



Evidence-Based Practices in Assessment Centres

Strengths, Concerns, and Challenges from a Global Survey

Presented by

Alyssa Gibbons, Colorado State University
Dan Hughes, a&dc

Colorado
State
University



a&dc
Assessment and
Development Consultants

Evidence-Based ACs?

- Assessment Centres enjoy a reputation as evidence-based practices (*Furnham, 2008; Lievens, 2002*).
 - Is this reputation warranted?
 - How well is research reflected in practice?
 - What are we doing well? What can we do better?
- We use a worldwide survey of AC practices to examine the state of the field from the lens of evidence-based practice.
 - For brevity, “AC” = assessment and/or development centre.

What Kind of Evidence?

“Evidence-based” can mean:

- Following the practices recommended by existing research and practitioner expertise for ACs in general.
- Collecting local evidence to support specific practices in a specific context.

Agenda

- Survey background & methodology.
- Are we following the general evidence we have?
 - Areas of strength and concern.
- Are we collecting the local evidence we need?
 - Areas of strength and concern.
- Are we practicing in areas where evidence is lacking?
- Conclusions.

Survey Background

Survey Goals

- Obtain a current snapshot of AC/DC practices **worldwide**.
- Incorporate both practitioner and academic viewpoints in survey design.
- Compare to previous surveys, eg:
 - *Povah, Crabb & McGarrigle (2008)*
 - *Krause & Thornton (2009)*
- Also capture emerging trends:
 - Technology, cultural adaptations



Respondents

- Recruited AC practitioners from professional associations, personal contacts, word of mouth, and social media.
 - Notably, **International Congress & ACSG!**

Internal: **44%**

HR Director
HR Manager
Training Director or Manager
Psychologist
Other HR Professional

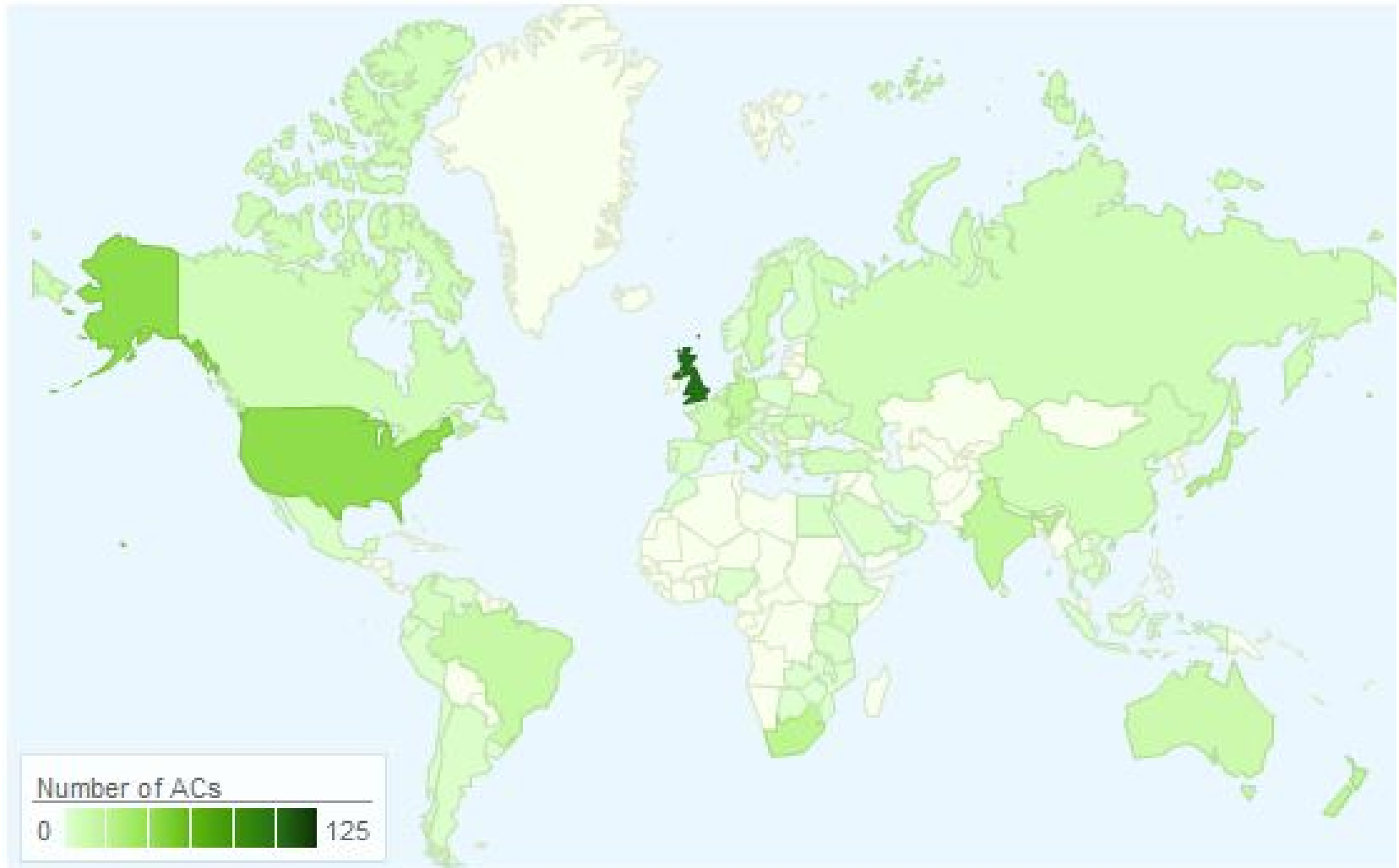
External: **56%**

Psychologist
HR Consultant

Method

- Respondents asked to describe **one specific AC or DC** they knew well.
 - **Not** mentally averaging across many ACs.
 - Could refer colleagues to describe other ACs.
 - In total, **511** unique ACs described.
- Online survey
 - Administered in English via Survey Monkey.
 - Anonymous format.
 - Approximately 59 questions (branching design).
 - Data collected August – November 2011.

Locations of ACs



Geographic and Cultural Diversity

- ACs implemented in **82** countries.
 - **77%** operated within one country only.
 - **23%** operated in multiple countries.

- Most common:
 - United Kingdom: **32%**
 - United States: **13%**
 - South Africa: **6%**

Are We Following the Evidence We Have?

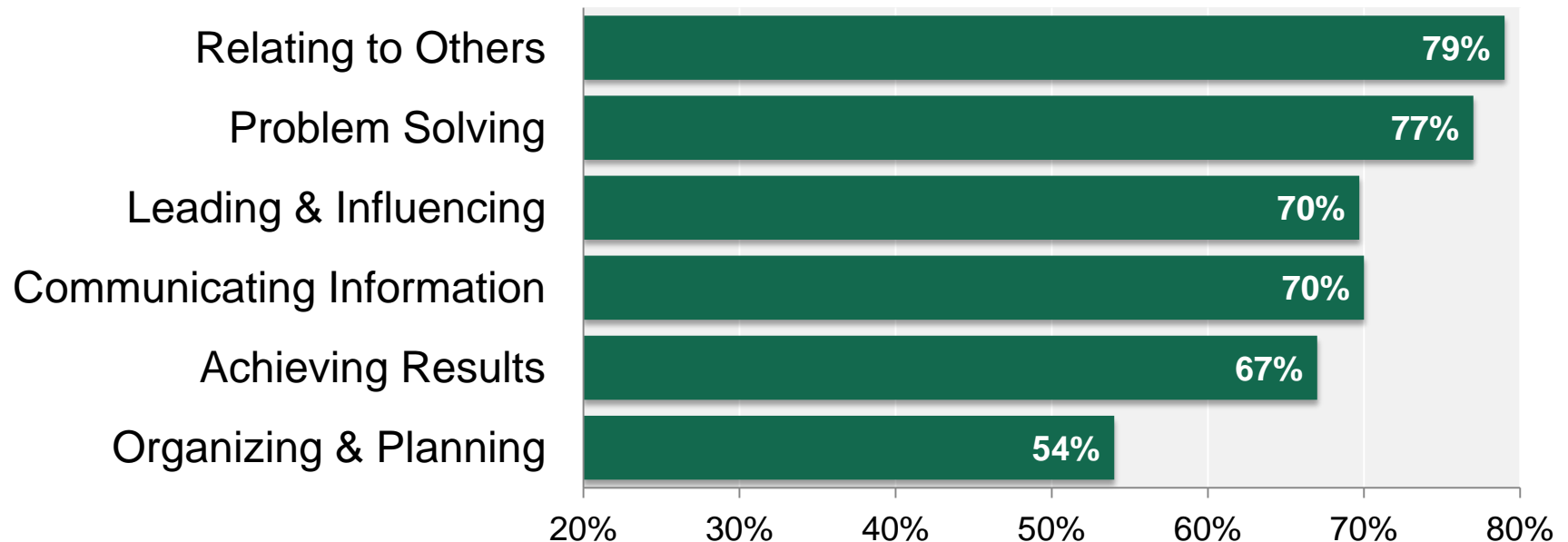
Strengths: Assessors

- **92%** of assessors receive some form of training.
 - Up from 80% in prior surveys (*Povah et al., 2008*).
- Training is empirically supported:
 - Dimensions: **94%** (*Scheicher, Day, Mayes, & Riggio, 2002*)
 - ORCE: **77%** (*Thornton & Zorich, 1980*)
- **98%** provide assessors with rating aids:
 - Definitions: **91%**
 - Example behaviors: **80%**
 - Checklists: **58%** (*Reilly, Henry, & Smither, 1990*).
 - Behaviorally anchored rating scales: **45%**

Strengths: Competencies

- **99%** use well-established competencies (at least one):

Assessed "to a substantial extent"



- All of these competency categories show predictive validity in AC meta-analyses (*Arthur et al., 2003; see also Meriac et al., 2008*).

Strengths: Competencies

- **77%** assess between 3 and 6 competencies per exercise (*Gaugler & Thornton, 1989*).
 - Although more competencies assessed across the AC as a whole ($M = 6.7$).
- Many fit the emerging **mixed-model** perspective (*Hoffman, 2012*):
 - **49%** collect both exercise and dimension/competency information.
 - **35%** use both to form the OAR.

Strengths: Additional Techniques

- When simulation exercises are combined with other techniques, they are most often techniques that show **incremental validity**:
 - Personality tests (**67%**)
 - Cognitive ability tests (**57%**)

Concerns: Hybrid Centers

- Research suggests that ratings made for decision-making are different from ratings made for development (*Murphy & Cleveland, 1995*).
 - Different errors and biases; distinguishing among dimensions.
- Often recommended that ACs focus on **either** assessment **or** development (*Thornton & Rupp, 2005*).
- However...
 - **28%** = primarily selection (but some feedback).
 - **26%** = primarily development (but some decision-making).
 - **24%** = equal emphasis
 - Only **23%** were single-purpose.

Concerns: Short ACs, Short Training

- **61%** of ACs last one day or less.
 - **74%** use 4 exercises or fewer.
 - **77%** assess each competency only 2 or 3 times.
- **34%** of assessors received less than 1 day of training, or no training at all.
- In **24%** of ACs, each assessor was responsible for more than 2 participants.

Concerns: Assessors & Feedback

- Evidence suggests that combining psychologists and managers produces the greatest validity, but only **25%** of ACs use both as assessors (*Gaugler et al., 1987*).
 - **16%** use neither managers nor psychologists (HR only)!
 - **34%** use assessors from only a single perspective.
 - **78%** use internal or external HR personnel – different perspective from either psychologists or managers.
- Although prompt feedback is important for development, **38%** of **developmental** ACs have more than a week delay in providing feedback.

Are We Collecting the Evidence We Need?

Strengths: Job Analysis

- **97%** of ACs reported using at least one job analytic technique.
 - **64%** used 3 techniques or more!
- Most common techniques:
 - **54%** reviewed the existing job description.
 - **52%** reviewed the existing competency model.
- But these techniques were rarely used alone.
 - **92%** of ACs using these techniques also used at least one other technique (e.g., interviews with management, interviews with job incumbents).

Strengths: Customization

- Only **12%** of ACs used off-the-shelf, prewritten exercises without customizing or adapting them to the organization.
- Job analysis + customization: real efforts to collect local data about the specific organization and target the AC to fit it.

Concerns: Assessor Evaluation

- Only **45%** of ACs require assessors to be formally certified.
 - Possible confusion about what “certification” means.
- **21%** of assessors are not evaluated **at all**.

Concerns: Outcome Evaluation

- **86%** of respondents evaluated their ACs.
 - But this evaluation was often limited to reactions of participants, assessors, and other stakeholders.
- Only **42%** conducted some type of validation analysis.
 - Content, predictive, concurrent, internal structure, or external construct validity.
 - Content analysis was by far the most common (**25%**) (*cf. Murphy, 2009*).

Concerns: Outcome Evaluation

- Although developmental ACs are supposed to produce changes in performance, only **18%** evaluate this change.
- Only **13%** conduct adverse impact or fairness analyses.
 - **50%** of North American ACs.
 - **23%** of multi-regional ACs.
- Only **15%** formally examine Return on Investment.

Where Are We Lacking Research?

Gaps: Culture

- **23%** of the ACs in our sample were administered in multiple countries.
 - **13%** in 2 or more major geographic regions.
- Many ACs (**74%**) make adaptations for culture.
 - Exercise content (**51%**), dimension definitions (**51%**), feedback processes (**42%**).
 - But how do we determine which adaptations to make?
- Lots of commentary on how culture *might* affect assessment (e.g., Bernthal & Lanik, 2010; Briscoe, 1997; Lievens & Thornton, 2005; Lanik & Gibbons, 2011).
 - Few if any published empirical studies (one: Melchers & Annen, 2010).

Gaps: Technology

- **57%** of ACs used at least one technology feature.
 - **23%** use automated or semi-automated reports.
 - **21%** use video recording.
 - **20%** use real-time phone interaction with participants.
- And many plan to add tech features within 2 years:
 - Computerized entry of behavioral observations (**28%**).
 - Automated or semi-automated reports (**26%**).
 - Automated or semi-automated scoring (**23%**).
- Users perceive a “small” to “medium” positive impact of technology overall.
 - $M = 3.38$ on a 5-point scale.
 - Most positive impact on **efficiency** of running the centre.

Gaps: Technology

- Specific features correlated with **perceived** benefit of technology:
 - Automated scheduling: $r = .23$ with overall impact, $r = .39$ with cost.
 - Automated scoring: $r = .21$ with overall impact.
 - Online simulations: $r = .23$ with impact on realism.
 - Automated reports: $r = .24$ with impact on duration.
- Very little published research evaluating technology features.
 - Video recording doesn't have much of an effect (*Ryan et al., 1995*).
 - Computerized exercises *can* predict criteria above cognitive ability (*Lievens, van Keer, & Volckaert, 2010*).
 - But what about phone interactions, delayed interactions, virtual reality, etc.?

Gaps: Integrated Exercises

- **39%** of ACs used integrated or “day in the life” exercises with multiple components.
 - Average of **4.5** components.
- Technology makes integrated exercises easier to deliver (*Howard, 2008*).
- But we know little about how the interdependence of integrated components affects performance, rating accuracy, and outcomes.

Gaps: Development

- **62%** of ACs involve a substantial development component.
 - **84%** of ACs – including selection emphasis ACs – provide feedback beyond pass/fail.
- But development is seldom evaluated as an outcome (*Rupp et al., 2006*).
 - What AC design features produce development?
 - Is a good AC a good DC?
 - Which differences matter?

Conclusions

Are ACs Evidence-Based?

- Where research evidence is clear and unambiguous, most ACs tend to follow it.
 - We have the best evidence about internal design features (dimensions, exercises), assessor training.
 - Most ACs follow research recommendations in this area.
- But logistical and practical considerations sometimes override evidence.
 - Pressure to do more with less.
 - Short assessor training, hybrid centers, etc.

Where is More Evidence Needed?

- Although some are rigorous about evaluating their own ACs, gathering evidence about effectiveness/validity of individual ACs is not as widespread as it could be.
- Dearth of evidence on several critical issues facing ACs of today:
 - Technology.
 - Culture.
 - Impact on development.

Moving Forward: Closing the Gaps

- More and better basic research.
- More and better evaluation of operational ACs.
- More multi-AC studies:
 - Generalizability and replicability.
- Establish effects of design choices on **bottom line outcomes**.
 - Validity, performance improvement, ROI.

Questions?

You can download the full Research Report for this survey at www.adc.uk.com/ACsurvey