A Review of AC Decision Making (Wash Up) Options: Striving for Fairness

Presenters:

Sean Keeley - Global Psychometric Lead at IBM Workforce Science and Analytics

Matt O'Sullivan - Senior Consultant at IBM Workforce Science and Analytics

Smarter Workforce

Objective

- Conference theme – “Managing fairness and diversity issues in Assessment Centres”

- Review and discuss the mechanical (arithmetic) decision making approach for ratings in terms of its strengths and weaknesses in enhancing fairness of the assessment centre process.

Agenda

- Current situation in regards to mechanical (arithmetic) versus consensus based decision making
- Issues and approaches of going to a mechanical approach
- A review of AC data looking at mechanical vs. consensus approaches
- Considerations and conclusions for the mechanical approach to AC decision making
- Discussion

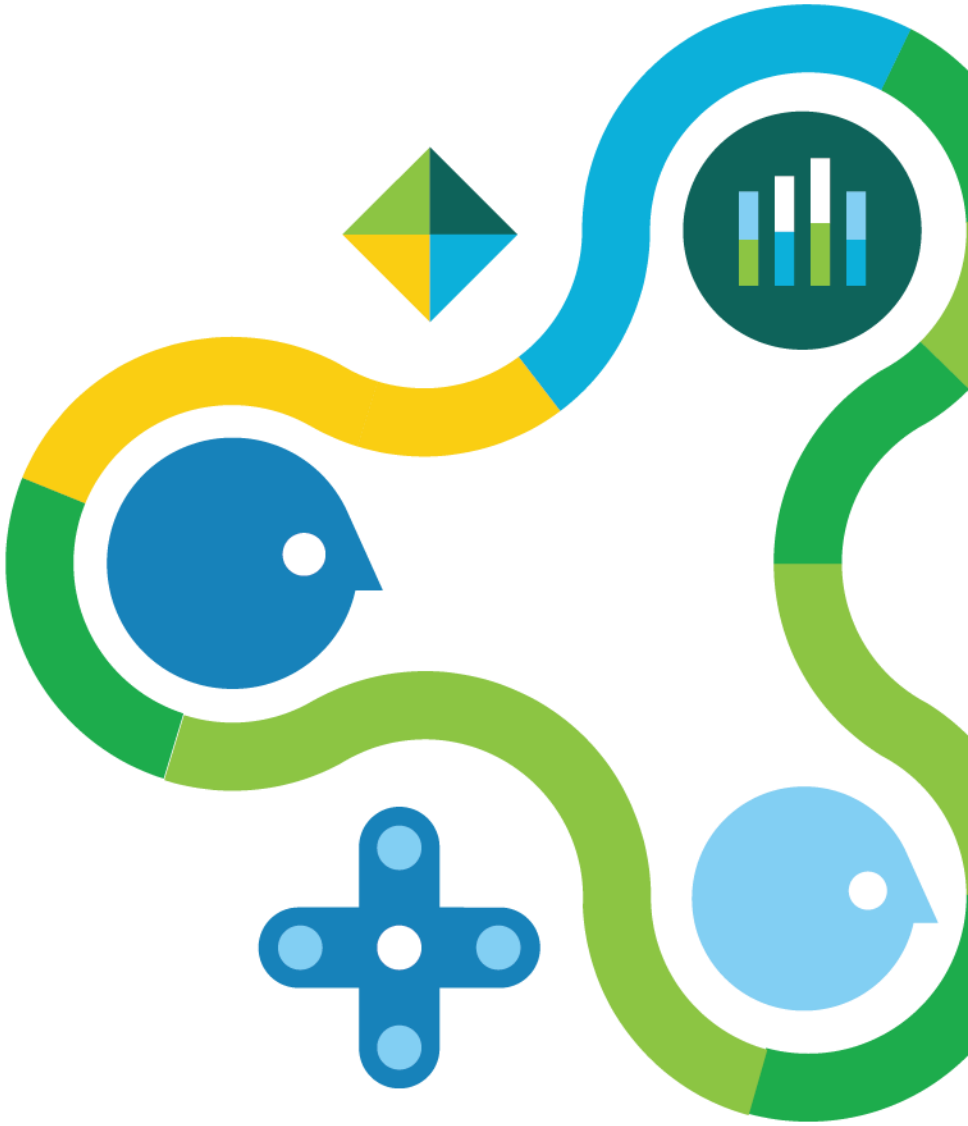
Assessment Centre fairness…

- Clarity and validity of assessment criteria
- Fair and valid exercises
- Objective and accurate assessments by assessors
- Proper AC administration and scheduling
- Objective, evidence based decision making

Decision Making – Focal Point

- AC Standards – these are management decisions related to the overall assessment (pass) standards (typically made in advance)
- Individual assessor decisions – these are the core assessment decisions on evidence and ratings during a centre
- Wash up (integration) decisions – these are the decisions related to the integration of ratings/evidence from exercises and what this means for final dimension ratings and AC outcomes (pass/fail)

Current Situation



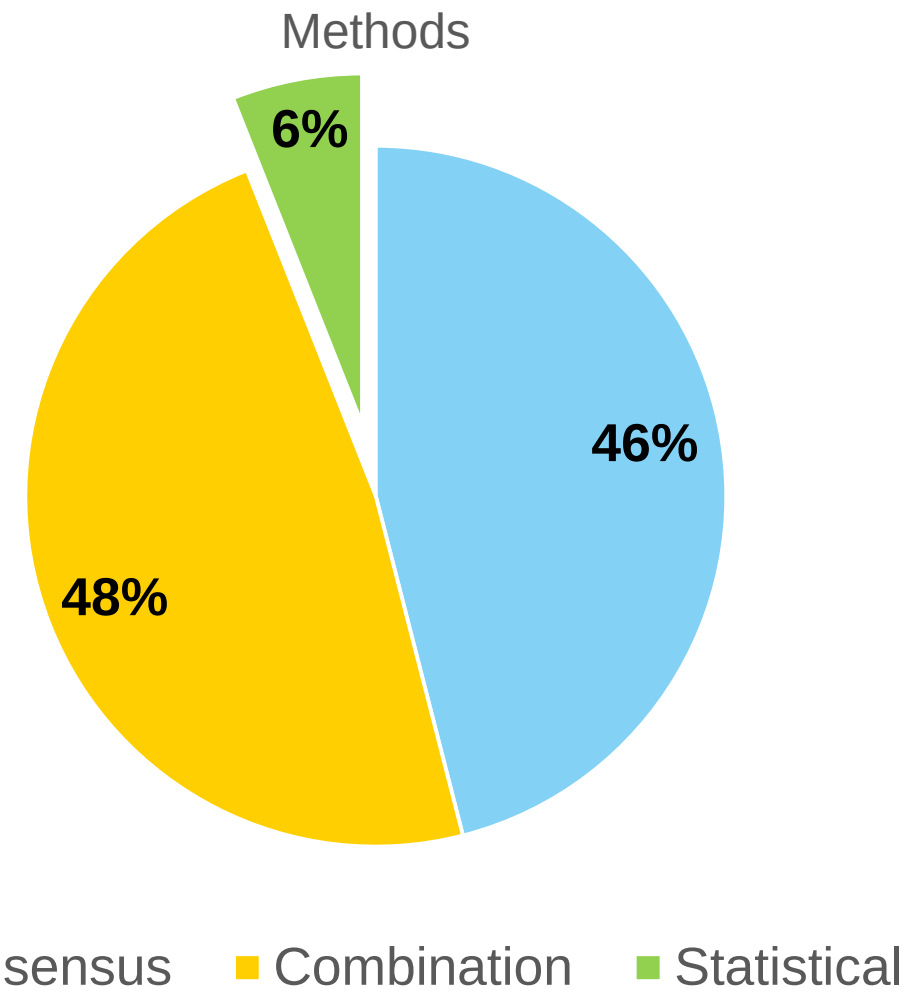
AC Standards related to Integration of Ratings

- ISO standard 10667 (2011) – The recent ISO standards Assessment Service Delivery focuses only on ‘ensuring that qualified personal are involved in interpreting the results and that there is clear rationale for interpretation and integration of different assessment data along with evidence to support this data’.

- International Task Force on Assessment Guidelines 2015: “The integration of each individual’s behaviours (individual dimension scores aggregated across exercises; exercise-specific scores; or potentially, depending on the purpose of the assessment centre, across-exercise scores aggregated into an overall assessment rating) must be based on pooled information from assessors or through a statistical integration process. The process used should be carried out in accordance with professionally accepted standards.”

- 2014 BPS guidelines Design and Delivery of Assessment Centres – “Rules for determining the Overall Centre Rating shall be defined based on the standards required in the job or role as identified through the job analysis.” “7.4 Arithmetic approaches shall be used to determine the Overall Centre Rating whenever the Centre is designed to facilitate selection decisions.”
  - “Research indicates that arithmetic combinations of scores are associated with much higher validities that consensual methods of determining final scores through discussion by Centre staff”.

Break Down of the Wash Up (Integration)



Ref: The 2012 A&DC Global Survey of Assessment Practice

Recent Research

- Chris Dewberry (2012) – qualitative study highlighting the prevalence of different, hidden influences (e.g. majority rule, hidden weightings, outside criteria, influential assessors etc.) on AC wash up meetings that reduced its aim for objectivity in decision making.

- Dilchert, S. and Ones, D. (2009) – Meta-analysis looking at consensus OARs, summed AC dimensions scores, optimally weights & summed AC dimension score to determine if there was incremental validity over cognitive tests and personality.

- Kuncel, N., Klieger, D., Connelly, B., and Ones, D. (2013) – mechanical integration of assessment data significantly improved the predictive power of selection tools compared to clinical, holistic methods (e.g. consensus meetings).

What's good about the consensus-based wash up?

– Opportunity to quality assure assessor evaluations

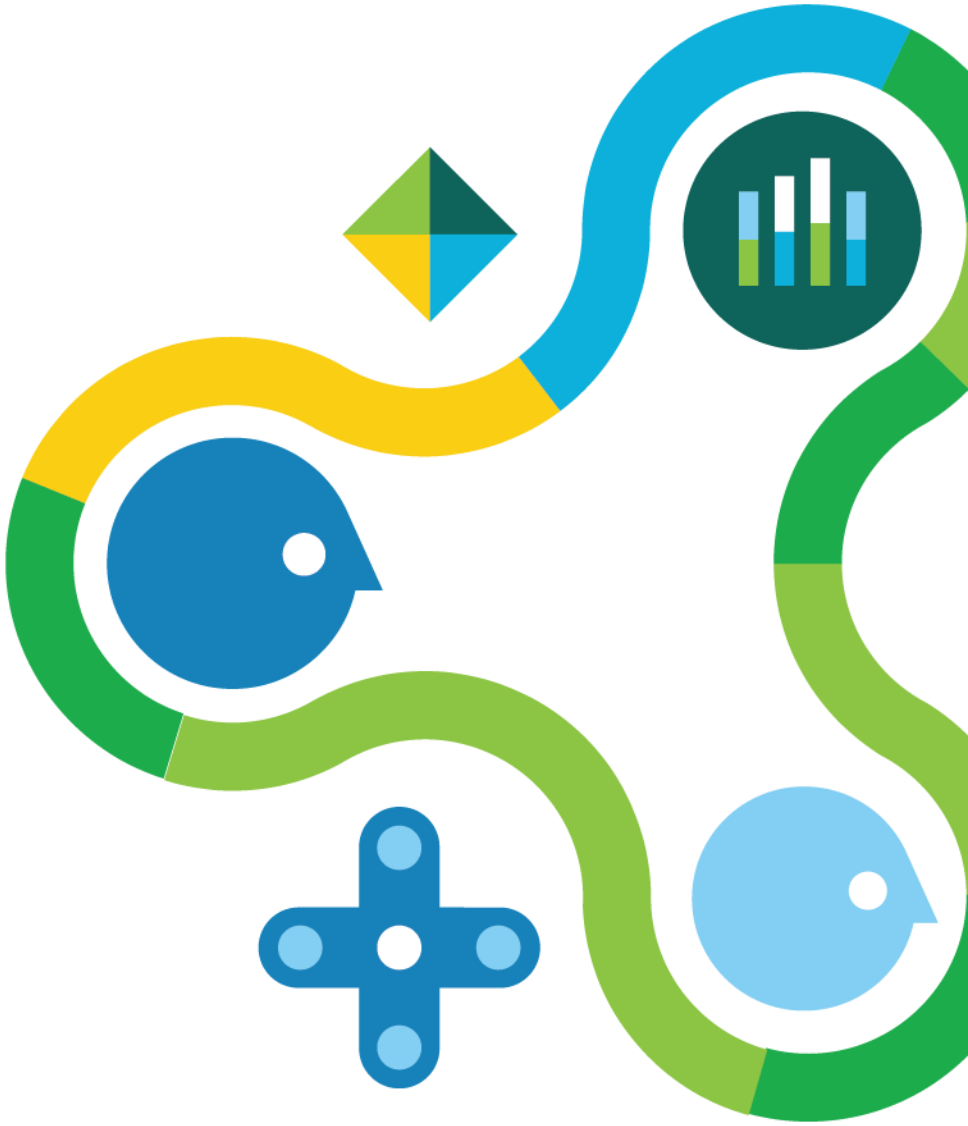
– Forum to carefully weigh up evidence across exercises

– A discussion of evidence not covered by indicators

– A chance to take account of exercise admin issues

– Opportunity to discuss development themes

Going Mechanical



A Mechanical (Arithmetic) Approach

Pros and cons

- Pros: less resource time, increased predictive validity, reduced subjectivity
- Cons: reduce/remove collective judgement/experience of assessor team

Key Options

1. Arithmetic Average – a set of dimension exercise scores averaged to produce final dimension scores

2. Weighted Average – a set of dimension exercise scores weighted to produce final dimension scores

3. Total Matrix Score – dimension exercise scores from matrix summed to produce an overall total score

Challenges

- Robust design, client expectations, quality assurance, job analysis, and scoring system.

Typical Matrix Scenarios

Typical Matrix Combinations											
5 point scale with 3 as the acceptable rating											
	Measured	Ex. 1	Ex. 2	Ex. 3	Ex. 4	Ex. 5	Whole	Decimal	Total Dim. Score	Gain / Lose	Consensus
Combination 1	2	3				2	3	2.5	5	Up 0.5	?
Combination 2	3		2	2	4		3	2.7	8	Up 0.3	2 or 3?
Combination 3	3	1		2		5	3	2.7	8	Up 0.3	2 or 3?
Combination 4	3	1	3		4		3	2.7	8	Up 0.3	3
Combination 5	3		3	3	3		3	3.0	9	-	3
Combination 6	3	2	3			4	3	3.0	9	-	3
Combination 7	3	2		2		5	3	3.0	9	-	2 or 3?
Combination 8	3	2		3		5	3	3.3	10	Down 0.3	3 or 4
Combination 9	3	3	3		4		3	3.3	10	Down 0.3	3 or 4

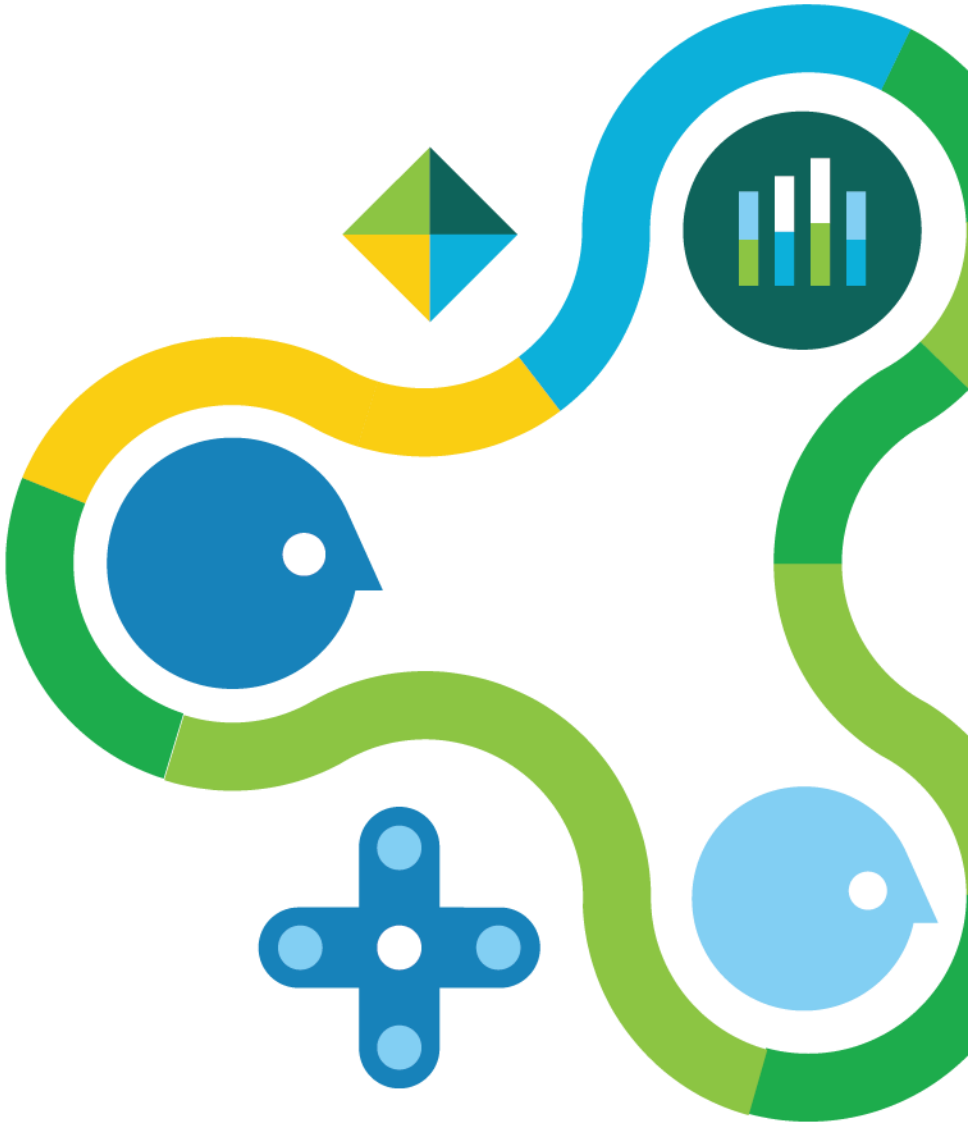
Extreme Examples

Best vs Worst

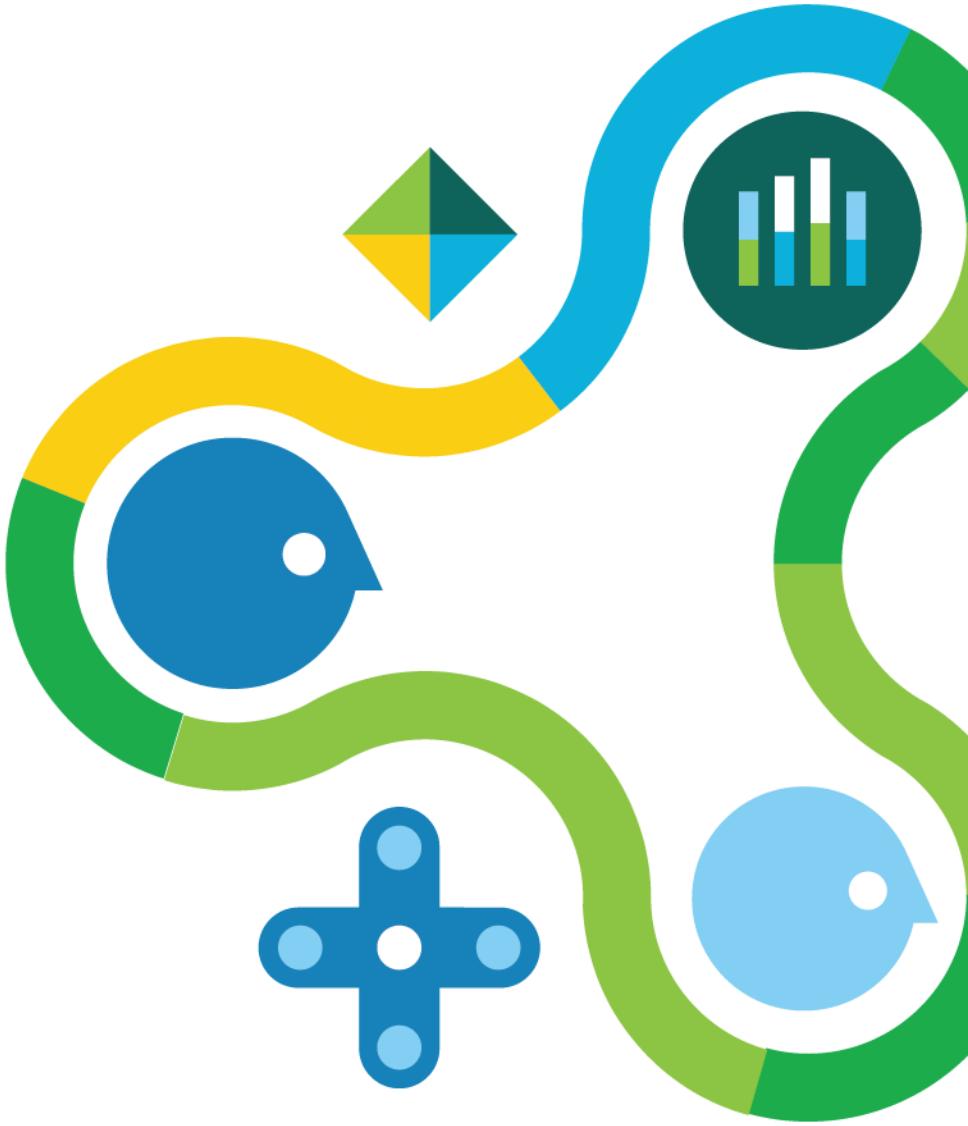
Pat Sample	5 point scale with 3 as acceptable rating								
Matrix	Ex. 1	Ex. 2	Ex. 3	Ex. 4	Ex. 5	1 decimal	Final Rating	Total	Gain / Lose
Competency 1		2	2	4		2.7	3	8	0.3
Competency 2		2	2	4		2.7	3	8	0.3
Competency 3	2	2			4	2.7	3	8	0.3
Competency 4	2		2		4	2.7	3	8	0.3
Competency 5	2	2		4		2.7	3	8	0.3
Competency 6	2		2		4	2.7	3	8	0.3
					Totals	**16**	**18**	**48**	**1.8**

Jane Bloggs	5 point scale with 3 as acceptable rating								
Matrix	Ex. 1	Ex. 2	Ex. 3	Ex. 4	Ex. 5	1 decimal	Final Rating	Total	Gain / Lose
Competency 1		3	3	4		3.3	3	10	-0.3
Competency 2		3	3	4		3.3	3	10	-0.3
Competency 3	3	3			4	3.3	3	10	-0.3
Competency 4	3		3		4	3.3	3	10	-0.3
Competency 5	3	3		4		3.3	3	10	-0.3
Competency 6	3		3		4	3.3	3	10	-0.3
					Totals	**20**	**18**	**60**	**-1.8**

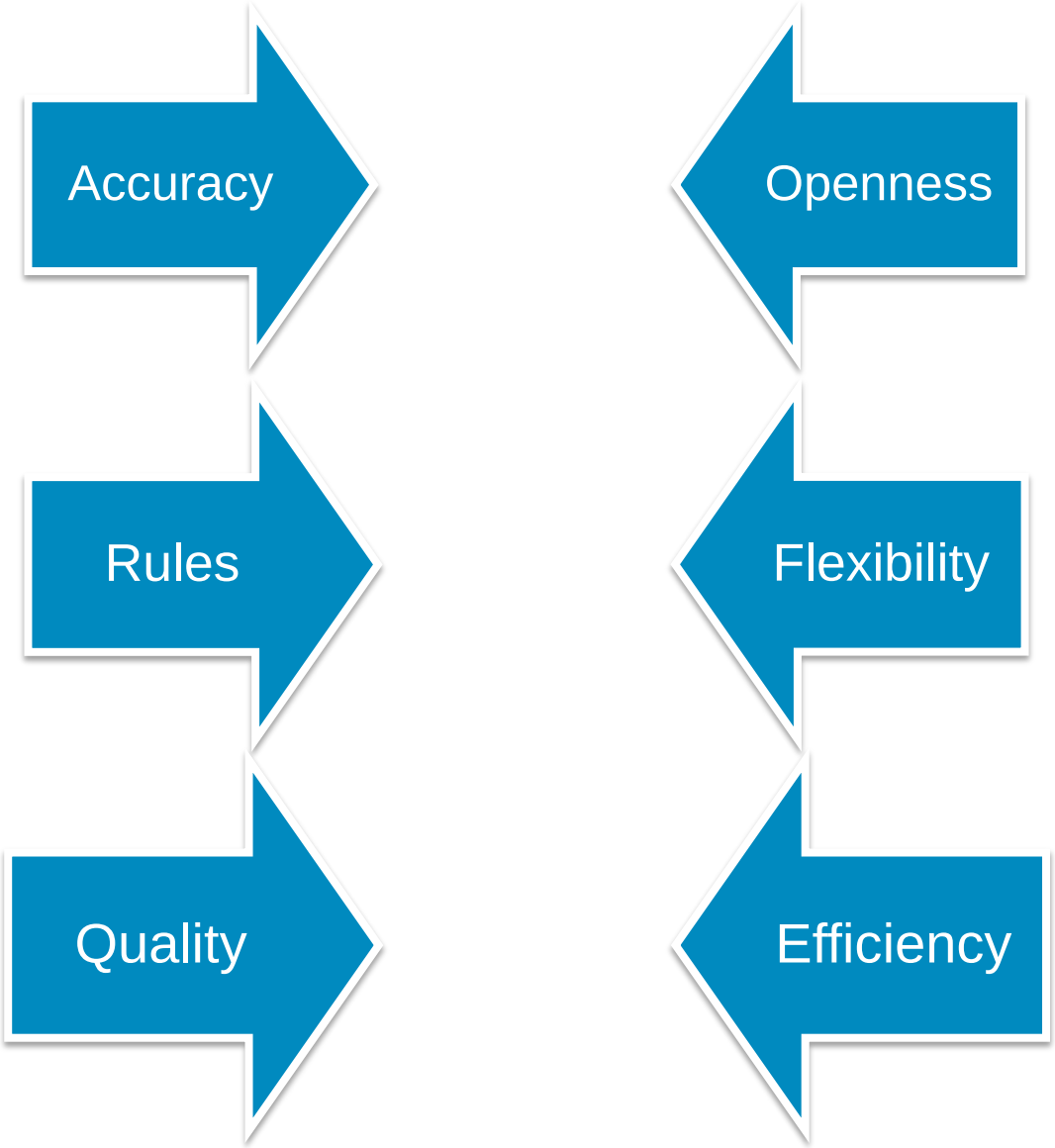
Data Set Findings



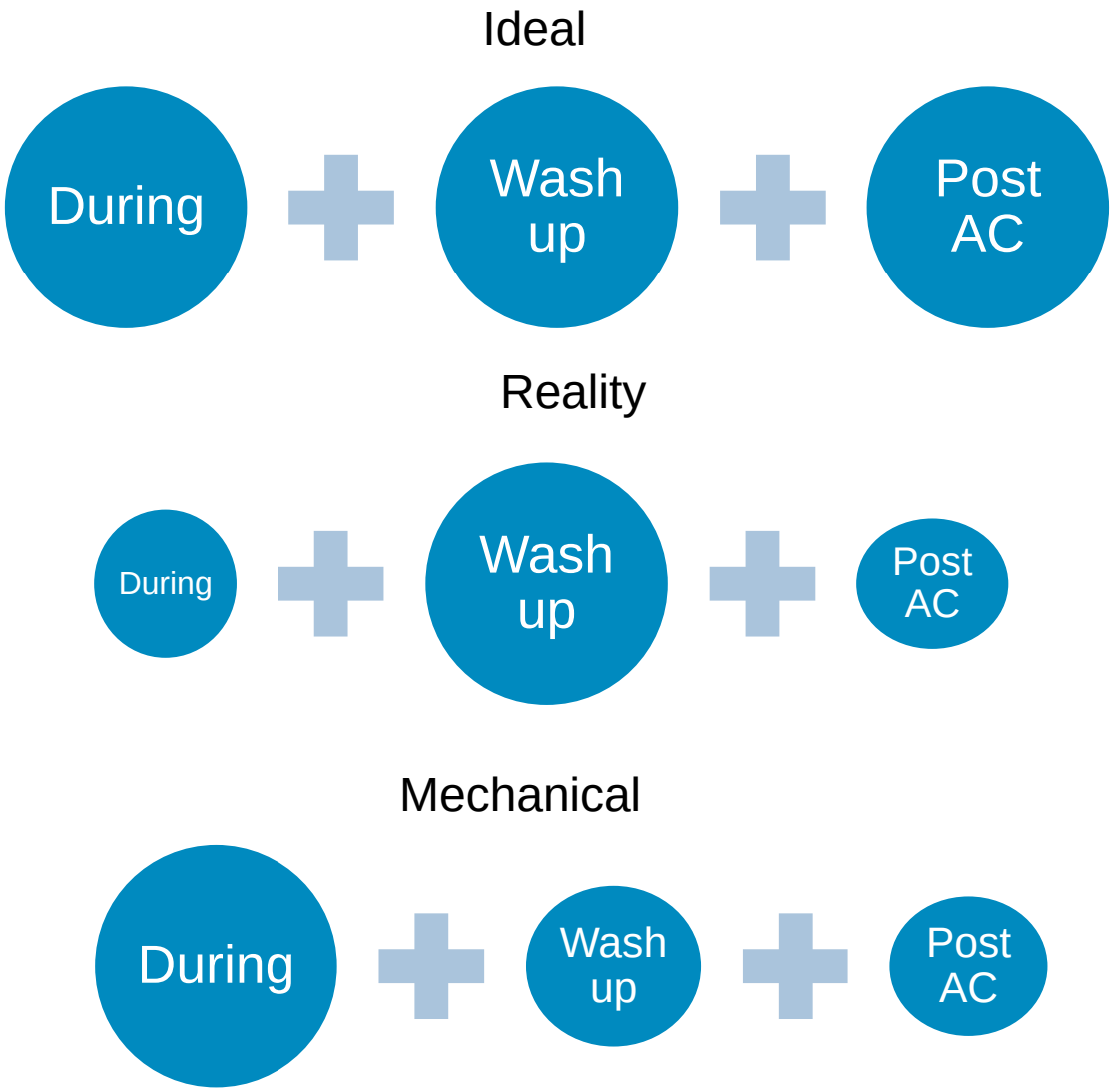
Conclusions



Managing the Tensions



What does mechanical mean for quality assurance of delivery?



Fitting the Process to Suit the Circumstances

- Mechanical processes can:
  - Provide more predictive validity
  - Reduce opportunity for potential biases/decision making errors
- So may be very appropriate when:
  - Assessor teams are inexperienced or mixed (internal/external)
  - Exercises have been evaluated with high validity/reliability
- Wash up sessions can still play a part:
  - Acting as a final check on data, interpretation or delivery issues or to capture developmental themes
- A consensus approach may be more appropriate if:
  - If assessors are very experienced and the exercises are new
  - Used to deal with exceptional cases (guided by decision rules)

Final Question

What is your approach / position on mechanical versus consensus based decision making for AC ratings?