

Assessment centers in the virtual age: validity and fairness in gender and age

Yumna Zafar Usmani

Glasford International, Bologna, Italy

Gerardo Petruzziello

*Department of Psychology “Renzo Canestrari”, University of Bologna,
Bologna, Italy*

Barbara Rizzo

Glasford International, Bologna, Italy, and

Marco Giovanni Mariani

*Department of Psychology “Renzo Canestrari”, University of Bologna,
Bologna, Italy*

Abstract

Purpose – This study focused on integrating digital methodologies in personnel selection processes and explored the psychometric properties of virtual assessment centers as alternatives to traditional ones. We evaluated the validity of a virtual application of an assessment center and its fairness in gender-based evaluations.

Design/methodology/approach – We collected data from 120 managers at an Italian company undergoing evaluations through an online assessment center. This virtual platform administered tests in English across three exercise phases to gauge competencies such as teamwork, decision-making, execution and assertiveness. The study employed Pearson correlation indexes, principal component analysis, McDonald’s omega, binary logistic regression, MANOVA, chi-square and the 4/5th rule for statistical analyses to assess the validity and gender fairness of the virtual assessment method.

Findings – The results indicate that virtual assessment centers show promising validity and do not exhibit adverse impacts based on gender or age. This suggests that they are a fair method for evaluating candidates. These results support the potential of digital methodologies to serve effectively in personnel selection processes, offering theoretical and practical implications for future application and research.

Originality/value – This article contributes novel insights into the evolving field of digital personnel selection processes. By systematically evaluating the validity and gender fairness of virtual assessment centers, the research addresses a significant gap in existing literature, providing a foundational basis for further exploration into virtual assessment methodologies.

Keywords Virtual assessment centers, Personnel selection, Gender, Age

Paper type Research paper

Introduction

Personnel evaluation for selection, training, and development is vital to any organization. Assessment centers (AC) have a long history (e.g. [Ansbacher, 1941](#); [Taft, 1948](#); [Vernon, 1950](#); [Thornton & Byham, 1982](#)) and nowadays are widely popular as many organizations use them for personnel evaluation and selection ([Hausknecht & Heavey, 2017](#)). Scholars have viewed ACs as a central mode of personnel evaluation since their inception over 100 years ago ([Thornton & Rupp, 2006](#); [Baczynska, Skoczeń, Thornton, & Chen, 2024](#)). They have maintained their popularity because of their numerous strengths. Scholars have associated them with content validity, negligible adverse effects, and high criterion-related validity (e.g.



performance prediction) (Vernon, 1950; Schmitt, Gooding, Noe, & Kirsch, 1984; Gaugler, Rosenthal, Thornton, & Bentson, 1987; Iles, 1992; Thornton & Rupp, 2006) and supported it as a valid selection method (Rupp, Thornton & Gibbons, 2008; Arthur, Woehr, & Maldegen, 2000). The AC technique has evolved with the emergence of online, digital, and virtual ACs – terms often used interchangeably in prior research (e.g. Howland, Rembisz, Wang-Jones, Heise, & Brown, 2015; Avni & Luria, 2022) – as part of the broader adoption of technology and digital processes driven by the fourth industrial revolution and accelerated by the pandemic. In this study, we will refer to them collectively as virtual assessment centers (Virtual ACs).

The advancement of digital assessment techniques in the professional world is necessary to keep up with fast-paced innovations. It requires research to evaluate the techniques' robustness in sight of massive use in the future. On the other hand, the need to study the validity of these new assessment forms has become increasingly important in research. Early studies laid the groundwork by exploring the transition from paper-and-pencil tests to computer-based tests and from in-person interviews to remote interviews, highlighting the critical need for continued validation of these digital methodologies (Sackett & Lievens, 2008).

Based on these premises, this work had two goals. First, we sought to provide initial virtual ACs' construct validity evidence regarding internal and external validity (Grimm & Widaman, 2012). We explored whether we could synthesize the AC results in an overall rating (Thornton & Gibbons, 2009) and whether the AC results align with the promotion to the middle/high management level.

Second, considering the persistent gender disparities in career progression (e.g. Murphy, Callander, Dohan, & Grandis, 2021) and the possible advantage of younger people in using technologies for personnel selection (Woods, Ahmed, Nikolaou, Costa, & Anderson, 2019), we aimed to assess and establish the methodology's adverse impact and focus on gender and age differences by examining the statistical significance of disparities between groups and investigating the potential influence of lead assessors' gender on the equity of the selection process. We shall contribute to the literature about digital selection, which needs scientific support, and provide indications for reliable and fair use of virtual ACs during digital selection procedures by exploring their characteristics.

Background and research questions

People use assessment centers in a variety of settings, for example, businesses, governmental organizations, educational institutions, and armed forces (Arnoneit, Schuler, & Hell, 2020) to assist organizations in certification of employees' competencies, promotions, selection for supervisory or management positions, and identification of employees with talent or high potential. Lately, the dive into the digital era has affected ACs, as we witnessed online assessments surge (Woods, Ahmed, Nikolaou, Costa, & Anderson, 2019), such as online applications, psychometric testing, digital interviews, gamified assessments, and social media mining. Both candidates and organizations favor digital techniques due to the higher comfort of use compared to physical assessment forms and the lower financial and geographical constraints (Templer & Lange, 2008; Howland, Rembisz, Wang-Jones, Heise, & Brown, 2015; Woods, Ahmed, Nikolaou, Costa, & Anderson, 2019). The use of digital tools in ACs has marked a technique's evolution. However, the field still needs a thorough exploration of this evolution, even though scholars have studied specific methods commonly used in ACs, such as tests and interviews, in their digital forms (e.g. Van Iddekinge, Raymark, Roth, & Payne, 2006; Sears, Zhang, Wiesner, Hackett, & Yuan, 2013). A limited body of research has recently referred to these forms of AC as virtual assessment centers (henceforth, virtual ACs), namely, several activities aimed at evaluating people for selection or development purposes within organizations (e.g. in-baskets, leaderless group discussions, management games) supported by a wide range of digital platforms and online tools (e.g. online social networks, e-learning systems, immersive technology, gamification, simulation and biodata (Howland, Rembisz,

Wang-Jones, Heise, & Brown, 2015; Lara-Prieto & Niño-Juárez, 2021). Herewith, virtual ACs differ from face-to-face ACs in that they do not require assesses to be physically present in the same location.

Validity of ACs

According to Grimm and Widaman (2012), the embodiment of an instrument's validity is construct validity, which is the tool's general capacity to measure the construct it intends to measure. Deeper down, we may articulate the construct validity along two axes: (1) internal validity (including content-related validity, dimensionality, reliability, and invariance) refers to the internal relationships among the indicators constituting a tool; (2) external validity (including criterion-related, discriminant and convergent validities, along with interpretation and implications of a tool's use), which pertains to the relations between a tool and external criteria.

Assessment centers owe their popularity to their validity. High consensus exists on how the composite score derived from an AC tends to be a valid reflection of the required job criteria, for example, future job performance (Hermelin, Lievens, & Robertson, 2007; Arthur & Day, 2011; Sackett, Shewach, & Keiser, 2017). Identifying ACs' critical dimensions is one of the key guidelines mentioned in the *Guidelines and Ethical Considerations of Assessment Center Operations* (Rupp *et al.*, 2015). This affirms the significance of AC's content validity. Therefore, organizations consistently use content-related validation methods to meet professional requirements (Woehr & Arthur, 2003). Moreover, scholars attribute ACs to notable high criterion-related validity, predictive validity correlations ranging from 0.37 to 0.52 to work-related criteria (Dilchert & Ones, 2009; Rupp, Thornton & Gibbons, 2008; Thornton & Gibbons, 2009), and incremental validity (Dayan, Kasten, & Fox, 2002). This infers that ACs provide distinctive future performance-related insight (Heimann, Ingold, Lievens, Melchers, Keen, & Kleinmann, 2021). Moreover, Heimann *et al.* (2021) showed that ACs were better suited to provide insight into personality traits. However, there has been less promising support for the construct-related validity of ACs (Meriac, Hoffman, Woehr, & Fleisher, 2008).

As a relatively novel format, scholars have not extensively studied virtual ACs for their psychometric properties. However, evidence supports measurement equivalence between traditional and online assessments (Lievens & Harris, 2003). Avni and Luria (2022) compared conventional and virtual ACs and found similar levels of validity and reliability. This is in line with other research that supports the efficacy of digital assessments (Langer, König, & Krause, 2017; Woods, Ahmed, Nikolaou, Costa, & Anderson, 2019). There is also a gap in exploring the criterion-related validity, namely how well a measure directly measures the construct or target behavior (Rosenbaum, 1989) of ACs and their content (e.g. incidents and exercises) (Hoffman, Kennedy, LoPilato, Monahan, & Lance, 2015). Overall, there is room for further research exploring the validity of online assessments like virtual ACs. As we attempted to fill this gap, our first research question was:

RQ1. What indications of construct validity (internal and external validity) do digital ACs offer?

The concept of adverse (disparate) impact

Gender biases exist and have been a source of hindrance to objectivity in employment-related decisions (Heilman & Eagly, 2008). Indeed, data that drives decisions essentially comes from humans. This information source may lead to unfairness. Biases based on stereotypes can affect decision-making tools, leading to different assessment standards for candidates and employees based on their gender. More specifically, they may adversely impact women's chances of being recruited or promoted (Hoyt, 2012). For instance, according to Moscatelli, Menegatti, Ellemers, Mariani, and Rubini (2020), competence is the main criterion against which ACs evaluate men.

In contrast, women must show excellent performance in three criteria, i.e. competence, morality, and sociability to make selection decisions that are in their favor.

We focused on two predominant types of unfairness concerns, i.e. disparate impact and disparate treatment. Adverse impact refers to the diverse rate of selection in recruitment, promotion, and/or other pertinent employment-related decisions that impact a particular group adversely or negatively, which may regard race, religion, gender, or ethnic group (In Outtz, 2010). Adverse impact reflects unintentional discrimination, largely governed by subconscious stereotypes. In contrast, disparate treatment is conscious and intentional discrimination in decision-making directed at employees because of their characteristics (Zafar, Valera, Gomez Rodriguez, & Gummadi, 2017). According to Prati *et al.* (2019), a significant factor influencing adverse impact is the gender of those responsible for evaluating the candidates (i.e. assessors or selectors). Specifically, male selectors tend to favor and recruit female candidates, while female selectors remain neutral. Both male and female selectors appear to be guided by differing standards when it comes to the assessment of male and female applicants. To prevent organizations from being overtly discriminatory and to promote equal employment opportunities, organizations have operationalized adverse impact in the form of selection ratios, commonly referred to as the “four-fifths rule” or “80% rule.” This rule states that the selection rate for any minority group should be at least 80% of the selection rate for the majority group. If the selection rate for the minority group is less than 80% of that of the majority group, this may indicate an adverse impact, suggesting potential discrimination. However, the ultimate indicator of adverse impact is the difference between the concerned groups and whether that difference is statistically significant and justifiable. Given that mixed teams have a higher probability of delivering superior decisions and contextual and task performances (Van de Voorde, Paauwe, & Van Veldhoven, 2012; Paauwe, Guest, & Wright, 2013; Prati *et al.*, 2019), it remains important to gauge whether the organization’s decision-making can discriminate or not, for instance through an adverse impact analysis. In this research, we focused on the employees’ gender and specifically assessed if the methodology utilized for selecting candidates for Group Talent works at the expense of the members of a certain gender.

Therefore, we hypothesized:

RQ2. Does the methodology of digital ACs exhibit any adverse impacts concerning gender differences?

Integrating digital tools into ACs may disproportionately benefit younger assesses. Indeed, younger individuals tend to have higher familiarity and comfort with digital technologies, as they have grown up in an environment rich in such technologies. This familiarity can lead to better performance in selection procedures that utilize digital platforms (Woods, Ahmed, Nikolaou, Costa, & Anderson, 2019). Moreover, younger assesses are often more adaptable to new digital innovations and technological changes. This adaptability can be beneficial in digital selection methods that are continually evolving (Folger, Brosi, Stumpf-Wollersheim, & Welp, 2021). Finally, proficiency in handling digital tools and platforms could also improve the performance of younger assesses. Their ease in navigating and utilizing digital assessment tools can lead to a more accurate representation of their skills and competencies during selection procedures (Newhouse & Njiru, 2009).

Therefore, we hypothesized:

RQ3. Do digital ACs exhibit any differential impacts on assesses across age groups?

Overall, the study aimed to clarify whether we can consider the virtual AC a sound technique for helping organizations reach their goals connected to people evaluation selection and development. Our research questions sought to explore whether virtual ACs evaluate people’s characteristics (internal validity), whether this evaluation is connected with relevant criteria (external validity) (RQ1), and whether virtual ACs are immune from adverse impact implications related to discriminating standards regarding gender (RQ2) or to the discrepant

levels of familiarity between younger and older assesses (RQ3). In the following part, we present the methodology adopted to answer these RQs.

Method

Context

To answer our research questions, we employed a digital AC developed to identify the group talent from an Italian Group of Companies. We standardized the AC procedure and administered it to all the participants uniformly. The exercises and assessors remained unchanged for all the participants to the AC except for one particular exercise, namely in-basket or business case presentation, between the two participants' levels (see 3.3. section). The final steps consisted of scoring and analysis between the assessors and the relevant organization members.

Sample

The sample consisted of 120 employees from an Italian Group of Companies. About 55.8% of the participants were men and 43.3% were women. Of them, 91 (75.8%) were managers, while 29 (24.2%) were senior managers. From the Italian Group of Companies, 16 participants belonged to Company 1 and 2 each, 4 participants were from Company 3, 13 were from Company 4, 4 participants belonged to Company 5, while Company 6 had 24, which was the greatest number of participants, 5 participants belonged to Company 7, and lastly, 24 participants were from the Corporate. The participants' ages ranged from 28 years to 51 years, with an average age of 37.85 years (SD = 5.13 years). Moreover, there was a total of 15 lead assessors. Of them, 11 (73.3 %) were women, and 4 (26.7%) were men.

Procedure for the virtual AC

We selected the participants from each company's internal human resources department in collaboration with the direct supervisor responsible for selecting the best performer. The selection criteria for the candidates partaking in the AC depended on two factors, i.e. fluency in English (minimum B2 level) and compliance with the Company values. The AC was an individually administered procedure conducted on digital platforms in English. It consisted of several exercises, namely the BEI interview, in-basket or business case presentation, and direct report meeting. We conducted them according to a standardized schedule, and they took about 4.5 hours to complete. We designed the framework around eight competencies derived from the Company's Soft Skills Model, structured as follows:

- (1) Execution: the application of organizational principles, knowledge, and expert skills to deal with critical issues and achieve results;
- (2) Proactivity: the ability to anticipate business trends and act accordingly;
- (3) Decision-making: the ability to make informed choices with calculated risks after context assessment (organizational goals and local market's understanding);
- (4) Adaptability: the capacity to adjust under varying circumstances and changing goals;
- (5) People management: the ability to create a learning environment, adopt effective communication, and customize content for an engaged workforce;
- (6) Teamwork: the ability to share and communicate knowledge for high-quality performance with colleagues;
- (7) Innovation: the ability to implement ideas backed by expertise and insights;
- (8) Openness and inclusion: the ability to include every member by ensuring open access to information and utilizing efficient communication.

The Lead Assessor initiated the AC with a presentation on Microsoft Teams and showed the agenda. Three stages followed after the meeting.

BEI interview: It consisted of a 75-min interview with the lead assessor. During the interview, the participants completed the Hogan personality test and cognitive test Matrigma or HRBI along with a Word document that included their profile summaries, including their education level and professional experience. We also generated insight into the candidates' future aspirations to gauge motivation in the candidate.

In-basket/business case presentation: This stage presented an individual problem-solving exercise. It consisted of different levels of complexity in terms of the business cases presented to the participants according to their manager or senior manager role. The participants had 90 minutes to solve the business case individually and 30 minutes to present to the Lead Assessor without any discussion.

Direct report meeting: It consisted of an interactive exercise and was uniform for both managerial groups. It comprised a business simulation in which the participants faced a real-life situation experienced by a particular employee. The participants had a 40-min time limit, with 20 minutes to assess the situation and 20 minutes to discuss the situation with the second assessor.

Following the administration, we utilized the BARS method to score each exercise. We collected the scoring on an integration grid that we also utilized for the integration of results and its subsequent scoring on the algorithm that we prepared to tabulate four score levels: a total competency score, which was the average of the candidate's score in each competency, a potential score which determined the potential level, an international attitude score which determined the extent to which a candidate was keen in being a part of different cultures, and global opportunities. We measured it by specific questions during the BEI interview. Those questions were designed to gauge past international experiences, the person's ability to coherently elaborate on those experiences and their value, and the willingness to continue to invest in their global expertise. Lastly, we gauged the career aspiration score during the BEI interview to assess the level of ambition a candidate has regarding their professional growth. Together, all these scores determined the status validation of the participant as a Group Talent.

Furthermore, determining a participant's designation as Group Talent entailed three distinct validation levels during its final phase. Initially, we held a meeting between the company's human resources business partner and the lead assessors to review and integrate the assessment scores with historical data on the applicant's performance. In the subsequent stage, we gathered the human resources business partner, lead assessors, and the participants' direct supervisors, where we further scrutinized the results from the first meeting and combined them with the supervisors' assessments. We made the ultimate decision in a board-level meeting attended by senior managers, senior HR executives (C-level), and a director, focusing on promotion decisions. Three key factors informed this comprehensive evaluation process, i.e. participants' professional histories, their AC scores, and the evaluations provided by their supervisors. Thus, a holistic review resulted in the final identification of participants as Group Talent. However, despite being comprehensive, the AC's validity criterion was not entirely independent since one-third relied on the AC scores.

Statistical analysis

We subjected the data obtained from the digital AC to a series of statistical analyses using IBM SPSS Statistics 27. We calculated the means, standard deviations, and explorative analyses to gauge the data distribution. To address the RQ1, we analyzed the following elements as indexes of the internal validity: (1) the dimensionality through the correlation between the competencies assessed during the virtual AC and through a principal component analysis; (2) reliability with Cronbach's alpha and McDonald's omega. Furthermore, we assessed external validity with criterion-related validity based on a binary logistic regression and sensitivity analysis (Grimm & Widaman, 2012).

To address RQ2, we calculated the Chi-squared test and the four-fifth rule to get an insight into the fairness of the digital AC as a selection tool specifically related to gender and to assess adverse impact. Finally, we adopted MANOVA to analyze the impact of gender (participant and assessor) in response to RQ2 and age in response to RQ3.

Results

RQ1: Evidence of construct validity

The first research question pertained to AC’s validity in its virtual format. Table 1 reports the mean and standard deviations of the competencies. The mean scores generally reside in the middle point of the scale, which is in line with the normal distribution. The standard deviations were medium level. The skewness values suggest distributions that are very close to symmetry. They were almost evenly balanced on both sides of the mean, with only subtle tendencies toward positive or negative skew. The kurtosis values in the range of 0.8 to -0.41 indicated a moderately platykurtic distribution.

Regarding internal validity, Table 2 shows the correlations between the competencies, specifically highlighting the highest and lowest correlations. Every competence was significantly correlated with each other, which implies that the dimensions fit to be a part of the virtual AC. Execution exhibited the strongest significant positive correlation with adaptability ($r = 0.48, p < 0.05$) and had the weakest yet significant positive correlation

Table 1. Descriptive statistics and Shapiro–Wilk test results for competency ratings (scale 1–5)

	Mean	SD	W	Df	p	Min	Max	Skewness	Kurtosis
Execution	3.23	0.60	0.97	120	0.01	1.70	4.50	-0.43	-0.26
Proactivity	3.33	0.60	0.98	120	0.13	1.50	4.60	-0.39	0.91
Decision-making	3.05	0.56	0.98	120	0.23	1.80	4.40	-0.07	-0.48
Adaptability	3.17	0.46	0.97	120	0.03	1.70	4.20	-0.48	0.54
People management	3.04	0.50	0.98	120	0.38	1.80	4.20	-0.02	-0.28
Teamwork	3.08	0.46	0.98	120	0.27	1.90	4.40	0.12	0.33
Innovation	3.04	0.48	0.98	120	0.08	1.80	4.00	-0.22	-0.41
Openness and inclusion	2.99	0.50	0.98	120	0.07	1.60	4.30	-0.01	0.35
Competency Total	3.09	0.38	0.97	120	0.03	1.90	4.10	-0.30	0.78

Note(s): SD = standard deviation; W = Shapiro–Wilk test statistic; Df = degrees of freedom and p = significance level (p-value) of the Shapiro–Wilk test
Source(s): Authors’ own elaboration

Table 2. Correlations between competencies

	The lowest	The highest	Mean
1. Execution	People management (0.21*)	Adaptability (0.48**)	0.49**
2. Proactivity	Decision-making (0.35**)	Adaptability (0.48**) Innovation (0.48**)	0.50**
3. Decision-making	People management (0.21*)	Execution (0.44**)	0.46**
4. Adaptability	Decision-making (0.41**)	Teamwork (0.60**)	0.57**
5. People management	Execution (0.21*)	Openness and inclusion (0.61**)	0.50**
6. Teamwork	Proactivity (0.36**)	Openness and inclusion (0.67**)	0.59**
7. Innovation	People management (0.38**)	Teamwork (0.54**)	0.54**
8. Openness and inclusion	Execution (0.38**)	Teamwork (0.67**)	0.57**

Note(s): Table 2 shows Pearson correlations between the competencies measured in the digital AC. More specifically, it displays the highest and lowest correlation among the competencies. * $p < 0.01$ and ** $p < 0.05$
Source(s): Authors’ own elaboration

($r = 0.21, p < 0.01$) with people management. Proactivity had the strongest positive correlations with adaptability ($r = 0.48, p < 0.05$) and innovation ($r = 0.48, p < 0.05$) and the weakest yet significant and positive correlation with decision-making ($r = 0.35, p < 0.05$). Decision-making showed the strongest and significantly positive correlation with execution ($r = 0.44, p < 0.05$) and a significantly positive but weakest correlation with people management ($r = 0.21, p < 0.05$). Adaptability exhibited the strongest significant positive correlation with teamwork ($r = 0.60, p < 0.05$) and exhibited the weakest yet significant positive correlation with decision-making ($r = 0.41, p < 0.05$). People management had the strongest significant positive correlation with openness and inclusion ($r = 0.61, p < 0.05$), whereas it had a significant yet weakest correlation with execution ($r = 0.21, p < 0.01$). Teamwork had the strongest significant positive correlation with openness and inclusion ($r = 0.67, p < 0.05$) and the weakest yet significant positive correlation with proactivity ($r = 0.36, p < 0.05$). Innovation had the strongest and most significant correlation with teamwork ($r = 0.54, p < 0.05$) and the weakest yet significant positive correlation with people management ($r = 0.38, p < 0.05$). Lastly, openness and inclusion exhibited the strongest significant positive correlation with teamwork ($r = 0.67, p < 0.05$) and the weakest yet significant positive correlation with execution ($r = 0.38, p < 0.05$).

Based on the eigenvalue greater than 1.0 rule, the scree plot, and the parallel analysis, the results of the principal component analysis showed the presence of one principal component. This component explains a significant 51.1% of the variance, which suggests that it captures a major underlying factor influencing all these attributes. Moreover, Bartlett's test of sphericity was significant ($\chi^2(28) < 0.001$), indicating that the variables were related and suitable for PCA. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was high (0.859) and thus confirmed that the data was suitable for PCA. The measure of sampling adequacy (MSA) for individual items ranged from 0.832 to 0.903, which is considered good. The PCA results demonstrate that the component had significant loadings, ranging from 0.61 for decision-making to 0.81 for teamwork, indicating that these attributes contribute meaningfully to the primary component (Table 3).

The results support a unidimensional model, allowing competencies to be summarized into a single overall rating score. The reliability coefficient was 0.86, based on McDonald's omega, and 0.85, based on Cronbach's alpha.

Regarding external validity, we utilized logistic regression to analyze the relationship between the assessed competencies and being a member of the Group of Talent. Table 4 displays the results of the binary logistic regression with the eight competencies. The analysis indicated that of all the competencies, execution ($B = 1.28, p = 0.03$) and adaptability ($B = 2.00, p = 0.02$) were the best determinants of status validation. A second logistic regression showed that the total competency score was the best predictor of the criteria ($B = 9.13, SE = 1.68, p < 0.001$).

Table 3. Principal component analysis

	Component 1
Teamwork	0.81
Openness and inclusion	0.78
Adaptability	0.77
Innovation	0.73
People management	0.67
Proactivity	0.67
Execution	0.65
Decision-making	0.61

Source(s): Authors' own elaboration

Table 4. Binary logistic regression results: relationship between assessed competencies and membership in the group of talent

	B	SE.	df	p
Execution	1.28	0.60	1	0.03
Proactivity	1.00	0.61	1	0.10
Decision-making	0.15	0.57	1	0.78
Adaptability	2.00	0.87	1	0.02
People management	1.11	0.78	1	0.15
Teamwork	1.16	0.99	1	0.24
Innovation	1.07	0.77	1	0.16
Openness and inclusion	0.95	0.88	1	0.28

Note(s): B = regression coefficient; SE = standard error; df = degrees of freedom and $p = p$ -value

Source(s): Authors' own elaboration

As [Table 5](#) shows, the total competency score accurately predicted that 64 participants would attain Group Talent status and 40 would not. The overall accuracy of these predictions was 87%. Specifically, the model was more effective at correctly identifying those who would achieve Group Talent status, with an accuracy of 89%, compared to an 83.0% accuracy rate for correctly predicting those who would not attain this status. However, the model generated eight false negatives, erroneously predicting that these candidates would not be identified as Group Talent when, in fact, they did achieve this status. Similarly, there were eight false positives, where the model incorrectly identified candidates as potential Group Talent who ultimately did not achieve this designation. We computed the sensitivity analysis by dividing the level of potential and total competency score into high and low and comparing it with Group Talent status. The sensitivity analysis showed a value of 100% as a true positive rate and 15% as a false positive rate.

RQ2: Fairness in gender

For research questions 2 and 3 that gauge the fairness and adverse impact related to gender, the first step was to conduct a chi-squared test to ascertain whether the status validation probability differed between the male and female participants of the virtual AC. [Table 6](#) displays that 52 females and 67 males attained status validation as GT. Results did not show a significant interaction between gender and status validation, $X^2(1) = 0.306, p = 0.403$.

[Table 8](#) presents the results of the MANOVA analysis conducted to examine the effects of participants' gender on various competencies. [Table 7](#) reveals that gender did not have a significant impact on any of the competencies assessed. The overall effect of gender on the combined competencies was not significant, $F(16, 222) = 0.614, p = 0.874$; Wilks' lambda = 0.917. This indicates that the participants' gender did not play a significant role in determining their competencies.

We conducted MANOVA to analyze the interaction between the genders of the assessees and the assessors with respect to competencies. The analysis revealed no significant

Table 5. Logistic regression results: status validation prediction of the total competency score

Observed	Predicted Group talent		Percentage correct
	No	Yes	
No	40	8	83
Yes	8	64	89
Overall percentage			87.0

Source(s): Authors' own elaboration

Table 6. Gender and age respect to status validation as GT

Gender	Status validation as GT		Total
	No	Yes	
F	22	30	52
M	25	42	67
Total	47	72	119
<i>Age</i>			
28–37	21	42	63
38–51	25	30	55
Total	46	72	118

Source(s): Authors' own elaboration**Table 7.** MANOVA results: effect of gender on competency assessments

	F	df	p
Execution	0.012	1	0.913
Proactivity	0.060	1	0.807
Decision-making	1.508	1	0.222
Adaptability	1.935	1	0.167
People management	0.609	1	0.437
Teamwork	1.634	1	0.204
Innovation	0.192	1	0.662
Openness and inclusion	0.571	1	0.451

Source(s): Authors' own elaboration**Table 8.** MANOVA results: effect of genders, participants and assessors on competency assessments

	F	df	p
Execution	0.032	1	0.859
Proactivity	0.016	1	0.900
Decision-making	0.059	1	0.809
Adaptability	0.149	1	0.700
People management	0.007	1	0.935
Teamwork	0.449	1	0.504
Innovation	0.003	1	0.954
Openness and inclusion	0.457	1	0.501

Source(s): Authors' own elaboration

interaction, $F(2, 120) = 1.259$, $p = 0.226$; Wilks' lambda = 0.834, between the genders of both parties and the competencies assessed (Table 8).

We compared the selection rates for men and women to assess the bias possibility in selecting candidates for the Talent Group, with 62.6% for men and 57.6% for women. An impact ratio analysis resulting in 92% indicated no adverse impact according to the 80% rule, as the selection rate for the less-favored group (women) was not less than 80% of the most-favored group (males), thus suggesting a fair selection process based on the virtual AC too.

RQ3: Fairness in age

Results did not show a significant association between age and status validation, $X^2(1, 118) = 1.814$, $p = 0.178$.

The MANOVA analysis indicated that age did not significantly impact any of the competencies assessed. As shown in [Table 9](#), the overall effect of age on the competencies was not significant, $F(8, 109) = 1.277, p = 0.263$; Wilks' lambda = 0.914. This suggests that the age of the participants did not play a significant role in determining the competencies assessment.

As a final control, we adopted an ANOVA to test if the combination of gender and age of participants could impact the total competency score. The ANOVA results indicated that there were no statistically significant main effects of gender ($F(1, 114) = 0.00953, p = 0.922, \eta^2 = 0.000$) or age ($F(1, 114) = 0.05275, p = 0.819, \eta^2 = 0.000$) on total competency score. The interaction between gender and age also did not reach statistical significance ($F(1, 114) = 1.48027, p = 0.226, \eta^2 = 0.013$). This suggests that neither the gender, participants' age nor the combination of these two factors had a significant impact on their total competency scores within the analyzed sample.

Discussion

We conducted the study on a cohort of Italian organizations that had implemented a virtual AC for evaluating 120 employees. The primary aim was to explore the efficacy of the AC in its virtual incarnation, focusing on its validity and fairness as a selection tool. Three fundamental research questions drove the research. First, we sought to determine whether the virtual AC displayed general indications of validity as a selection mechanism. Second, we investigated whether the virtual AC was an appropriate tool for selecting candidates across genders. Lastly, the research aimed to ascertain whether the virtual ACs exhibited any differential impacts on assesses belonging to various age groups.

The findings showed preliminary evidence of general indications of virtual AC's validity. Regarding internal validity, the findings supported a score based on a unidimensional index and showed good reliability. The results inherent in external validity also showed good and promising validity indications. In summary, the overall rating showed that it could predict the validation process results with a relevant sensitivity. The results we obtained about the validity of the virtual ACs allowed us to align it with conventional ACs ([Eurich, Krause, Cigularov, & Thornton, 2009](#)) and the existent literature that confirms the validity of the online selection procedures ([Groene, Knorr, Vogel, Hild, & Hampe, 2022](#)).

We also addressed the lack of research on the adverse impact and fairness of digital selection procedures. Results suggest fairness in terms of gender. We found no evidence of adverse impact. This sets up a good robustness signal for digital selection procedures, and virtual ACs specifically, boding well for their use. This finding is in contrast to [Lievens, Schollaert, and Keen \(2015\)](#) research that reported gender subgroup differences when studying digital situational judgment tests in police candidates, which is a predominantly men-dominated profession.

Table 9. MANOVA results: effect of age on the competencies

	F	df	p
Execution	0.119	1	0.730
Proactivity	1.666	1	0.199
Decision-making	3.034	1	0.084
Adaptability	0.012	1	0.913
People management	0.211	1	0.647
Teamwork	0.081	1	0.777
Innovation	0.205	1	0.652
Openness and inclusion	0.736	1	0.393

Source(s): Authors' own elaboration

Finally, the participants' age did not influence the competencies assessment. This result contrasts with a study on a digital assessment performed with computer simulation games that found a better performance for men than for women and younger than for older applicants (Melchers & Basch, 2022). The results may vary because adopting these technologies requires skills equally present across age groups in the roles analyzed. Alternatively, the hypothetical advantages of younger participants – such as familiarity with digital technologies, adaptability to innovations, and technological proficiency – may have been offset by the job experience of older participants. Regardless, nowadays, people may perceive virtual ACs as easy to use and useful for participants, as shown by Pattnaik and Padhi (2021).

Theoretical and practical implications

The findings have both theoretical and practical implications. The study revealed general indications of validity and fairness for both genders. Through this, we can theoretically imply that a well-designed and valid virtual AC reduces the chances of adverse impact. Researchers have widely established the importance of conventional ACs by demonstrating their validity, reliability, and overall significance (Woehr & Arthur, 2003; Rupp, Thornton, & Gibbons, 2008; Thornton & Gibbons, 2009; Dilchert & Ones, 2009). However, despite living in the industrial era where digitalization has become widespread, literature on virtual ACs is scarce. Thereby, this research contributes to understanding how a virtual assessment works and is an important starting point for future research on AC's digitalization. The study provides initial validity evidence on a fair virtual AC, because it had no adverse impact when both the participants and the assessors considered gender. The pioneering nature of the research serves as its strength. Furthermore, the present research not only studies the adverse impact of virtual ACs on personnel selection decisions but also adds to the existing evidence of how digitalization can reduce bias, especially in the presence of human assessors (Aichi, Bassiri, Benmokhtar, & Belouaad, 2018).

A key practical benefit of an AC with no adverse impact and fair assessment is the ability to promote equal opportunity by integrating diversity management. The fairness of virtual AC reinforces the need to assess adverse impacts and use inclusive assessment practices. This can translate into how future virtual ACs can consist of a more diverse group of candidates and assessors. Including a pool of assessors from a diverse range of backgrounds ensures a thoroughly comprehensive and potentially fair evaluation process. The composition of a diverse assessor panel is a strong message to the candidates on the organization's stance on diversity and its commitment to working on equal opportunities.

From the organizational point of view, to gain a more thorough insight, the organization can keep track of the employee's performance after the AC. This would have reinforcing benefits since the post-assessment data would enable the organization to measure the AC's efficacy. Although the research shows positive results regarding the fairness of virtual ACs, it is important to pay close attention to these issues consistently. When training assessors for virtual ACs, we must emphasize several key aspects to ensure fairness and accuracy. Assessors should receive training in minimizing potential biases, particularly those arising from video-based assessments, as previous research (e.g. Van Iddekinge, Raymark, Roth, & Payne, 2006; Sears, Zhang, Wiesner, Hackett, & Yuan, 2013) indicates that ratings in digital formats often differ from face-to-face evaluations. Training should also focus on maintaining consistency in applying evaluation criteria across different platforms to ensure the validity of virtual ACs. Moreover, assessors should develop skills to interpret subtle non-verbal cues, which may be less apparent in digital interactions, and mitigate any potential adverse impacts related to demographic factors such as age and gender.

Limitations and future research recommendations

Readers should view the study and its findings while remembering its limitations. The primary limitation of the present research is the lack of generalizability because we based it on a small

sample size derived from a single organizational case. To overcome the problem of lack of generalizability, future research should have access to the data of multiple virtual ACs and a larger sample size.

The second limitation was the absence of a job performance variable. Since we based this research on an organizational case, we deemed a significant portion of the data unusable to protect the company's confidentiality. With the participants' performance data, this research would have claimed stronger validity evidence (e.g. criterion validity). Future research should include variables that measure performance after the administration of the virtual AC to completely gauge the methodology's efficacy.

As highlighted in previous research, the third limitation is the need for a more structured comparison between the validity of virtual AC ratings and traditional face-to-face ACs. Studies, including [Sears, Zhang, Wiesner, Hackett, and Yuan \(2013\)](#) and [Van Iddekinge, Raymark, Roth, and Payne \(2006\)](#), suggest that ratings for digital interviews, particularly video-based ones, are slightly lower than those for face-to-face interactions but still maintain validity.

The fourth limitation of this study lies in the criteria for inclusion in the Group Talent, which was not entirely independent of the scores obtained from the virtual AC. However, these scores constitute only one-third of the evaluation sources. Furthermore, we divided decisions regarding Group Talent status from a comprehensive process encompassing various phases involving multiple roles of the human resource department and the line.

Fifth, our study analyzed the validity of the virtual AC by focusing on the total competency score and overall evaluation. The current scientific debate, as discussed by [Dewberry \(2024\)](#) and [Bosneag and Iliescu \(2024\)](#), revolves around whether ACs should rely on dimension-based or exercise-based indicators. This opens a potential area for further exploration in virtual ACs, where scholars could assess the efficacy of both approaches in digital environments to ensure robust and fair selection practices.

Future research into applicant reactions could provide valuable insights into virtual AC. These reactions could be positive or negative, influenced by various factors. Scholars have found digital assessments to increasingly appeal to candidates ([Folger, Brosi, Stumpf-Wollersheim, & Welpel, 2021](#)). However, others see them as lacking in procedural justice compared to traditional methods in certain situations ([Balcerak & Woźniak, 2021](#)). Moreover, there is a degree of skepticism about them from recruiters' perspectives ([Basch & Melchers, 2021](#)). This underscores the importance of considering applicant reactions in evaluating the usefulness and relevance of assessments.

This complex landscape calls for a multi-faceted exploration of how people perceive digital assessments and how these perceptions influence the selection process's fairness and efficacy. Importantly, integrating the findings of [Baczynska, Skoczeń, Thornton, and Chen \(2024\)](#) on age's moderating effects offers a valuable lens through which scholars may examine the interaction between individual differences and digital assessment outcomes. This approach broadens our understanding of digital ACs' utility and applicability and enriches the discourse on personalizing assessment methodologies to accommodate diverse applicant profiles.

Conclusion

Conventional ACs are valuable and effective but scholars have associated them with practical and logistical problems and biased judgments ([Lievens & Thornton, 2017](#)), which research could address by digitalizing them and conducting them using virtual ACs. Our research presents preliminary evidence of validity and fairness for online ACs. However, it is important for future studies to (1) explore multiple virtual ACs, (2) compare virtual and conventional ACs to gauge their effectiveness, and (3) find ways to assess fairness in more subgroups. The research evidence contributes to the knowledge of practitioners and researchers, and the article serves as a stepping stone for further exploration and understanding. Our results affect the policy field because virtual ACs promote diversity and inclusion based on gender and age.

Moreover, virtual ACs can eliminate geographic and physical barriers and thus enable organizations to attract more diverse candidates. Hence, these findings align with broader discussions on the legal and regulatory environment for employee selection across different jurisdictions, which emphasize the need for fair and inclusive selection methods (Sackett *et al.*, 2010). Finally, the implications regard the economic field, too, because virtual ACs can offer significant cost savings by reducing expenses associated with travel, venue rental, and logistical arrangements.

References

- Aichi, Y., Bassiri, M., Benmokhtar, S., & Belouaad, S. (2018). The digital assessment center and the evaluation of soft skills. *Asia Life Sciences*, 1, 35–42.
- Ansbacher, H. L. (1941). German military psychology. *Psychological Bulletin*, 38(6), 370–392. doi: [10.1037/h0056263](https://doi.org/10.1037/h0056263).
- Armoneit, C., Schuler, H., & Hell, B. (2020). Use, validity, practicality, and acceptance of personnel selection methods in Germany 1985, 1993, 2007, 2020: The continuation of a trend study. *German Journal of Work and Organizational Psychology*, 64(2), 67–82. doi: [10.1026/0932-4089/a000311](https://doi.org/10.1026/0932-4089/a000311).
- Arthur, W., & Day, E. A. (2011). Assessment centers. in S. Zedeck (Ed.), *APA handbook of industrial and organizational psychology, volume 2: selecting and developing members for the organization* (pp. 205–235). American Psychological Association. doi: [10.1037/12170-007](https://doi.org/10.1037/12170-007).
- Arthur, W. A., Woehr, D. J., & Maldegen, R. (2000). Convergent and discriminant validity of assessment center dimensions: a conceptual and empirical re-examination of the assessment center construct-related validity paradox. *Journal of Management*, 26(4), 813–835. doi: [10.1016/S0149-2063\(00\)00057-X](https://doi.org/10.1016/S0149-2063(00)00057-X).
- Avni, E., & Luria, G. (2022). Validity and reliability of virtual versus face-to-face assessment center. *Academy of Management Proceedings*, 2022(1). doi: [10.5465/ambpp.2022.14415abstract](https://doi.org/10.5465/ambpp.2022.14415abstract).
- Baczynska, A., Skoczeń, I., Thornton, G., & Chen, S. (2024). Personality metatraits and managerial AC dimensions in assessment center performance: the moderating effect of age in strong and weak VUCA simulations. *Central European Management Journal*, 32(2), 179–198. doi: [10.1108/CEMJ-06-2023-0274](https://doi.org/10.1108/CEMJ-06-2023-0274).
- Balcerak, A., & Woźniak, J. (2021). Reactions to some ICT-based personnel selection tools. *Economics & Sociology*, 14(1), 214–231. doi: [10.14254/2071-789x.2021/14-1/14](https://doi.org/10.14254/2071-789x.2021/14-1/14).
- Basch, J. M., & Melchers, K. G. (2021). The use of technology-mediated interviews and their perception from the organisation's point of view. *International Journal of Selection and Assessment*, 29(3–4), 495–502. doi: [10.1111/ijsa.12339](https://doi.org/10.1111/ijsa.12339).
- Bosneag, I., & Iliescu, D. (2024). Dimension- or task-based assessment centers? A direct comparison study of two measurement approaches. *Psihologia Resurselor Umane*, 22(1). doi: [10.24837/pru.v22i1.545](https://doi.org/10.24837/pru.v22i1.545).
- Dayan, K., Kasten, R., & Fox, S. (2002). Entry-level police candidate assessment center: an efficient tool or a hammer to kill a fly?. *Personnel Psychology*, 55(4), 827–849. doi: [10.1111/j.1744-6570.2002.tb00131.x](https://doi.org/10.1111/j.1744-6570.2002.tb00131.x).
- Dewberry, C. (2024). Assessment centers do not measure competencies: why this is now beyond reasonable doubt. *Industrial and Organizational Psychology*, 17(2), 154–175. doi: [10.1017/iop.2024.5](https://doi.org/10.1017/iop.2024.5).
- Dilchert, S., & Ones, D. S. (2009). Assessment center dimensions: individual differences correlates and meta-analytic incremental validity. *International Journal of Selection and Assessment*, 17(3), 254–270. doi: [10.1111/j.1468-2389.2009.00468.x](https://doi.org/10.1111/j.1468-2389.2009.00468.x).
- Eurich, T. L., Krause, D. E., Cigularov, K., & Thornton, G. C. III (2009). Assessment centers: current practices in the United States. *Journal of Business and Psychology*, 24(4), 387–407. doi: [10.1007/s10869-009-9123-3](https://doi.org/10.1007/s10869-009-9123-3).
- Folger, N., Brosi, P., Stumpf-Wollersheim, J., & Welpel, I. M. (2021). Applicant reactions to digital selection methods: a signaling perspective on innovativeness and procedural justice. *Journal of Business and Psychology*, 37(4), 735–757. doi: [10.1007/s10869-021-09770-3](https://doi.org/10.1007/s10869-021-09770-3).

- Gaugler, B. B., Rosenthal, D. B., Thornton, G. C. III, & Bentson, C. (1987). Meta-analysis of assessment center validity. *Journal of Applied Psychology*, 72(3), 243–259. doi: [10.1037/0021-9010.72.3.493](https://doi.org/10.1037/0021-9010.72.3.493).
- Grimm, K. J., & Widaman, K. F. (2012). Construct validity. in H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds), *APA handbook of research methods in psychology, vol. 1. Foundations, planning, measures, and psychometrics*. American Psychological Association, 621–642. doi: [10.1037/13619-033](https://doi.org/10.1037/13619-033).
- Groene, O. R., Knorr, M., Vogel, D., Hild, C., & Hampe, W. (2022). Reliability and validity of new online selection tests for midwifery students. *Midwifery*, 106, 103245. doi: [10.1016/j.midw.2021.103245](https://doi.org/10.1016/j.midw.2021.103245).
- Hausknecht, J. P., & Heavey, A. L. (2017). Selection for service and sales jobs. in J. L. Farr, & N. T. Tippins (Eds), *Handbook of employee selection*. Routledge. 2nd ed., 781–96. New York. doi: [10.4324/9781315690193-35](https://doi.org/10.4324/9781315690193-35).
- Heilman, M. E., & Eagly, A. H. (2008). Gender stereotypes are alive, well, and busy producing workplace discrimination. *Industrial and Organisational Psychology*, 1(4), 393–398. doi: [10.1111/j.1754-9434.2008.00072.x](https://doi.org/10.1111/j.1754-9434.2008.00072.x).
- Heimann, A. L., Ingold, P. V., Lievens, F., Melchers, K. G., Keen, G., & Kleinmann, M. (2021). Actions define a character: assessment centers as behavior-focused personality measures. *Personnel Psychology*, 75(3), 675–705. doi: [10.1111/peps.12478](https://doi.org/10.1111/peps.12478).
- Hermelin, E., Lievens, F., & Robertson, I. T. (2007). The validity of assessment centres for the prediction of supervisory performance ratings: a meta-analysis. *International Journal of Selection and Assessment*, 15(4), 405–411. doi: [10.1111/j.1468-2389.2007.00399.x](https://doi.org/10.1111/j.1468-2389.2007.00399.x).
- Hoffman, B. J., Kennedy, C. L., LoPilato, A. C., Monahan, E. L., & Lance, C. E. (2015). A review of the content, criterion-related, and construct-related validity of assessment center exercises. *Journal of Applied Psychology*, 100(4), 1143–1168. doi: [10.1037/a0038707](https://doi.org/10.1037/a0038707).
- Howland, A. C., Rembisz, R., Wang-Jones, T. S., Heise, S. R., & Brown, S. (2015). Developing a virtual assessment center. *Consulting Psychology Journal: Practice and Research*, 67(2), 110–126. doi: [10.1037/cpb0000034](https://doi.org/10.1037/cpb0000034).
- Hoyt, C. L. (2012). Gender bias in employment contexts: a closer examination of the role incongruity principle. *Journal of Experimental Social Psychology*, 48(1), 86–96. doi: [10.1016/j.jesp.2011.08.004](https://doi.org/10.1016/j.jesp.2011.08.004).
- Iles, P. (1992). Centres of excellence? Assessment and development centres, managerial competencies and human resource strategies. *British Journal of Management*, 3(2), 79–90. doi: [10.1111/j.1467-8551.1992.tb00037.x](https://doi.org/10.1111/j.1467-8551.1992.tb00037.x).
- Langer, M., König, C. J., & Krause, K. (2017). Examining digital interviews for personnel selection: applicant reactions and interviewer ratings. *International Journal of Selection and Assessment*, 25(4), 371–382. doi: [10.1111/ijsa.12191](https://doi.org/10.1111/ijsa.12191).
- Lara-Prieto, V., & Niño-Juárez, E. (2021). Assessment center for senior engineering students: in-person and virtual approaches. *Computers & Electrical Engineering*, 93, 107273. doi: [10.1016/j.compeleceng.2021.107273](https://doi.org/10.1016/j.compeleceng.2021.107273).
- Lievens, F., & Harris, M. M. (2003). Research on internet recruitment and testing: current status and future directions. in C. L. Cooper, & I. T. Robertson (Eds), *International review of industrial and organisational psychology*, 18, 131–165. Chichester. Wiley. doi: [10.1002/0470013346.ch4](https://doi.org/10.1002/0470013346.ch4).
- Lievens, F., & Thornton, G. C. (2017). Assessment centers: recent developments in practice and research. *The blackwell handbook of personnel selection*, 243–264. doi: [10.1002/9781405164221.ch11](https://doi.org/10.1002/9781405164221.ch11).
- Lievens, F., Schollaert, E., & Keen, G. (2015). The interplay of elicitation and evaluation of trait-expressive behavior: Evidence in assessment center exercises. *Journal of Applied Psychology*, 100(4), 1169–1188. doi: [10.1037/apl0000004](https://doi.org/10.1037/apl0000004).
- Melchers, K. G., & Basch, J. M. (2022). Fair play? Sex-, age-, and job-related correlates of performance in a computer-based simulation game. *International Journal of Selection and Assessment*, 30(1), 48–61. doi: [10.1111/ijsa.12337](https://doi.org/10.1111/ijsa.12337).

- Meriac, J. P., Hoffman, B. J., Woehr, D. J., & Fleisher, M. S. (2008). Evidence of the validity of assessment center dimensions: analysis of the incremental criterion-related validity of dimension ratings. *Journal of Applied Psychology, 93*(5), 1042–1052. doi: [10.1037/0021-9010.93.5.1042](https://doi.org/10.1037/0021-9010.93.5.1042), PMID:18808224.
- Moscatelli, S., Menegatti, M., Ellemers, N., Mariani, M. G., & Rubini, M. (2020). Men should be competent, women should have it all: multiple criteria in the evaluation of female job candidates. *Sex Roles, 83*(5-6), 269–288. doi: [10.1007/s11199-019-01111-2](https://doi.org/10.1007/s11199-019-01111-2).
- Murphy, M., Callander, J., Dohan, D., & Grandis, J. (2021). Women’s experiences of promotion and tenure in academic medicine and potential implications for gender disparities in career advancement. *JAMA Network Open, 4*(9), e2125843. doi: [10.1001/jamanetworkopen.2021.25843](https://doi.org/10.1001/jamanetworkopen.2021.25843).
- Newhouse, C., & Njiru, J. (2009). Using digital technologies and contemporary psychometrics in the assessment of performance on complex practical tasks. *Technology, Pedagogy and Education, 18* (2), 221–234. doi: [10.1080/14759390902992626](https://doi.org/10.1080/14759390902992626).
- Outz, J. L. (Ed.) (2010). *Adverse impact*. Routledge. doi: [10.4324/9780203848418](https://doi.org/10.4324/9780203848418).
- Pauwwe, J., Guest, D. E., & Wright, P. M. (Eds) (2013). *HRM and performance*. John Wiley and Sons. Chichester.
- Pattnaik, S., & Padhi, M. (2021). Challenges in assessment centres: lessons from experience. *Management and Labour Studies, 46*(3), 313–336. doi: [10.1177/0258042X211002503](https://doi.org/10.1177/0258042X211002503).
- Prati, F., Menegatti, M., Moscatelli, S., Kana Kenfack, C. S., Pireddu, S., Crocetti, E., . . . Rubini, M. (2019). Are mixed-gender committees less biased toward female and male candidates? An investigation of competence-morality- and sociability-related terms in performance appraisal. *Journal of Language and Social Psychology, 38*(5-6), 586–605. doi: [10.1177/0261927x19844808](https://doi.org/10.1177/0261927x19844808).
- Rosenbaum, P. R. (1989). Criterion-related construct validity. *Psychometrika, 54*(4), 625–633. doi: [10.1007/bf02296400](https://doi.org/10.1007/bf02296400).
- Rupp, D. E., Thornton, C. G., & Gibbons, A. M. (2008). The construct validity of the assessment centre method and usefulness of dimensions as focal constructs. *Industrial and Organisational Psychology: Perspectives on Science and Practice, 1*(1), 116–120. doi: [10.1111/j.1754-9434.2007.00021.x](https://doi.org/10.1111/j.1754-9434.2007.00021.x).
- Rupp, D. E., Hoffman, B. J., Bischof, D., Byham, W., Collins, L., Gibbons, A., . . . Thornton, G. (2015). Guidelines and ethical considerations for assessment center operations. *Journal of Management, 41*(4), 1244–1273. doi: [10.1177/0149206314567780](https://doi.org/10.1177/0149206314567780).
- Sackett, P. R., & Lievens, F. (2008). Personnel selection. *Annual Review of Psychology, 59*(1), 419–450. doi: [10.1146/annurev.psych.59.103006.093716](https://doi.org/10.1146/annurev.psych.59.103006.093716).
- Sackett, P. R., Shen, W., Myers, B., Lievens, F., Schollaert, E., Van Hoye, G., . . . & Aguinis, H. (2010). Perspectives from twenty-two countries on the legal environment for selection. in J. L. Farr, & N. T. Tippins (Eds), *Handbook of employee selection*. Psychology Press. 651–676. doi: [10.4324/9780203809808](https://doi.org/10.4324/9780203809808).
- Sackett, P. R., Shewach, O. R., & Keiser, H. N. (2017). Assessment centers versus cognitive ability tests: Challenging the conventional wisdom on criterion-related validity. *Journal of Applied Psychology, 102*(10), 1435–47. doi: [10.1037/apl0000236](https://doi.org/10.1037/apl0000236).
- Schmitt, N., Gooding, R. Z., Noe, R. A., & Kirsch, M. (1984). Meta analysis of validity studies published between 1964 and 1982 and the investigation of study characteristics. *Personnel Psychology, 27*(3), 407–422. doi: [10.1111/j.1744-6570.1984.tb00519.x](https://doi.org/10.1111/j.1744-6570.1984.tb00519.x).
- Sears, G. J., Zhang, H., Wiesner, W. H., Hackett, R. D., & Yuan, Y. (2013). A comparative assessment of videoconference and face-to-face employment interviews. *Management Decision, 51*(8), 1733–1752. doi: [10.1108/MD-09-2012-0642](https://doi.org/10.1108/MD-09-2012-0642).
- Taft, R. (1948). Use of the group situation observation method in the selection of trainee executives. *Journal of Applied Psychology, 32*(6), 587–594. doi: [10.1037/h0061967](https://doi.org/10.1037/h0061967).
- Templer, K. J., & Lange, S. R. (2008). Internet testing: equivalence between proctored lab and unproctored field conditions. *Computers in Human Behavior, 24*(3), 1216–1228. doi: [10.1016/j.chb.2007.04.006](https://doi.org/10.1016/j.chb.2007.04.006).

- Thornton, G. C., & Gibbons, A. M. (2009). Validity of assessment centres for personnel selection. *Human Resource Management Review*, 19(3), 169–187. doi: [10.1016/j.hrmr.2009.02.002](https://doi.org/10.1016/j.hrmr.2009.02.002).
- Thornton, G. C. III, & Rupp, D. R. (2006). *Assessment centers in human resource management: Strategies for prediction, diagnosis, and development*. Lawrence Erlbaum. Mahwah, NJ. doi: [10.4324/9781410617170](https://doi.org/10.4324/9781410617170).
- Thornton, G. C., III, & Byham, W. C. (1982). *Assessment centers and managerial performance*. New York: Academic Press.
- Van de Voorde, K., Paauwe, J., & Van Veldhoven, M. (2012). Employee well-being and the Hrm-organisational performance relationship: a review of quantitative studies. *International Journal of Management Reviews*, 14(4), 391–407. doi: [10.1111/j.1468-2370.2011.00322.x](https://doi.org/10.1111/j.1468-2370.2011.00322.x).
- Van Iddekinge, C. H., Raymark, P. H., Roth, P. L., & Payne, H. S. (2006). Comparing the psychometric characteristics of ratings of face-to-face and videotaped structured interviews. *International Journal of Selection and Assessment*, 14(4), 347–359. doi: [10.1111/j.1468-2389.2006.00356.x](https://doi.org/10.1111/j.1468-2389.2006.00356.x).
- Vernon, P. E. (1950). The validation of civil service selection board pro-cedures. *Occupational Psychology*, 24, 75–95.
- Woehr, D. J., & Arthur, W. (2003). The construct-related validity of assessment center ratings: a review and meta-analysis of the role of methodological factors. *Journal of Management*, 29(2), 231–258. doi: [10.1177/014920630302900206](https://doi.org/10.1177/014920630302900206).
- Woods, S. A., Ahmed, S., Nikolaou, I., Costa, A. C., & Anderson, N. R. (2019). Personnel selection in the digital age: a review of validity and applicant reactions, and future research challenges. *European Journal of Work & Organizational Psychology*, 29(1), 64–77. doi: [10.1080/1359432x.2019.1681401](https://doi.org/10.1080/1359432x.2019.1681401).
- Zafar, M. B., Valera, I., Gomez Rodriguez, M., & Gummadi, K. P. (2017). Fairness beyond disparate treatment & disparate impact. In *Proceedings of the 26th International Conference on World Wide Web*. doi: [10.1145/3038912.3052660](https://doi.org/10.1145/3038912.3052660).

Corresponding author

Marco Giovanni Mariani can be contacted at: marcogiovanni.mariani@unibo.it