



Traditional and Technological Assessment Centers: Capturing the Same Constructs, Creating the Same Outcomes?

Alyssa M. Gibbons, *Colorado State University*

Martin Lanik, *Global Assessor Pool*

38th International Congress on Assessment Center Methods
October 22-24, 2014, Alexandria, VA

Colorado
State
University



GLOBAL ASSESSOR POOL



Overview



- A trend, a research gap, and a critical question.
- A study: how closely can we compare a traditional and a high-technology assessment center?
 - Design choices and challenges.
 - Preliminary results: within-participants comparisons, convergence across methods, relationships with other variables, participant reactions.
- A few lessons learned, and still more questions.



Technology in ACs is Here to Stay

- ▶ The majority of ACs (~64%) now use some form of technology. *(Gibbons, Hughes, Riley, Thornton, & Sanchez, 2013)*
 - ▶ More often to support the administrative side than to enhance or facilitate the participant experience.
 - ▶ But participant-experience technologies are increasingly common, e.g.:
 - ▶ Delivering online or computer-based simulations. *(Lievens, Van Keer, & Volckaert, 2010)*
 - ▶ Role plays via phone, email, etc. *(Gowing, Morris, Adler, & Gold, 2008)*
 - ▶ Use of animation and virtual reality. *(Hatfield, Gurira, & Harvey, 2013; Hawkes, 2013)*
- ▶ Technology issues are prominently featured in the new Guidelines!



Defining a Difference



- Infinite possible variations – hard to draw clear lines.
- For our purposes here:
 - **A traditional AC** is one that brings candidates, assessors, and role players together for in-person, real time interaction.
 - **A technological AC** is one that uses technology (e.g., phone, internet) to allow candidates, assessors, and role players to interact when separated by time and/or space.



Obvious Practical Differences



Technological

- Expensive to develop.
 - In the beginning...
- Increase flexibility of administration.
 - Time, location, etc.
- Appeal to tech-savvy participants.

Traditional

- Expensive to execute.
- Less dependent on reliable technology.
 - Internet connectivity, etc.
- Familiar for less tech-savvy participants.



Similar Successes



- ▶ Technological ACs **can** elicit positive participant reactions. *(e.g., Hawkes, 2013)*
 - ▶ But so can the traditional kind! *(e.g., Dodd, 1977; Macan et al., 1994)*
 - ▶ Is there a difference? For whom?
- ▶ Technological ACs **can** show criterion-related validity. *(Lanik, Dvergstal, Dvorak, Gibbons, 2014; Lievens et al., 2010).*
 - ▶ But, of course, so can the traditional kind! *(e.g., Gaugler et al., 1987; Hardison & Sackett, 2004; Hermelin, Lievens, & Robinson, 2007)*
- ▶ Very few direct comparisons.
 - ▶ A few exceptions: *Denunzio, Marira, & Collins, 2013; Fay, Frame, Kenworthy, & Lopez, 2013*



A Critical Question

- The question so far has been “are technological ACs any good?”
 - The question we need to ask now is “what do we change when we introduce technology?”
- 



Possible Differences



- Subtle differences in overt behavior may tap different aspects of underlying constructs.
 - E.g., written vs. phone vs. face-to-face communication.
 - Different strategies to accomplish the same tasks?
- Different attentional demands & information resources for assessors.
 - Written records vs. recordings vs. real-time observation.
 - Nonverbal cues tend to be influential, but are they important?
 - The tantalizing possibilities of metadata.



Current Project



- Can we directly compare a technological AC and a traditional AC, holding constant as many elements as possible yet maintaining fidelity to each approach?
- Are we measuring the same constructs in the same way?
 - Compare mean differences, convergence, relationships with other constructs.
- When participants experience both, how do they compare?



Holding Constant



- Participants
 - Within-subjects design – all participants completed both versions.
- Competencies & behavioral indicators.
- Assessors
 - Drawn from the same pool.
 - All received the same training.
- Time
 - Participants completed both on the same day, in the same location.



Three (Conflicting?) Design Aims

- As parallel as possible.
 - Same competencies, same definitions, same behavioral indicators.
 - Phone role play ~ in-person role play.
- Unique features of each method.
 - Traditional included an LGD.
 - Conflict situation addressed by email ~ in person.
- Holistic experience for participants.
 - Keep the same organizational context.
 - All materials delivered through email client.



Details



- 37 participants.
 - Managers and aspiring managers.
- Participants or their organizations paid a nominal fee for a one-day DAC.
 - Morning: GAP eSimulator™
 - Afternoon: Parallel traditional DAC
- Completed surveys:
 - Pre-program Big Five personality measure (IPIP REF HERE).
 - Post-program reactions.
- Assessors were graduate students in IO psychology or related fields.
 - 10+ hours training.

Mean Differences?

	Technological		Traditional	
	M	SD	M	SD
Strategy Execution	1.89	0.45	1.62	0.35
Analytical Thinking	2.04	0.34	2.12	0.36
Leading Change	2.12	0.45	1.90	0.35
Communicating with Impact	2.29	0.32	2.39	0.25
Coaching Others	2.13	0.37	2.44	0.30
Emotional Intelligence	1.90	0.37	2.27	0.32
Leading Teams	1.77	0.44	1.79	0.38

- Higher scores in the **technological AC** on strategy/change competencies.
- Higher scores in the **traditional AC** on interpersonal competencies.

Convergence?

Traditional

Technological

	SE	AT	LC	CI	CO	EI	LT
Strategy Execution	0.43	0.41	0.47	0.27	0.33	0.07	0.21
Analytical Thinking	0.60	0.60	0.56	0.44	0.42	0.28	0.24
Leading Change	0.51	0.52	0.44	0.38	0.31	0.23	0.15
Communicating with Impact	0.30	0.36	0.21	0.55	0.09	0.45	0.18
Coaching Others	0.44	0.43	0.44	0.37	0.31	0.08	0.10
Emotional Intelligence	0.37	0.30	0.29	0.20	0.37	0.14	0.19
Leading Teams	0.59	0.56	0.51	0.37	0.31	0.21	0.09

- Some competencies converge, but many don't.
- Average convergent $r = .37$
- Average discriminant $r = .52$ (technological), $.50$ (traditional).

Relationships with Personality Measures

Technological AC	A	C	N	E	O
Strategy Execution	0.15	-0.02	0.13	0.24	0.10
Analytical Thinking	0.10	-0.01	0.07	0.14	0.12
Leading Change	0.06	-0.08	0.15	0.18	0.20
Comm. with Impact	0.20	0.03	0.32	0.23	0.22
Coaching Others	0.15	-0.13	0.16	0.02	-0.03
Emotional Intelligence	0.23	0.12	0.11	-0.01	0.10
Leading Teams	0.35	0.07	0.15	0.07	0.16

All N = 34

bold = $p < .05$

Technological AC:
average absolute $r = .13$

Traditional AC:
average absolute $r = .19$

Traditional AC ratings
showed more and
stronger relationships with
personality traits.

Traditional AC	A	C	N	E	O
Strategy Execution	0.30	-0.24	-0.15	0.12	0.20
Analytical Thinking	0.12	-0.26	0.10	0.23	0.37
Leading Change	0.11	-0.36	-0.01	0.37	0.30
Comm. with Impact	0.04	-0.23	0.05	0.40	0.39
Coaching Others	-0.08	-0.18	-0.13	0.22	0.11
Emotional Intelligence	0.02	-0.05	0.06	0.40	0.34
Leading Teams	0.07	-0.04	0.15	0.23	0.15



Apples to Oranges?

- So far, results suggest some noticeable differences!
 - Caveat: small sample, but trends make sense.
- Is this just a failure of parallelism?
- Focus on comparing the most similar exercise.
 - Coaching role-play.
 - Only the character and medium (phone vs. face-to-face) differed.

Role Play Exercise: Mean Differences?

	Technological		Traditional	
	M	SD	M	SD
Leading Change	2.04	0.60	2.29	0.61
Communicating with Impact	2.40	0.30	2.48	0.39
Coaching Others	2.23	0.44	2.38	0.32
Emotional Intelligence	1.94	0.42	2.23	0.41

- Again, higher scores in the **traditional AC** on interpersonal competencies.

Role Play Exercise: Convergence?

Technological	Traditional			
	LC	CI	CO	EI
Leading Change	0.24	0.31	0.48	0.30
Communicating with Impact	0.22	0.26	0.41	0.25
Coaching Others	-0.08	0.04	0.34	0.01
Emotional Intelligence	0.13	0.12	0.39	0.28

- Average convergent $r = .28$
- Average discriminant $r = .65$ (technological), $.65$ (traditional).

Role Playing Exercise: Relationships with Personality Measures

All $N = 34$

bold = $p < .05$

Technological AC	A	C	N	E	O
Leading Change	-0.06	-0.03	0.06	0.04	-0.11
Comm. with Impact	-0.11	-0.03	0.15	0.06	0.07
Coaching Others	0.15	-0.12	0.13	-0.08	-0.14
Emotional Intelligence	0.18	0.14	0.04	-0.11	0.02

Technological AC:
average absolute $r = .09$

Traditional AC:
average absolute $r = .19$

Again, traditional AC ratings showed more and stronger relationships with personality traits.

Traditional AC	A	C	N	E	O
Leading Change	-0.37	-0.24	0.06	0.40	0.06
Comm. with Impact	-0.34	-0.26	-0.09	0.24	0.20
Coaching Others	-0.19	-0.21	-0.09	0.29	0.07
Emotional Intelligence	-0.09	-0.05	-0.02	0.38	0.21



Participants' Reactions



- Captured after receiving feedback.
- 3 items asking participants to compare the two sets of activities directly.
 - Part of a larger survey that did not differentiate the two.
- Considerable attrition.
 - 14 surveys completed out of 37; 38% response rate.
- Also captured some qualitative reactions during the program and in-person debriefing.



Participants' Reactions: Post-Program Surveys

	Realistic?	Challenging?	Enjoyable?
Technological	1	9	1
Both	12	4	7
Traditional	1	1	6

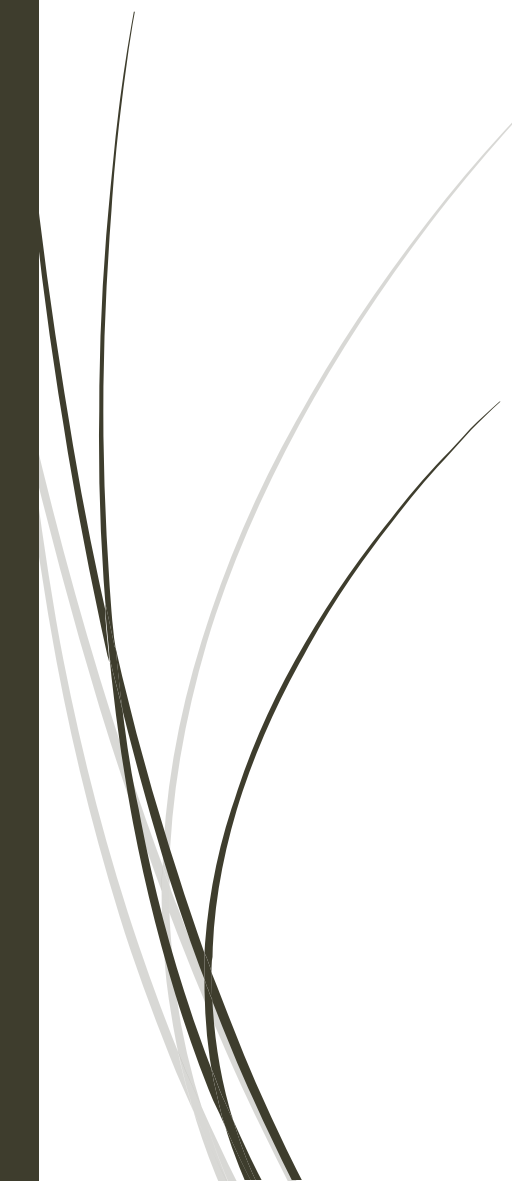


Participants' Reactions: Qualitative Comments

- ▶ Traditional session felt slower, easier after the technological session.
 - ▶ Not as dynamic, tasks not as integrated.
 - ▶ Afternoon analytical exercise perhaps too subtle.
- ▶ Mixed views on the in-person vs. phone interactions.
- ▶ Fidelity of technological session was higher for some participants than others.
 - ▶ Desk jobs, virtual communication vs. hands-on, field management.



Participants' Reactions: Qualitative Comments

- Participants liked the (potential) convenience of the technological AC.
 - Assessors preferred it, too!
 - Glitches a little more noticeable in the technological AC.
- 



Lessons Learned: Parallelism is a Challenge

- Many considerations – degrees and dimensions of parallelism.
- Balancing strict parallelism with quality of participant experience is challenging in a within-subjects design.
- Some features of each kind of AC are simply hard to parallel!
 - Group interactions, dynamic exercises.
- A pure within-subjects comparison is difficult to come by.

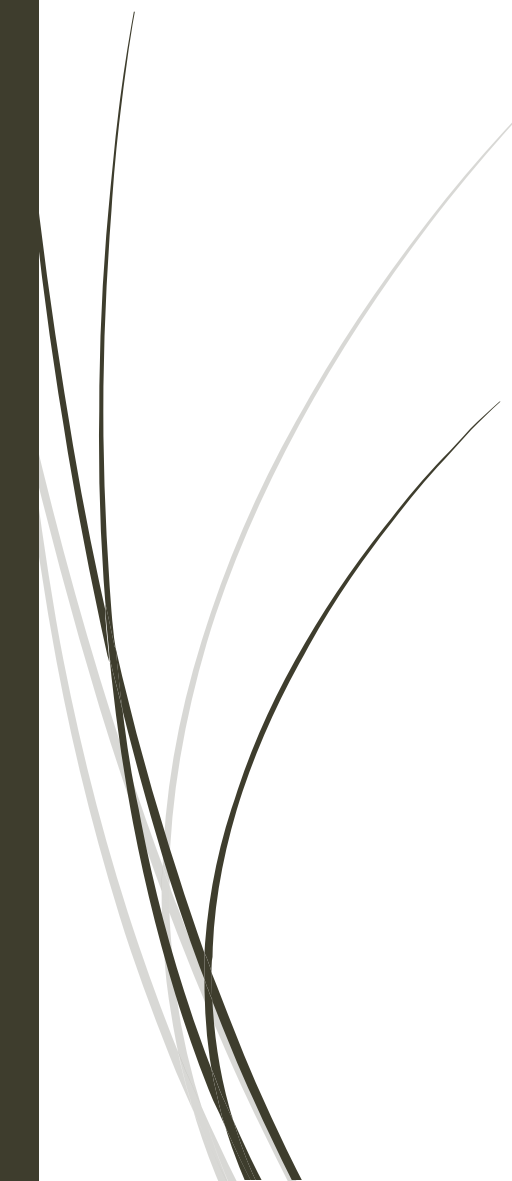


Lessons Learned: Are the Constructs the Same?

- Small sample precludes strong conclusions about measurement equivalence.
- But there are some intriguing trends.
 - Higher scores on interpersonal skills in person, higher scores on analytical skills online.
 - Little convergence across the two forms.
 - Personality traits were more strongly related to traditional AC ratings.
- Why? Are we measuring different aspects of the same construct, missing relevant behavior, or reducing contamination?



Lessons Learned: Putting the Genie Back in the Bottle?

- Convenience of technological ACs is hard to resist.
 - For participants, assessors, and administrators.
 - As high-tech ACs become more common, may need a more compelling case to justify in-person assessment.
 - Particularly when technology is a regular and essential part of the job.
- 



Thank you! Any questions?

We gratefully acknowledge assistance from Kyla Dvorak, Jaclyn Menendez, and Yi Ren in collecting and managing the data for this project.