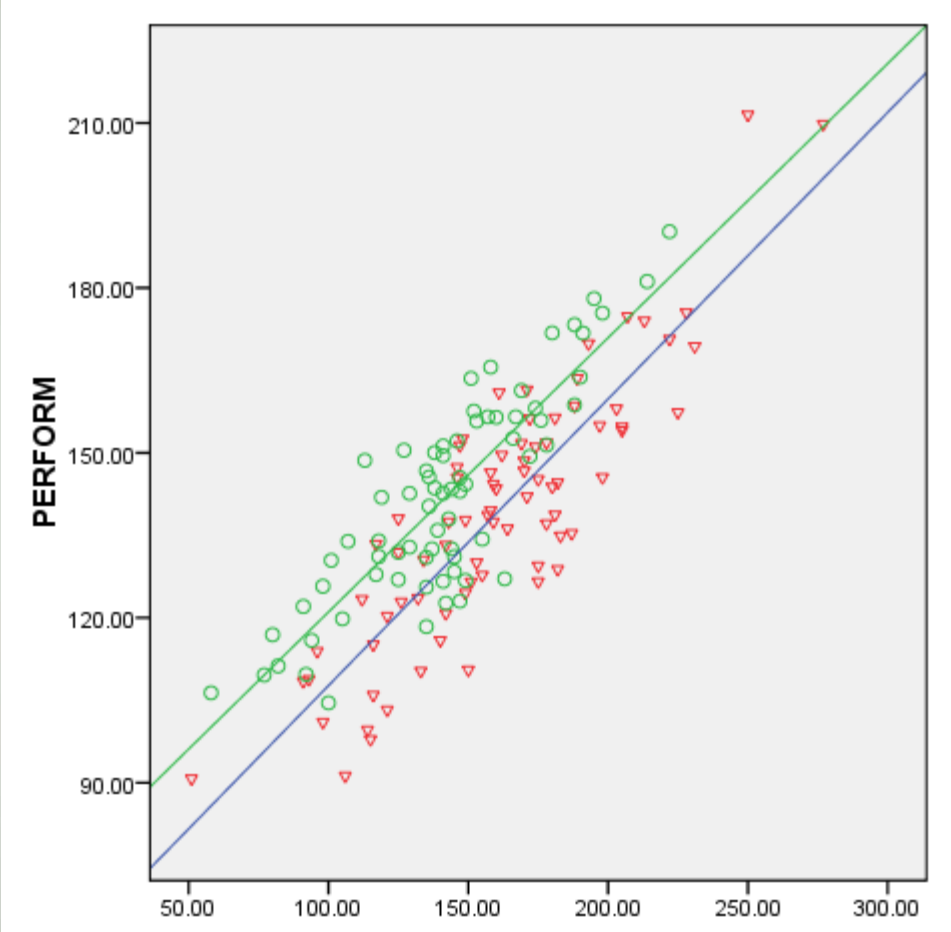
Validity (continued)

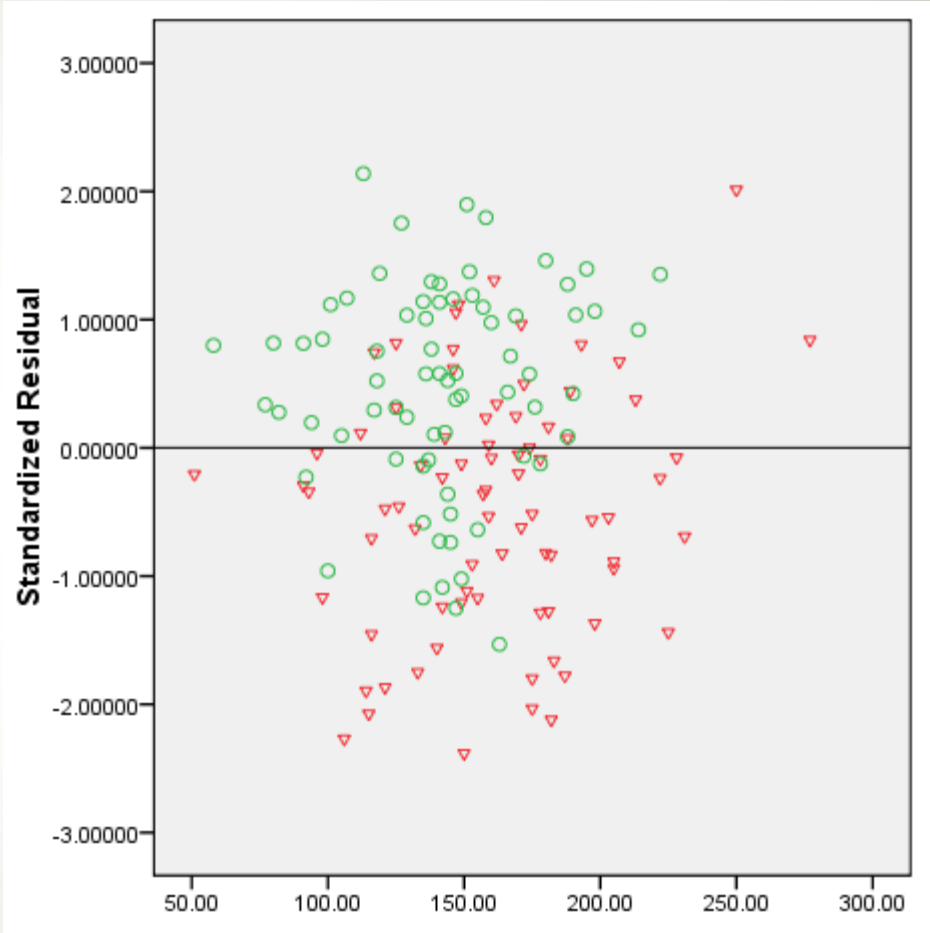
- Validation design
  - Does validation involve concurrent design before predictive design?
  - More convergent and discriminant validity evidence needed using multiple methods
- Too little programmatic validation studies - consider more validity generalisation studies (VGA)
- More alternative validation strategies needed, especially in small-sample designs

Fairness and Bias

- Adverse impact analyses, e.g. SR, AIR?
- Group difference scores (standardised mean differences, *D*) reported in manual?
- Measurement bias evidence (CFA, IRT)?
- Predictive bias evidence?

Predictive bias analysis





Utility

- Post implementation ROI evaluation?
- Incremental validity of ACs
- Coupling AC intervention with HR metric evaluation?

Legality

- ?

Claims


Validity
Legality
Reliability
CRITERIA
Acceptability
Fairness
Utility

Ethical pitfalls

Essential Elements of ACs

1. Job Analysis / Competency Modeling – to determine dimensions important to job success
2. Behavioural Classification – Overt behaviours, classified into meaningful and relevant categories
3. Assessment Techniques - Link from behaviours to dimensions to simulations
4. Multiple Assessments - e.g. tests, interviews, questionnaires, simulations

UNIVERSITY OF JOHANNESBURG

University of Cape Town · iYunivesithi yaseKapa · Universiteit van Kaapstad

LEMASA
Unlocking potential.

Essential Elements (continued)

5. Simulations – sufficient number of simulations, at least 1 simulation. Participants must **overtly display** behaviour; **demonstrate** response

6. Assessors / Observers – Multiple assessors observe and assess a participant

7. Assessor Training – Must show demonstrated competence

UNIVERSITY OF JOHANNESBURG

University of Cape Town · iYunivesithi yaseKapa · Universiteit van Kaapstad

LEMASA
Unlocking potential.

Essential Elements (continued)

8. Recording and Scoring Behaviour

9. Data Integration – Integration of each participants' behaviours – pooled information from assessors or statistical integration process

Discussion

Criteria, Practices and Ethical Pitfalls when Selecting Assessment Instruments for your Assessment Centre

Gert Roodt (University of Johannesburg, South Africa)
Francois de Kock (University of Cape Town, South Africa)
and **Sandra Schlebusch** (LEMASA, South Africa)

Speakers Corner
ACSG 14 – 15 March 2013
ICACM, Spier, Stellenbosch

UNIVERSITY OF JOHANNESBURG

University of Cape Town · iYunivesithi yaseKapa · Universiteit van Kaapstad

LEMASA
Unlocking potential.