**Virtual Assessment Center Versus Face-to-Face Assessment Center: Validity and Reliability Assessment**

**Abstract**

Technological advances in every aspect of life in the last decade have changed personnel selection processes. This study compares virtual assessment centers (VACs) and face-to-face assessment centers (FTF-ACs) in terms of validity, reliability and psychometrics. It reports two closely-related studies. The first is a field study that included 11,157 candidates participating in an FTF-AC or a VAC in a selection process for military positions. The study compared the assessments of the candidates in these two ACs while focusing on the assessment distributions, reliability, and structural validity. Many similarities between these two ACs: their distributions were similar; the average assessments were similar, or slightly higher in some of the VACs’ dimensions; reliability between the assessors was high in both ACs, and the validity appeared similar. The second study addresses the results of functional assessments by supervisors of 123 organization’s internal participants who performed a VAC in order to examine its validity. The study shows predictive validity for only some of the assessed dimensions. These results point to similarities in psychometric aspects between VACs and FTF-ACs. Recommendations for human resource managers and for further research are discussed.

*Key words:* *Virtual Assessment Center, reliability, validity, virtual simulation, virtual leaderless group discussion*

**Introduction**

Over the last decade, there has been significant progress and development in innovative technologies to assist in organizational selection processes (Langer et al., 2018; Woods et al., 2020). The COVID-19 pandemic, with its social distancing guidelines, has further accelerated the transition to virtual selection processes (Jones & Abdelfattah, 2020; Joshi et al., 2020). Virtual selection makes the process faster and more efficient, gives organizations access to a wider pool of candidates, and, by reducing the barriers of distance, saves on costs and time (Chapman & Rowe, 2001; Chapman & Webster, 2001, 2003; Galen Kroeck & Magnusen, 1997). However, while the pace of development and the adoption of virtual selection tools has been extensive and rapid, scientific research in the area has lagged behind, resulting in a significant gap between actual practice and the theoretical basis for virtual assessment centers (Blacksmith et al., 2016; Woods et al., 2020).

Woods et al. (2020) point out the potential risks of using new selection technologies without a proper understanding of their effects and implications. This has led researchers in the field to call for the urgent development of a distinct literature on selection technologies, on the assumption that technological selection tools are fundamentally different from traditional tools, and, as such, present unique challenges (Chamorro-Premuzic et al., 2016; Tippins, 2015; Woods et al., 2020). An extensive literature review focused on studies performed to reduce this gap in the research (e.g., Woods et al., 2020) found that existing studies examined only some technology-based selection tools (e.g., web-based tests or video-conference interviews [Stone et al., 2013]). Moreover, there seems to be no published research focusing on virtual assessment centers (VACs). In the context of the COVID-19 pandemic especially, it is very likely that the use of video technology for assessment centers (ACs) will expand, as has happened with virtual interviews (Sears et al., 2012) and online call–based focus groups (Stewart & Shamdasani, 2017).

This study, the first to scientifically examine VACs, seeks to close some of the gaps that exist in the literature regarding VACs. The goal is to examine how technology, and the transition to virtual environments affect the distributions, reliability, and validity of ACs. Two related field studies comparing face-to-face assessment centers (FTF-ACs) with parallel virtual assessment centers (VACs) are presented. The first study focuses on the effects of VACs in comparison with FTF-ACs in terms of evaluators’ assessments of candidates. It compares descriptive and statistical data of assessments in a VAC with the assessments in a FTF-AC focusing on reliability between assessors and structural validity. The second study focuses on predictive validity. This was done by examining the correlation between assessments of the candidates in the VAC and their success indicators in their eventual jobs as rated by their supervisors.

**Assessment Center**

The term assessment center, referring to both the process for personnel selection and the physical site in the case of FTF-AC, has been in use for over fifty years and the process is an accepted recruitment method in organizations throughout the world (Kleinmann & Ingold, 2019). The goal of the selection process is to find the most suitable individuals in terms of their potential to contribute to organizational goals and objectives (Stone et al., 2013). At an AC, candidates undergo standard assessments of behaviors in a variety of exercises that simulate work-related situations (e.g., role-plays and group discussion). The uniqueness of the AC is its interactive nature, where there is social communication among the participants in the context of exercises which evoke actual behaviors. This is in contrast to other selection tools, such as questionnaires or interviews, which do not involve actual communication and are, instead, based on the candidate’s self-report (Kleinmann & Ingold, 2019). Assessors at an AC, who have undergone dedicated training, systematically observe candidates, recording their observations and assessments of candidates’ overt behavior in terms of specific dimensions that are predetermined as being of relevance for the target position. The results of the AC are assessments in various dimensions (e.g., topic presentation ability and leadership ability) and an overall assessment of the candidates’ ability to succeed in their future position (International Task Force on Assessment Center Guidelines, 2009; Kleinmann & Ingold, 2019; 2008; Thornton & Gibbons, 2009).

This study compares the assessments of an FTF-AC and a VAC. While the purpose of these two selection tools is the same, the method they use for gathering information is completely different. At traditional FTF-ACs, candidates physically arrive at the selection site and perform various individual and group tasks with other candidates under observation. In a VAC, assessments are based on advanced virtual technology that allows candidates and evaluators to interact without being in the same physical place. The communication between the participants occurs on virtual platforms (for example Zoom or Skype) where all the participants are connected in real time. As part of the assessment, candidates perform group and individual tasks, such as group exercises, topic presentation exercises or role-playing exercises. In a VAC, like an FTF-AC, assessors observe candidates’ performance and assess them according to predefined metrics. The use of the term “virtual assessment center” in this study does not refer to conducting remote tests or video-based interviews but only to group exercises or simulations.

**Video conference–based communication**

Virtual communication based on video conferencing (VC) has been available for over 80 years, but only since the early 21st century has this technology begun to be used practically to link distant interlocutors. Facetime, Google Chat and Skype are just some of the many options available for remote communication using technology (Nehls et al., 2015). A growing number of organizations have begun to use virtual technologies in the field of recruitment and selection of employees, including for interviews that serve as an adjunct or as an alternative to face-to-face interviews (e.g., Vadi et al., 2016). This is to deal with increasing pressure to expand recruitment and selection processes while becoming more efficient, reducing recruitment and selection costs, and saving significant time (Chapman & Rowe, 2001; Chapman & Webster, 2001, 2003). This trend, which has been intensified by COVID-19, raises urgent questions concerning the validity of virtual communications as a substitute for meeting in person in the context of hiring processes. Are virtual communications an ideal substitute to meeting in person? This article will try to answer this question while focusing on whether or not virtual communications are a valid substitute for in-person assessments.

**Study 1– Reliability and validity of virtual assessment centers   
Assessments in a VAC compared to in an FTF-AC**

Studies suggest differences between VC-based interviews and face-to-face interviews which may also produce differences between the assessments of candidates in these two types of interviews (Chapman et al., 2003; Chapman & Rowe, 2001; Sears et al., 2013). Previous studies have found performance differences between VC interviews and face-to-face interviews which may have particularly negative implications for screening processes that combine both face-to-face selection and a VC-based selection (Basch et al., 2020). Because both VAC and video-based interviews are based on real-time communication without a physical encounter and low information wealth (Croes et al., 2019; Wegge, 2006), it can be assumed that the impact of video technology on assessments in interviews and VAC will be comparable.

Studies examining these differences in assessments present an inconclusive picture. While a study by Chapman & Rowe (2001) found that interviewers rated candidates’ performance better in video interviews than in face-to-face interviews, other studies indicated a completely opposite trend (Basch et al., 2020; Blacksmith et al., 2016; Sears et al., 2013). The reasons for these contrasts are unclear, with some explanations in the studies focusing on how the virtual environment affects interviewers’ evaluations, and others focusing on how it affects candidates’ behavior (Basch et al., 2020). According to some studies, candidates’ behavior was influenced by the VC because it made it difficult for them to convey non-verbal cues and to make a positive impression (Blacksmith et al., 2016; Chapman & Rowe, 2001). In contrast, other studies indicate that young candidates felt more confident in the virtual environment that is natural for them, resulting in reduced anxiety levels and improved performance in video job interviews (Horn & Behrend, 2017; McColl & Michelotti, 2019; Powell et al., 2018; Valkenburg & Peter, 2011). In addition, people with social anxiety and introverted people prefer virtual interviews and tend to feel that they can express themselves better in front of others online, as opposed to in face-to-face interivew (McKenna et al., 2002).

Other explanations for the differences between face-to-face and virtual selection on assessors suggested that nonverbal communication between people in close contact produces psychological closeness (Croes et al., 2019). An online environment may also inhibit the development of interest, warmth, emotion, and personal relationships between participants and increase the likelihood of unpleasant and even hostile feelings forming (Croes et al., 2019; Walther, 2012). In contrast, analyses conducted within the framework of Attribution Theory (Kelley, 1973) uncovered that VC assessments are more effective than face-to-face assessments. The theory concerns how humans attempt to decode the behaviors of others by attributing them to either internal or external causes. In contrast to the noted tendency of attributing low performance to the interviewee rather than to the interview situation (Jones & Nisbett, 1972), in the context of VC interviews, interviewers may perceive the online format as being the primary cause for poor performance (Taylor & Fiske, 1975). In line with this, even where candidates’ performances are low, regardless of VC, assessors tend to believe that their performances were influenced by external factors (i.e., technology) and compensate them accordingly by over-correcting their assessments (Chapman & Webster, 2001). Interviewers may perceive the candidates in the VC as having a disadvantage over the candidates in the face-to-face conversation, and this may cause them to overcompensate the candidates in the VC with more positive assessments of these candidates (Wegener & Petty, 1995).

Other studies have found that, although interviewers reported greater difficulty in recognizing the candidate in a video interview, compared to in a face-to-face interview, their assessments in the video interview were similar to their assessments in the face-to-face one (Straus et al., 2001). Distinguished medical institutions in the United States that have examined the significance of moving from face-to-face interviews to VC-based interviews have found that it is an acceptable alternative that saves time and costs (Pasadhika et al., 2012; Vadi et al., 2016).

Since the studies presented in the reviews above were concluded, FTF-AC technology has developed significantly (Stone et al., 2015) and there have been improvements that have narrowed the gaps between face-to-face communication and that based on VC. For example, the use of high-resolution cameras, wide screens, and high-speed internet connections has expanded (Basch et al., 2020; Bohannon et al., 2013), and the synchronization between audio and video has improved, contributing to the ability to transmit non-verbal messages through VC (Joshi et al., 2020). Additionally, interviewers and candidates today have more experience in VC technology than in the past (Basch et al., 2020).

The older studies presented above mostly concerned VCs conducted in conference rooms in designated locations equipped with VC technologies (e.g., Chapman & Rowe, 2001; Croes et al., 2019; Sears et al., 2013). Today’s candidates have video technologies available on their personal devices, enabling them to undergo the selection process from the comfort and familiarity of their own homes (as is the case in this study) (McCarthy et al., 2017; Toldi, 2011). In addition, the assessment procedures used in an AC take much longer than an interview, therefore providing richer information, meaning that the online environment may be not be as significant a limiting factor for VACs as it is for online interviews. Considering this, it was hypothesized that technological advances, along with long observation times at the AC, would reduce the differences between a video-based VAC and an FTF-AC and the distribution of the assessments would be similar, leading to:

***Hypothesis 1****: Assessments in a VAC will be similar to assessments in an FTF-AC.*

**Reliability of measurement at a VAC**

Assessors at the AC play a key role because their assessments determine the candidates’ scores, on the basis of which decisions are made regarding recruitment. The reliability and accuracy of the scores given by the assessors is of major importance for the integrity of AC assessments (Kleinmann & Ingold, 2019). According to Classical Test Theory (Ghiselli et al., 1981), the score given by the assessor, known as the Observed Score, represents the difference between the “true score” and the measurement error. Based on the assumptions of this theory, the reliability of the measurement represents the degree of unsystematic and unpredictable variance or random error associated with the indices. The characteristics of different selection methods (in this study, FTF-AC and VAC) may affect the measurement error differently. This raises the question of whether there is a difference in the reliability of assessments between a VAC and an FTF-AC.

To address this issue, we referred to Funder’s (1995) Realistic Accuracy Model, which deals with circumstances in which an accurate assessment of a person’s psychological characteristics in a social environment can be made. This model can also be applied to selection processes (e.g. Christiansen et al., 2005). According to this model, accuracy in assessments is influenced by four elements: relevance, availability, detection and a correct use of relevant behavioral cues. Only when each of the four elements is successfully implemented can assessors provide accurate assessments of candidates (Lievens et al., 2015).

We examined how the four elements of this model are reflected in a VAC:

1. *Relevance* concerns the appropriateness of a VAC in evoking relevant candidates' behavioral traits similar to those evoked in an FTF-AC. The VAC presented in this study was developed by a team of occupational psychologists to ensure that tasks would evoke behaviors relevant to the assessment.
2. *Availability* refers to the extent to which candidates’ non-verbal cues are observable by assessors. In a VAC, assessors can see only the upper body, which may make it difficult for the assessor to interpret candidate behaviors (Blacksmith et al., 2016). In contrast, in an FTF-AC setting, many interactions, such as those occurring between participants during breaks, etc., would be missed by the assessors, whereas in a VAC, the entire assessment is available to the assessors. In addition, *all* the assessors present in a VAC have access to the candidates’ behaviors at the same time, as there is no interaction during breaks or waiting times that may not be available to all assessors (Kleinmann & Ingold, 2019). For these reasons, there are no significant differences in terms of access to information between a face-to-face and a VC interaction (Jabotinsky & Sarel, 2020).
3. *Detection* concerns the assessors’ ability to detect and take note of relevant behaviors in candidates and elaborate them immediately and consciously. While the comprehensibility of behaviors in the VAC may be compromised by reduced non-verbal communication, leading to misconceptions about the candidates’ social skills (Blacksmith et al., 2016), there are fewer participants in the VACs in this study than in the FTF-ACs , making it easier to detect relevant inputs. In addition, in VACs, assessors can take more detailed notes (Chapman & Rowe, 2001).
4. In this study *correct use and assessment of observed relevant behaviors* at the two ACs was ensured by subjecting assessors to professional training and mentoring.

In this study, two of the aspects concerned in the above model (availability and detection) are, by necessity, applied differently in the VAC and the FTF-AC, but they are applied in such a way that compensates for any salient differences in the results. These four elements were successfully implemented both in the VAC and in the FTF-AC. It was hypothesized that the reliability of the evaluators’ assessments in the VAC would be similar to that from the FTF-AC. Findings that reinforce this hypothesis can be seen in a study on an asynchronous video-based state-of-the-art test in which there was high agreement among assessors (Cucina et al., 2015). Similar results emerged from another study that found that observation and assessment in an AC that was video-captured does not affect the accuracy of the observation or evaluation (Ryan et al., 1995). Considering the existing literature and the measures put in place to compensate for any discrepancies between a VAC and an FTF-AC led to the next hypothesis:

***Hypothesis 2****: The inter-rater reliability (reliability of assessments performed by different assessors) of the assessments at a VAC will be similar to the interrater reliability of the assessments at an FTF-AC.*

**Ensuring the validity of a virtual assessment center**

A significant challenge in the selection of personnel is ensuring that the essence of the assessed abilities in different selection tools that are designed to select candidates to the same job description is similar. For example, can a quality like *leadership* actually be assessed in a VAC in the same way as it can in an FTF-AC? In order to answer this question, one of the aims of this study is to compare the structural validity of a VAC with that of an FTF-AC. The assumption is that, in terms of core abilities and dispositions, the assessments will be similar.

In general, it is accepted that social behavior is co-determined by both personality traits and situational context, and, only when situations are functionally equal, can behavioral observations exhibit a valuable degree of consistency (Wright & Mischel, 1988). Greater consistency can be expected if the exercises used in virtual assessment centers and face-to-face assessment centers place similar demands on candidates (Sackett & Harris, 1988). In this study, the exercises used in the VAC and in the FTF-AC were constructed identically, the only variable being the constraints imposed by the virtual medium in the former. When exercises from a VAC and an FTC-AC that assess the same ability place similar demands on candidates, one can expect more consistency, and a higher correlation of assessments between exercises (Sackett & Harris, 1988).

Previous studies examining the correlations between AC assessments and other external variables have consistently found that AC assessments are related to measures of cognitive ability and personality (Collins et al., 2003; Shore et al., 1990), but these correlations have been found to be low (Damitz et al. al., 2003). For example, relatively low correlations have been found between the interpersonal sensitivity dimension and the cognitive dimension. The low correlation was expected because the measurement of the interpersonal dimension concerns a candidate’s style when communicating with others (Shore et al., 1990). In contrast, strong correlations have beenre found between interpersonal abilities and interpersonal personality traits in AC assessments (e.g., extroversion and emotional stability) (Spector et al., 2000).

In view of the lack of research examining the impact of VAC on the nature of the assessed abilities, our hypothesis is necessarily preliminary given the consistency of the findings, as presented, indicating that similar dimensions are generally more strongly related than different dimensions (Bray & Grant, 1966), it is expected that this trend will also hold true regarding the abilities assessed in a VAC. Thus, it is suggested that the context of an AC operating in a virtual environment, in which young candidates feel more at ease (Valkenburg & Peter, 2011), will not affect the essence of the measured dimension. We predict that the essence of the assessed abilities in VAC and FTF-AC that are designed to select candidates for the same job description will be similar, and also the construct validity of the two ACs will be similar, leading to the following hypothesis:

***Hypothesis 3:*** *The construct validity of a VAC will be similar to the construct validity of FTF-AC.*

**Methodology**

**Procedure and the participants**

The participants in this naturalistic field study were candidates for a variety of positions in the Israel Defense Forces (IDF). All the participants were women (the AC in this study is for women only) in the age range 16.2 to 24.5 (M = 17.3, SD = 0.5). In this longitudinal study, data were collected from the participants on two separate test days, six months apart. These represented two different phases of the military service selection system. The first selection day, attended by all participants (N=11,157) was conducted face-to-face. Some participants took part in the second day of the selection process in person in FTF-AC format (N=6,992). However, owing to the outbreak of COVID-19, for the remainder of the candidates, the process was conducted online through a VAC (N=4,165).

The assessors were graduates of a diagnostic position training program or an army training program, or were students in the field of social sciences, in the age range of 22–36 (M = 27.40, SD = 2.92). The assessors, who were well-trained and worked for a large recruitment company, tested the candidates in several dimensions (e.g. teamwork, leadership). Assessment dimensions were identical in both the VAC and the FTF-AC, and were assessed by means of similar exercises. Data for this study were provided by the Department of Military Behavioral Sciences in the form of a file with each participant’s identification number deleted to ensure privacy and anonymity.

**Description of the selection in the virtual and face-to-face assessment centers**

On the first selection day, the candidates reported to the selection site to perform a cognitive test on a computer, and attend a personal interview. The second selection day, on which some of the candidates performed on the selection site and some remotely, included an AC that lasted about three hours. The FTF-AC was performed at the selection site in the presence of candidates and assessors. For the VAC, the candidates and assessors were connected virtually. The candidates in the AC were randomly divided into different groups, each of which included two assessors with six candidates in the video-based selection, and eight candidates in the face-to-face selection.

During the AC, three exercises were performed:

1) A group exercise that tested teamwork and leadership ability which included several group tasks in a sequence, requiring participants to cooperate to solve a problem (e.g., group discussion or preparing a joint product together);

2) An oral presentation exercise in which each candidate delivered a short lecture to the group,

testing for presentation skills, such as oral expression, adjusting content, and creating interest;

3) A role-playing exercise testinginterpersonal sensitivity which involved two individual role-plays between candidates and assessors. Candidates were asked to play a pre-determined role in a mock situation that included an emotional or interpersonal problem and the assessor played the second role. The role-plays revealed important information about the interpersonal skills of the candidates – for example, their sensitivity and empathy toward other person. The exercises in the synchronous VC were performed using the video conferencing application Zoom with candidates and assessors connected to the VC from a PC or laptop with webcams set up so that the head and torso of each participant were clearly visible.

In line with the recommendations of Tenopyr (1977), a set of simple, well-defined abilities were identified for the AC which were tested using exercises designed to make the candidates’ abilities easily observable. For example, to evaluate leadership, the assessor examines how dominant the candidate is in the group and whether the other members of the group are listening to them, etc. Also, as part of these AC exercises, clear indicators were developed for each ability being assessed. For each of the three exercises, the assessor completed an assessment form for the candidates and assessed the relevant abilities for each exercise.

**Table 1**: Dimensions and exercises in the assessment center

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Sitting exercise | |  |  | | --- | --- | | Standing exercise |  | | |  |
| Role  play | Oral  presentation | Group  exercise | Exercises |
|  |  | X | Teamwork skills  Dimensions |
|  |  | X | Leadership skills |
|  | X |  | Presentation skills |
| X |  |  | Interpersonal sensitivity |

**Measures**

**Assessments**

Candidates were assessed in four different abilities, all of which were appropriate to the AC setting and were relevant to a variety of military roles. The first was teamwork ability, relating to the candidate’s involvement in the group and her investment in advancing the group mission, producing fruitful collaborations, and developing working relationships with others. The second ability was leadership, which relates to exercising effective authority over others, taking responsibility for the group task, and motivating the group to achieve its goals successfully. The third was presentation ability that relates to the candidates’ ability to deliver professional content in a clear and interesting manner. The fourth was the ability to care for other people, which refers to the ability to create service relationships while expressing empathy and sensitivity to the other.

The assessment was based on the behaviors of the candidates in three group and individual exercises at the AC. Performance was graded on a five-level scale (1 – “very low” to 5 – “very high”). The final score of each candidate in each dimension is the average of the scores of the two assessors who observed the candidate (apart from “ interpersonal sensitivity,” which was based on the assessment of a single evaluator in two different role-plays).

**“**General cognitive ability” was measured on the first selection day of the military assessment process using the same basic cognitive ability test given to all new recruits – called the Initial Psychotechnical Rating (IPR). This test has been used in many studies, and has been widely validated according to a variety of criteria (Fine et al., 2016; Luria et al., 2019; Tziner, 1988). The test consists of four sub-tests – two tests of analogies, formal and verbal (Mulholland et al., 1980), which measure the ability to deduce from a rule and apply it in other verbal or formal relations; a test examining arithmetic ability and based on the Wechsler Adult Intelligence Scale; and the fourth test examining comprehension ability and accuracy in executing instructions (based on an adaptation of a U.S. military intelligence test [Rabinowitz et al., 2000]). The test score range was 10–90 (M = 56.599, SD = 18.025).

**“**Adjustment**”** was measured on the first day of the military selection process using a structured interview designed to predict adaptability to military service among women, developed by the Israeli Army’s Behavioral Sciences Department. The interview was conducted by high school graduates, all female soldiers aged 18–20 who were selected and trained in a several-month-long course to perform this assessment process. The test was based on a common test applied to young men that has been used in several studies and is widely validated according to a variety of criteria (Luria et al., 2019a; Reeb, 1969; Tubiana & Ben-shakhar, 1982) and has recently been adapted for young women. The score range is 8–40 (M = 23.997, SD = 4.540).

**Results**

**Preliminary analyses**

The study includes two groups – one that performed the FTF-AC over several months prior to the outbreak COVID-19, and another that performed the VAC, over several months, follow the pandemic’s outbreak. We examined the differences in cognitive ability and adjustability between them (as measured in the first selection day of the military selection process) before examining the hypotheses to rule out other possible explanations for possible differences found between the groups. Because of the size of the groups, we calculated an effect size measurement and used Cohen’s (1988) rules of thumb defining d = 0.20 as a small effect, d = 0.50 as a medium effect and d = 0.80 as a large effect. First, we examined whether there were differences in cognitive ability between the groups and found no difference between candidates in the FTF-AC (M = 56.339, SD = 17.457) and candidates in the VAC (M = 57.129, SD = 19.121); (t [6300.490] = 2.039, p <0.05, d = 0.044). Secondly, we examined whether there were differences in the ability to adjust to the military framework and found no difference between candidates in the FTF-AC (M = 24.019, SD = 4.573) and candidates in the VAC (M = 23.957, SD = 4.481); (t [10697] = - 0.673 p> 0.5, d = -0.014).

**FTF-AC assessments compared to VAC assessments**

Campbell & Fiske (1959) proposed that studying prediction methods is best accomplished by comparing them when the predictor construct remains constant and only the method changes. Table 2 shows the averages and standard deviations of the candidates’ assessments in the various dimensions in the two ACs. Hypothesis 1, that assessments at the video-based VAC would be similar to assessments at the FTF-AC, was partially confirmed. For assessments in two dimensions – teamwork and interpersonal sensitivity – no differences in assessment averages were found between the VAC and the FTF-AC (see Table 2). In the assessments of the dimensions of –leadership and presentation – contrary to the hypothesis, small effects were found (see Table 2). The average of the assessments in leadership and presentation within the VAC is higher than the average of assessments in the same dimensions in the FTF-AC. Beyond examining the magnitude of the effect between the assessments’ averages, we also examined the form of distribution for the four dimensions in the two ACs. Appendix 1 presents the distributions of assessment scores in the various dimensions. It demonstrates that the distribution of assessments for the same dimension in the two different ACs is similar.

**Table 2:** Descriptive statistics and tests of between-subject effects

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Cohen’s d | p | t(X) | Face-to-face  assessment center | | | Virtual  assessment center | | |  |
|  |  |  | SD | M | N | SD | M | N |
| 0.27 | 0.00 | (18182.9)20.23 | 0.88 | 2.76 | 13,484 | 0.84 | 3.00 | 8,345 | Leadership skills |
| 0.02 | 0.10 | ) 17254.0)1.63 | 0.68 | 3.58 | 13,484 | 0.70 | 3.60 | 8,345 | Teamwork skills |
| 0.38 | 0.00 | (18821.6)27.97 | 0.80 | 2.81 | 13,745 | 0.73 | 3.11 | 8,338 | Presentation skills |
| 0.18 | 0.00 | (11078)9.57 | 0.67 | 3.13 | 6,933 | 0.63 | 3.26 | 4,147 | Interpersonal sensitivity |

**Interrater reliability in virtual and in face-to-face assessment centers**

The two assessors assessed the candidates in all dimensions, except for the interpersonal sensitivity dimension, in which candidates were assessed by only one assessor. Interrater reliability was calculated for the three dimensions of teamwork, leadership, and presentation in each of the ACs by calculating the correlations between assessors for each ability individually in the exercise they observed together. The correlations that were calculated were all highly significant (p<.001)). The leadership dimension was (N = 4171, r= .835) for the VAC and (N=6741, r=.854) for the FTF-AC. The teamwork dimension was (N= 4171, r=.818) for the VAC and (N=6741, r=.819) for the FTF-AC. The presentation dimension was (N=4,167, r=.824) for the VAC and (N=6,872, r=.864) for the FTF-AC. Much like with the FTF-AC, the correlations between the assessors at the VAC were higher than the value 0.8 and are therefore considered acceptable (Marcoulides, 1989). In addition, the difference between the reliabilities of the assessors for the same dimension in different ACs was small (the difference for leadership was 0.19, the difference for teamwork was 0.001 and the difference for presentation was 0.04). While the difference between teamwork reliability was not significant (Z = 0.153, P >.05), there was a significant difference in the leadership reliability (Z = 3.365, P <.01), and in the presentation reliability (Z = 7.116, P <.01).

According to these findings, Hypothesis 2 was partially confirmed. While the reliability between assessors’ assessments at the VAC was high in terms of Cohen’s (1988) rules of thumb, there were significant differences between them for some of the dimensions, similar to the reliability between their assessments at the FTF-AC. The reliability between the assessors in the presentation and leadership dimensions was significantly lower in the VAC, as compared to the reliability among the assessors in those dimensions in the FTF-AC. It is important to note that the reliability between the assessors in the teamwork dimension in the two ACs were similar and no significant difference was found between them.

**Construct validity of the VAC**

To test whether the validity of the construct of the dimensions in the VAC is comparable to the construct of the dimensions in the FTF-AC (i.e., whether the dimensions in the VAC measured similar capabilities to those measured in the FTF-AC), we mapped the dimensions in both ACs and compared between them in two stages. First, we focused on the internal construct of the dimensions in each AC by examining correlations between the dimensions and factor analysis. In the second phase, we focused on the correlations of the dimensions in each of the ACs versus stable external dimensions which were identical in both.

To compare the internal construct of the dimensions in both ACs, we calculated the correlations between the four assessed dimensions (see Table 3) and compared them between the two ACs (virtual and face-to-face). From Table 3, it can be seen that the correlations among the dimensions within the two assessment centers are medium-high for all pairs of dimensions. It also appears that the order of the correlation strengths between pairs of identical dimensions within each of the assessment centers is similar. For example, the pair of dimensions with the highest strength in the FTF-AC –teamwork and leadership – is also found with the highest strength in the VAC. Similarly, the pair of dimensions with the weakest correlation – teamwork and interpersonal sensitivity – shows the weakest correlation in both types of AC. This is in contrast to a situation where there is no similar trend between the two types of AC and where the correlations should appear in random order (Sawilowsky, 2002).

Second, to examine the degree of similarity in the construct validity of the two ACs, factor analyses were also performed for each AC separately to determine the degree of similarity in the way the dimensions in each are grouped into factors and the intensity of the loading of each index. Results of a principal component analysis with oblimin rotation are shown in Tables 4 and 5. In each factor analysis, two factors were defined, and in both, an eigenvalue of the first factor is greater than 2.5 and of the second less than 1. The correlation between the factors is similar, as the VAC is -.707 and the FTF-AC is -.749. In the FTF-AC, the first factor accounted for 53.26% of the variance and the second accounted for 5.47%, and in the VAC the first factor accounted for 60.20% of the variance and the second accounted for 7.05%.

The way in which the various dimensions are grouped into two factors in the two ACs is also similar. In both of the factor analyses, oral presentation skill has the highest loading for the first factor (above 0.85), followed by two additional indicators – interpersonal sensitivity, and leadership – which are loaded modestly on the first factor. In the two ACs, the dimensions are loaded negatively on the second factor, while teamwork loaded intensity on both (below -0.8), followed by an index of leadership in similar loading on both ACs.

**Table 3:** Correlations between assessment centers dimensions

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |
| 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | **Dimensions** |
|  |  |  |  |  |  |  |  | **Face-to-face assessment center** |
|  |  |  |  |  |  |  |  | 1. Leadership |
|  |  |  |  |  |  |  | .652\*\* | 2.Teamwork |
|  |  |  |  |  |  | .497\*\* | .608\*\* | 3. Oral presentation |
|  |  |  |  |  | .535\*\* | .469\*\* | .512\*\* | 4.Interpersonal sensitivity |
|  |  |  |  |  |  |  |  | **Virtual assessment center** |
|  |  |  |  |  |  |  |  | 5.Leadership |
|  |  |  | .711\*\* |  |  |  |  | 6.Teamwork |
|  |  | .523\*\* | .697\*\* |  |  |  |  | 7.Oral presentation |
|  | .502\*\* | .444\*\* | .465\*\* |  |  |  |  | 8.Interpersonal sensitivity |

N= 4165 for virtual, N=6992 for face-to-face

\*\**p* < .001 (two-tailed).

**Table 4:** Pattern Matrix in the VAC for two factors

|  |  |  |
| --- | --- | --- |
| Factor | | Dimensions |
| 2 | 1 |  |
| -.548 | .389 | Leadership |
| -.899 |  | Teamwork |
|  | .950 | Oral presentation |
|  | .382 | Interpersonal sensitivity |

Note – The table only includes loadings of at least 0.250.

**Table 5:** Pattern Matrix in the FTF-AC for two factors

|  |  |  |
| --- | --- | --- |
| Factor | | Dimensions |
| 2 | 1 |  |
| -.591 | .273 | Leadership |
| -.814 |  | Teamwork |
|  | .866 | Oral presentation |
|  | .461 | Interpersonal sensitivity |

Note – The table only includes loadings of at least 0.250

In order to complete the analysis of the validity of the construct of the VAC in relation to the FTF-AC, in the second phase, we also examined the relationship with external dimensions – cognitive ability and adjustment. Cognitive and adjustment abilities were examined in a standard and uniform way, using valid selection tools and in similar manner for all study candidates, both those who performed a VAC and those who performed an FTF-AC (see Table 6). The strength of the correlations between the various dimensions in each of the ACs and cognitive ability and adjustment were examined, rated and placed in order. The examination shows that the correlation strength order between the dimensions from each of the ACs, with similar adjustment (there are only two inversions in the order of the correlations strengths between the two ACs). Also, there is only one inversion in the order of the correlation strength between the dimensions in the two ACs and cognitive ability. Hence, all the examinations presented above provided support Hypothesis 3, as the construct validity of the VAC was found to be similar to the construct validity of the FTF-AC.

**Table 6:** Correlations between dimensions of assessment center and outcomes (cognitive ability and adjustment)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Cognitive ability** | | **Adjustment** | |  |
| Virtual | Face to face | Virtual | Face-to-face |  |
| **r** | **r** | **r** | **r** |  |
| .304\*\* | .296\*\* | .365\*\* | .373\*\* | Leadership |
| .263\*\* | .264\*\* | .299\*\* | .332\*\* | Teamwork |
| .370\*\* | .428\*\* | .359\*\* | .429\*\* | Oral presentation |
| .237\*\* | .277\*\* | .297\*\* | .355\*\* | Interpersonal sensitivity |

N= 4165 for virtual, N=6992 for face-to-face

\*\**p* < .001 (two-tailed).

**Discussion**

The comparison between candidates who performed the VAC and candidates who performed the FTF-AC revealed that their overall cognitive ability and adjustment were similar. This similarity is the basis on which comparisons between groups in subsequent analyses were possible. For all dimensions in the VAC, the assessments were the same or higher in relation to the assessments of those dimensions in the FTF-AC. Based on these findings, it appears that, in contrast to previous studies (Blacksmith et al., 2016; Chapman & Rowe, 2001), the candidates had no difficulty making a positive impression in the VAC. It is possible that, similar to results of studies conducted on virtual interviews (Horn & Behrend, 2017; McColl & Michelotti, 2019; Powell et al., 2018; Valkenburg & Peter, 2011), candidates feel more confident in what is a natural virtual environment for them – a VAC. Being in their natural environment decreases anxiety and improves performance. It is also possible that assessments improve in some dimensions because assessors attribute any poor performance in a VAC to the method in which it is performed (Taylor & Fiske, 1975) and tend to believe candidates were affected by difficulties that characterize the virtual environment, therefore compensating with a correction of their assessments (Chapman & Webster, 2001).

Focusing on the differences between assessments in the various dimensions revealed that while the assessments of teamwork and interpersonal sensitivity in the VAC were found to be similar to the FTF-AC, leadership and presentation ability assessments in the VAC were higher than in the FTF-AC. A possible reason may have to do with the fact that the essence of the “interpersonal sensitivity” dimension is an individual’s ability to express empathy and sensitivity to the other, and that the virtual environment may reduce candidates’ effectiveness in this regard. Nonverbal communication produces psychological closeness between people (Croes et al., 2019). Reduced potential for nonverbal communication in a VAC may make it difficult to develop interest, warmth, emotion, and personal relationships between participants (Croes et al., 2019; Walther, 2012).

The results of this study demonstrated high correlations between assessments in the VAC and the FTF-AC. These findings are of great importance, since reliability is the basis for selection validity and is critical for organizations to be able to trust the data and use them to make recruitment decisions. However, a significant difference was found between the reliability of the assessor’s assessments in the presentation and leadership dimensions in favor of the FTF-AC. It is, however, important to note that the raw differences in the correlations were minimal, and the reliability in the VAC was high, and above exactable in terms of reliability norms. In terms of structural validity, the correlations between the dimensions of the ACs and the factor analysis were similar in VAC and FTF-AC.

These findings contribute to personnel psychology as they present preliminary scientific evidence regarding the validity of VAC. Organizations that have adopted or are considering adopting VAC-based selection can assume, based on these findings, that there is a difference in assessments in a VAC and that the distribution is similar or slightly higher than in an FTF-AC. The reliability of assessments is high, and the validity of the construct is generally similar to FTF-AC. In addition, because the assessments of some of the dimensions are higher in VAC than in FTF-AC, combining the two types of ACs is not recommended, as candidates tested in a VAC will have an advantages over those tested in FTF-AC.

This research is a field study and, as such, is not without limitations. First, the population in the study were all young and female. While there is no reason to suspect that male candidates would differ, there may be difficulties in generalizing the study to older segments of the population who are less comfortable in online settings (Valkenburg & Peter, 2011). Replicating this study with an older, mixed-sex population would help mitigate this limitation. Secondly, the number of dimensions examined in this study is limited and the assessments in one of these, the teamwork dimension, are particularly high in the FTF-AC and therefore may be experiencing a “ceiling effect.” Dimensions missing from this study include organizational ability, persistence ability, and motivation. These shortfalls, combined with the limitation noted above in terms of the teamwork dimension, make it difficult to conclude whether the distributions of the assessments in the dimensions are the same between the two types of ACs or different. To expand on our understanding regarding differences between FTF-AC and VAC, further research that includes a variety of additional dimensions is recommended. Information from this study could shed light on the reasons for discrepancies present in some dimensions in assessments between the VAC and the FTF-AC.

**Study 2– The validity of the Virtual Assessment Center**

**Introduction**

The practice of developing predictors to improve occupational decisions regarding job applicants is an area of ​​great interest among occupational psychologists (Arthur Jr et al., 2008). From an organizational contribution perspective, the most important characteristic of the manpower assessment method is predictive validity; the ability to predict future job performance or job-relevant training (Sackett et al., 2017; Schmidt & Hunter, 1998). The use of highly predictive recruitment and selection methods leads to increased corporate profits (Hunter et al., 1990). Employee selection is critical to ensure that organizations are staffed with the right people, meaning those with the right attitude, knowledge, skills, and abilities (Hinkin & Tracey, 2010).

Studies that assess predictors are quite common in the literature (Hunter & Hunter, 1984; Schmitt et al., 1984). There is ample evidence of strong links between FTF-AC and performance and achievement in various activities, such as promotion, performance, and wage promotion (Thornton & Gibbons, 2009). These findings culminated in Schmidt & Hunter’s (1998) analysis summarizing 85 years of research on the validity of manpower selection processes. This analysis found that the average validity coefficients of an FTF-AC versus performance criteria at work is 0.37 (Schmidt & Hunter, 1998), and a later meta-analysis found that the average validity coefficients are 0.44 (Sackett et al., 2017).

The literature in the field has raised the need to make an important distinction between two groups of predictors (Arthur Jr. et al., 2008). The first group – method predictors – concerns the techniques or processes by which relevant behavioral information is collected (e.g. AC, or interview). The second group, predictor constructs, concerns the sampled behavioral domain (for example, leadership, ability, or adaptability). Studies conducted on ACs have to date examined predictor constructs and focused mainly on the predictive power of the various dimensions, while including ACs of different types together (e.g., Collins et al., 2003; Thornton & Gibbons, 2009). The novelty of this study is that it examines the predictive validity of a new predictive method – the VAC. To our knowledge, no study exists that examines the validity of VAC, although it has become a common practice since the COVID-19 pandemic. This study contributes to answering questions regarding the use of technology and its effects on the predictive validity of AC (Kleinmann & Ingold, 2019).

**Validating VAC**

One of the main differences between video-based virtual communication and face-to-face communication is the availability of non-verbal cues between participants (Joshi et al., 2020). In video conversations, fewer non-verbal behaviors, such as eye-contact and body language, can be observed in the candidates, which can make assessing candidates’ abilities more challenging (McColl & Michelotti, 2019; Sears et al., 2013). For example, it is more difficult to assess “soft skills” and interpersonal aspects in video-based selection (Sears et al., 2013).

Other studies have suggested that face-to-face communication can elicit information unrelated to the role, which can potentially distract assessors (Mennecke et al., 2000). In contrast, because nonverbal behavior in video conversation is reduced, decisions are based more on facts than on personality traits, so that when using video technology, the reference is more to facts and less to emotional aspects (Fullwood, 2006). Support for this was found in a study by Chapman and Rowe (2001) who found that some assessors felt they were better able to focus on verbal content in video-based interviews because technology limits potentially distracting non-verbal information**.** They found that some interviewers who conducted video-basedinterviews felt that they could evaluate the interviewees more objectively than interviewers who conducted face-to-face interviews.

Therefore, it seems that reducing non-verbal information in a VAC may, on the one hand, impair predictive validity due to the loss of non-verbal behavior, but, on the other hand, may also contribute to predictive validity due to assessors’ attention to more relevant information. In addition, a VC-based interview was found to allow interviewers to record more comments than a face-to-face interview (Chapman & Rowe, 2001). This finding is significant because there is evidence to suggest that recording notes can increase the accuracy and predictive power of assessments in interviews (Biesanz et al., 1999; Middendorf & Macan, 2002). Based on these findings, it can be hypothesized that, despite the disadvantages of VAC in terms of losing non-verbal information, its characteristics, including reducing distractions and ease of writing notes during behavioral observation, contribute to the assessor’s ability to assess candidates’ behavior. To explore this, we examined the predictive validity of assessments in a VAC by predicting job success and leadership. We also examined the validity of the construct between the VAC and the FTF-AC, based on the assumption that similar dimensions from different centers would be correlated. The foregoing discussion leads to the following hypotheses:

***Hypothesis 4****: Dimensions from a VAC will predict job success and leadership (predictive validity);* and

***Hypothesis 5:*** *Dimensions from a VAC will be correlated to similar dimensions of FTF-AC (congruent validity).*

**Methodology**

**Participants and Procedure**

One hundred and twenty-three participants were involved in this study, thirty-four of whom were omitted because they lacked data, and four more of whom were omitted because they did not meet the pre-defined minimum duration of acquaintance with the commander who filled out opinions on them. Eighty five qualifying participants remained, all women aged 18–21 with 12 years of education. Participants served as soldiers in a variety of junior positions in the military, including clerks, instructors, commanders and more.

These participants performed the AC twice. The first time was as part of the FTF-AC for army service selection. The second time occurred about 18 months later when they were already enlisted soldiers, but were asked to go through a VAC for research purposes, only with no direct consequences for them.

Participants performed the VAC that included a group exercise, a topic presentation exercise, and a role-playing exercise. The AC was performed from a home computer in a quiet environment as part of a group of about six participants and two assessors (for details on the VAC, see Methodology in Study 1). Participants were asked to behave naturally, as they would if they were to perform a “real” selection, and were assured that the use of the information would be for research purposes only. In addition, a questionnaire on the performance of the participants’ role was presented to their commanders in the army, in which they were asked to assess their actual performance in the relevant dimensions examined at the AC. The study was presented to and approved by the Ethics Committee (385/20).

**Measures**

***Assessments***

The assessors assessed the participants in the exercises in four different dimensions –– teamwork, leadership, presentation, and interpersonal sensitivity. Assessment was based on the behaviors of the candidates in a variety of group and individual exercises in the ACs. Candidates were rated on a five-point scale (1– “very low” to 5– “very high”). The final score of each participant in each dimension is the average of the assessments of the two assessors, with the correlation between them for all dimensions being p <0.01 r> 0.80. This excludes interpersonal sensitivity which was based on the assessment of a single assessor.

***Success in the job***

Assessments were performed by a supervisor in the form of questionnaire on participants’ performance in four dimensions assessed at the AC– presentation ability, teamwork, interpersonal sensitivity, and leadership. Supervisors were asked to rate participants’ abilities on the Likert scale (1 = very low to 5 = very high). For example, they were asked to evaluate the presentation and leadership abilities of the participants as demonstrated in the role-play. This questionnaire was completed by the direct supervisor shortly after conducting the VAC and passed to researchers directly without showing it to the participants.

***Informal leadership emergence***

This dimension was measured in an assessment process during basic military training (about 10 weeks after enlistment). Participants in basic training were asked to indicate who in the group had the potential to be a leader in a peer review questionnaire (also known as a sociometric questionnaire). To account for differences in group sizes, the number of selected leaders depended on the number of group members. The score ranged on a scale of 0–100 representing the percentage of the group that indicated the participant as suitable to be a leader. A score of 0 means that the participant was not selected, whereas a score of 100 means that all group members chose the same participant (Kalish & Luria, 2016; Luria et al., 2014, 2019b). In the current study, the range was from 0–95 (M = 17.79, SD = 23.433).

**Results**

**Predictive validity– correlations between VAC assessments and job performance**

Table 7 presents descriptive statistics and correlations between the four VAC dimensions and measures of outcomes (informal leadership emergence and job success). The findings indicate that there are correlations between some of the dimensions evaluated at the VAC and measures of outcomes (leadership emergence and job success) and therefore Hypothesis 4 was partially supported. The assessment of leadership in the VAC was associated with job evaluation of teamwork (r = .246, p <.05), and with an informal leadership emergence (r = .305, p <.05). The assessment of teamwork in the VAC was only related to informal leadership emergence (r = .354, p <.01). VAC assessment of presentation ability was found to be associated with almost all the job success outcomes measured: – leadership (r = .312, p <.01); teamwork (r = .302, p <.01); presentation (r = .322, p <.01), and informal leadership emergence (r = .292, p <.05). In contrast, the VAC assessment of interpersonal sensitivity was not associated with any of the outcome measures and none of the VAC assessment dimensions predicted the job evaluation of interpersonal sensitivity.

**Table 7:** Descriptive statistics and correlations between the VAC dimensions and outcome measures

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Performance evaluation** | | | |  |  |  |
| Informal leadership emergence | Interpersonal sensitivity | Oral presentation | Teamwork | Leadership | SD | M | **Dimensions**  **Virtual assessment center** |
| .305\* | -.148 | .086 | .246\* | .172 | .831 | 3.28 | Leadership (N=83) |
| .354\*\* | -.114 | .099 | .111 | .122 | .662 | 3.67 | Teamwork (N=85) |
| .292\* | .099 | .322\*\* | .302\*\* | .312\*\* | .709 | 3.68 | Oral presentation (N=72) |
| .154 | .094 | .179 | .082 | .163 | .712 | 3.81 | Interpersonal sensitivity (N=58) |

\**p* < .05, one-tailed.

\*\**p* < .01, one-tailed.

**Structural validity – correlations between VAC assessments and FTF-AC assessments**

Table 8 presents descriptive statistics and Table 9 presents the correlations between dimensions from the VAC and the FTF-AC. The dimensions from the VAC were found to be related to one or more dimensions from the FTF-AC, as detailed in Table 9. Examination of each dimension from the VAC as compared with the identical dimension from the FTF-AC revealed that, apart from the teamwork dimension, which was not significantly correlated to its identical dimension in the FTF-AC, all other dimensions had medium-high correlation with their identical dimensions – leadership r = .360, p <.01, presentation r = .488, p <.01, and the interpersonal sensitivity r = .248, p <.01. However, these dimensions from the VAC also showed similar correlations with other dimensions from a FTF-AC. Therefore, Hypothesis 5 was partially supported, as dimensions from the VAC correlated with similar dimensions from the FTF-AC but also similarly correlated to other (not identical dimensions). That is to say, it was possible to see a convergent validity with high correlations to identical dimensions but not a distinctive validity that expects low correlation to different dimensions. In addition, it can be seen that the only dimension from the FTF-AC that correlates with the four dimensions of the VAC is the presentation dimension.

**Table 8:** Descriptive Statistics of dimensions from the VAC and the FTF-AC

|  |  |  |
| --- | --- | --- |
| SD | M | **Dimensions** |
|  |  | **Virtual assessment center** |
| .831 | 3.28 | Leadership (N=83) |
| .662 | 3.67 | Teamwork (N=85) |
| .709 | 3.68 | Oral presentation (N=72) |
| .712 | 3.81 | Interpersonal sensitivity (N=58) |
|  |  | **Face to face assessment center** |
| .725 | 3.21 | Leadership (N=85) |
| .317 | 4.08 | Teamwork (N=85) |
| .734 | 3.48 | Oral presentation (N=85) |
| .567 | 3.81 | Interpersonal sensitivity (N=85) |

**Table 9**– Correlation between dimensions from the two assessment centers

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Interpersonal sensitivity | Oral presentation | Teamwork | Leadership | **Face-to-face assessment center**  **Virtual assessment center** |
| .289\*\* | .394\*\* | .049 | .360\*\* | Leadership (N=83) |
| .118 | .331\*\* | .131 | .296\*\* | Teamwork (N=85) |
| .347\*\* | .488\*\* | .188 | .513\*\* | Oral presentation (N=72) |
| .248\*\* | .394\*\* | .152 | .165 | Interpersonal sensitivity (N=58) |

\**p* < .05, one-tailed.

\*\**p* < .01, one-tailed.

**Discussion**

The results presented in this study contribute to an existing gap in the literature regarding the validity and reliability of VACs. This study provides a preliminary answer to the question of whether a VAC is a good personnel selection tool. The findings indicate that there were correlations between three dimensions assessed in the VAC (leadership, teamwork, oral presentation) and job outcomes – job success and leadership emergence, or both. However, there was no correlation between the VAC assessment of the interpersonal sensitivity dimension and job outcomes. The significance of these results for organizations is that, for the first time, there is research-based scientific evidence indicating that estimates can be based on most of the dimensions from a VAC in favor of decision-making regarding personnel selection. This means that organizations that recruit candidates with high ratings from VAC assessments (leadership, teamwork, and presentation) are likely to have employees with greater chances of succeeding in their job.

However, the finding that there was a lack of correlation between the VAC assessment of interpersonal sensitivity and job dimensions may be attributable to a number of factors. The lack of correlation could be due to the small number of participants in the study, the difficulty of the candidates to demonstrate this ability owing to reduced non-verbal behavior in a virtual communication leading to difficulty developing emotional and personal relationships between participants (Croes et al., 2019; Walther, 2012), or difficulty on the part of acting supervisors to make an evaluation in this matter. The VAC assessment of candidates’ presentation ability was correlated with multiple outcomes beyond the supervisors’ evaluations of the presentation ability itself. The higher a candidate’s performance in a topic presentation exercise in the VAC, the more successful the candidate will be in his or her job in terms of presentation, leadership, and teamwork.

The current study found a correlation between the dimensions assessed in the AC and informal leadership emergence. It demonstrates that VACs and FTF-ACs that take place before the candidate is working in the organization can predict the perceptions of other members of the leadership potential of the candidate. It is important to note that informal leadership emergence predicts effectiveness as formal leaders more than a year after testing (Luria et al., 2019). The current study demonstrates that it is possible to use VAC as a valid first step in the selection of future organizational leaders. Building a leadership “pipeline” is essential to the success of organizations and the current study demonstrates that such efforts can and should start in the first stages of employee selection.

An examination of the correlations between dimensions from the VAC and dimensions from the FTF-AC revealed that they were mostly positively correlated, except for the teamwork dimension from the FTF-AC, which was not related to any dimension in the VAC. Correlations were found both between a dimension from the VAC and the parallel dimension from the FTF-AC, demonstrating convergent validity in comparison to already validated measures (FTF-AC). However, some dimensions from the VAC also correlated with other dimensions from the FTF-AC (that are not their parallel dimension) and did not demonstrate discriminating validity. That is, the expected construct of correlations between dimensions of VAC and dimensions of FTF-AC was not found. There could be various reasons for the lack of discriminating correlations between dimensions from these two ACs. First, the four dimensions are related to interpersonal abilities and, because similar dimensions are logically more closely related than different dimensions (Bray & Grant, 1966), these dimensions are initially highly related in each of the ACs separately. In such a situation, it is impossible to discern discriminating and converging correlations in a VAC. Secondly, another possible reason for the lack of validity of the construct could be attributable to the long period of time that elapsed between the two ACs during which participants enlisted in the army and underwent various instruction courses designed to develop abilities and skills. The third reason may be the small number of participants involved in this study.

Along with the positive results of this preliminary study, there were also limitations. The first limitation was the small number of participants. The second limitation was that the participants who took part in the study knew that their performance at the ACs would not affect them, and this may have impaired the degree of authenticity of the behavior. The third limitation is that the candidates who participated in this study belong to a number of positions in the military with threshold conditions for acceptance, and also received training and occupational experience during their military service. Therefore, the participants likely had average to higher than average abilities, and, therefore, there was no representation of low abilities. In addition, while in the AC, there participants in the group had no prior acquaintance, in this study, some of the participants knew each other and it is possible that this influenced their behavior.

**General Discussion**

The revolutionary changes in recruitment and selection technologies, along with organizational responses to the COVID-19 epidemic (Jones & Abdelfattah, 2020; Joshi et al., 2020) raise the need for research analysis of previously unexplored video-based AC to produce theoretical and applied knowledge to help organizations make recruitment decisions. It is expected that in the coming years, an increasing number of organizations will move toward the implementation of VACs and, therefore, understanding the differences between VACs and FTF-ACs can contribute much to the evolution of selection process.

This study provides preliminary information about the use of VACs in organizations. In general, this study suggests that despite the significant difference in performance and transfer conditions between a VAC and an FTF-AC, they are similar in many respects, especially assessment distributions, predictive validity, and construct validity. These findings are of great importance because they can provide support for HR managerial decisions based on empirical evidence regarding the implementation of a VAC. It has been found that, the use of a VAC, much like an FTF-AC, along with the organizational benefits of saving time and money and expanding the scope of relevant candidates (Chapman & Webster, 2001, 2003), also allows organizations to make decisions about manpower based on reliable and valid data. However, in some dimensions, the VAC assessments are higher than in an FTF-AC and, therefore, it would not be valid for organizations to compare candidates who performed a VAC with candidates assessed in an FTF-AC.

These findings support the transition of organizations to VACs, but it is important to note that this is a preliminary study only. Deepening the study and expanding it, both to additional populations, with an emphasis on an older population than that in the current study, and to additional assessment dimensions (such as strategic decision making), will shed light on the differences found in this study between the various dimensions of assessment. A re-examination of the validity of the prediction in additional ACs against job success measures is recommended to deepen our understanding of differences in group dynamics between a face-to-face communication and a virtual communication and their impact on assessment. Also, in order to enhance the understanding of this new selection tool, it is important to delve deeper into the candidates' and assessors’ perceptions towards VAC, with an emphasis on comparing these perceptions with FTF-AC.

**Bibliography**

Arthur Jr, W., Villado, A. J., Arthur, W., & Villado, A. J. (2008). The Importance of

Distinguishing Between Constructs and Methods When Comparing Predictors in Personnel Selection Research and Practice. *Journal of Applied Psychology*, *93*(2), 435–442. https://doi.org/10.1037/0021-9010.93.2.435

Basch, J. M., Melchers, K. G., Kurz, A., Krieger, M., & Miller, L. (2020). It Takes More Than a Good Camera: Which Factors Contribute to Differences Between Face-to-Face Interviews and Videoconference Interviews Regarding Performance Ratings and Interviewee Perceptions? *Journal of Business and Psychology*, *Im*. https://doi.org/10.1007/s10869-020-09714-3

Biesanz, J. C., Neuberg, S. L., Judice, T. N., & Smith, D. M. (1999). When interviewers desire accurate impressions: The effects of notetaking on the influence of expectations. *Journal of Applied Social Psychology*, *29*(12), 2529–2549. https://doi.org/10.1111/j.1559-1816.1999.tb00124.x

Blacksmith, N., Willford, J., & Behrend, T. (2016). Technology in the Employment Interview: A Meta-Analysis and Future Research Agenda. *Personnel Assessment and Decisions*, *2*(1). https://doi.org/10.25035/pad.2016.002

Bohannon, L. S., Herbert, A. M., Pelz, J. B., & Rantanen, E. M. (2013). Eye contact and video-mediated communication: A review. *Displays*, *34*(2), 177–185. https://doi.org/10.1016/j.displa.2012.10.009

Bray, D. W., & Grant, D. L. (1966). The assessment center in the measurement of potential for business management. *Psychological Monographs*, *80*(17), 1–27. https://doi.org/10.1037/h0093895

Campbell, J. A. (1998). Participation in videoconferenced meetings: User disposition and meeting context. *Information and Management*, *34*(6), 329–338. https://doi.org/10.1016/S0378-7206(98)00073-1

Chamorro-Premuzic, T., Winsborough, D., Sherman, R. A., & Hogan, R. (2016). New Talent Signals: Shiny New Objects or a Brave New World? In *Industrial and Organizational Psychology* (Vol. 9, Issue 3).

Chapman, D. S., & Rowe, P. M. (2001). The impact of videoconference technology, interview structure, and interviewer gender on interviewer evaluations in the employment interview: A field experiment. *Joumal of Occupational and Orgam\ational Psycholog*, *74*(3), 279–298. https://doi.org/10.1111/1468-2389.00208

Chapman, D. S., Uggerslev, K. L., & Webster, J. (2003). Applicant Reactions to Face-to-Face and Technology-Mediated Interviews: A Field Investigation. *Journal of Applied Psychology*, *88*(5), 944–953. https://doi.org/10.1037/0021-9010.88.5.944

Chapman, D. S., & Webster, J. (2001). Rater correction processes in applicant selection using videoconference technology: The role of attributions. *Journal of Applied Social Psychology*, *31*(12), 2518–2537. https://doi.org/10.1111/j.1559-1816.2001.tb00188.x

Chapman, D. S., & Webster, J. (2003). The use of technologies in the recruiting, screening, and selection processes for job candidates. *International Journal of Selection and Assessment*, *11*(2–3), 113–120. https://doi.org/10.1111/1468-2389.00234

Christiansen, N. D., Wolcott-Burnam, S., Janovics, J. E., Burns, G. N., & Quirk, S. W. (2005). The good judge revisited: Individual differences in the accuracy of personality judgments. *Human Performance*, *18*(2), 123–149. https://doi.org/10.1207/s15327043hup1802\_2

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Collins, J. M., Schmidt, F. L., Sanchez-Ku, M., Thomas, L., McDaniel, M. A., & Le, H. (2003). Can basic individual differences shed light on the construct meaning of assessment center evaluations? *International Journal of Selection and Assessment*, *11*(1), 17–29. https://doi.org/10.1111/1468-2389.00223

Croes, E. A. J., Antheunis, M. L., Schouten, A. P., & Krahmer, E. J. (2019). Social attraction in video-mediated communication: The role of nonverbal affiliative behavior. *Journal of Social and Personal Relationships*, *36*(4), 1210–1232. https://doi.org/10.1177/0265407518757382

Cucina, J. M., Su, C., Busciglio, H. H., Harris Thomas, P., & Thompson Peyton, S. (2015). Video-based Testing: A high-fidelity job simulation that demonstrates reliability, validity, and utility. *International Journal of Selection and Assessment*, *23*(3), 197–209. https://doi.org/10.1111/ijsa.12108

Damitz, M., Manzey, D., Kleinmann, M., & Severin, K. (2003). Assessment center for pilot selection: Construct and criterion validity and the impact of assessor type. *Applied Psychology*, *52*(2), 193–212. https://doi.org/10.1111/1464-0597.00131

Fine, S., Goldenberg, J., & Noam, Y. (2016). Integrity testing and the prediction of counterproductive behaviours in the military. *Journal of Occupational and Organizational Psychology*, *89*(1), 198–218. https://doi.org/10.1111/joop.12117

Fullwood, C. (2007). The effect of mediation on impression formation: A comparison of face-to-face and video-mediated conditions. *Applied Ergonomics*, *38*(3), 267–273. https://doi.org/10.1016/j.apergo.2006.06.002

Funder, D. C. (1995). On the Accuracy of Personality Judgment: A Realistic Approach. *Psychological Review*, *102*(4), 652–670.

Galen Kroeck, K., & Magnusen, K. O. (1997). Employer and job candidate reactions to videoconference job interviewing. *International Journal of Selection and Assessment*, *5*(2), 137–142. https://doi.org/10.1111/1468-2389.00053

Ghiselli, E. E., Campbell, J. P., & Zedeck, S. (1981). *Measurement theory for the behavioral sciences*. San Francisco : W. H. Freeman.

Hinkin, T. R., & Tracey, J. B. (2010). What makes it so great?: An analysis of human resources practices among fortune’s best companies to work for. *Cornell Hospitality Quarterly*, *51*(2), 158–170. https://doi.org/10.1177/1938965510362487

Horn, R., & Behrend, T. (2017). Video killed the interview star: Does picture-in-picture affect interview performance? *Personnel Assessment and Decisions*, *3*(1). https://doi.org/10.25035/pad.2017.005

Howland, A. C., Rembisz, R., Wang-jones, T. S., Heise, S. R., & Brown, S. (2015). Developing a Virtual Assessment. *Consulting Psychology Journal: Practice and Research*, *67*(2), 110–126.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, *96*(1), 72–98. https://doi.org/10.1037/0033-2909.96.1.72

Hunter, J. E., Schmidt, F. L., & Judiesch, M. K. (1990). Individual Differences in Output Variability as a Function of Job Complexity. *Journal of Applied Psychology*, *75*(1), 28–42. https://doi.org/10.1037/0021-9010.75.1.28

International Task Force on Assessment Center Guidelines. (2009). Guidelines and Ethical Considerations for Assessment. *International Journal of Selection and Assessment*, *17*(3), 243–253.

Jabotinsky, H. Y., & Sarel, R. (2020). Shall We Meet? An Experimental Comparison of Video Conferences and Face-To-Face Meetings. *SSRN Electronic Journal*, 1–67. https://doi.org/10.2139/ssrn.3589431

Jones, R. E., & Abdelfattah, K. R. (2020). Virtual interviews in the era of Covid-19: A primer for applicants. *Journal of Surgical Education*, *77*(4), 733–734. https://doi.org/10.1016/j.jsurg.2020.03.020

Joshi, A., Bloom, D. A., Spencer, A., Gaetke-Udager, K., & Cohan, R. H. (2020). Video Interviewing: A Review and Recommendations for Implementation in the Era of COVID-19 and Beyond. *Academic Radiology*, *27*(9), 1316–1322. https://doi.org/10.1016/j.acra.2020.05.020

Kalish, Y., & Luria, G. (2016). Leadership emergence over time in short-lived groups : integrating expectations states theory with temporal person-perception and self servind bias. *Journal of Applied Psychology*, *101*(10), 1474–1486. https://psycnet.apa.org/fulltext/2016-29686-001.html

Kelley, H. H. (1973). The process of causal attribution. *American Psychologist*, *38*(February), 107–128. http://www.communicationcache.com/uploads/1/0/8/8/10887248/the\_processes\_of\_causal\_attribution.pdf

Kleinmann, M., & Ingold, P. V. (2019). Toward a Better Understanding of Assessment Centers: A Conceptual Review. *Annual Review of Organizational Psychology and Organizational Behavior*, *6*, 349–372. https://doi.org/10.1146/annurev-orgpsych-012218-014955

Landy, F. J. (1986). Stamp collecting versus science. validation as hypothesis testing. *American Psychologist*, *41*(11), 1183–1192. https://doi.org/10.1037/0003-066X.41.11.1183

Langer, M., König, C. J., & Fitili, A. (2018). Information as a double-edged sword: The role of computer experience and information on applicant reactions towards novel technologies for personnel selection. *Computers in Human Behavior*, *81*, 19–30. https://doi.org/10.1016/j.chb.2017.11.036

Lievens, F., Schollaert, E., Keen, G., Lievens, F., & Schollaert, E. (2015). The interplay of elicitation and evaluation of trait-expressive behavior : Evidence in assessment center exercises. *Journal of Applied Psychology*, *100*(4), 1169–1188.

Luria, G., Kahana, A., Goldenberg, J., & Noam, Y. (2019a). Contextual moderators for leadership potential based on trait activation theory. *Journal of Organizational Behavior*, *40*(8), 899–911. https://doi.org/10.1002/job.2373

Luria, G., Kahana, A., Goldenberg, J., & Noam, Y. (2019b). Leadership Development: Leadership Emergence to Leadership Effectiveness. *Small Group Research*, *50*(5), 571–592. https://doi.org/10.1177/1046496419865326

Luria, G., Kalish, Y., & Weinstein, M. (2014). Learning disability and leadership: Becoming an effective leader. *Journal of Organizational Behavior*, *35*(6), 747–761. https://doi.org/10.1002/job

Marcoulides, G. A. (1989). The application of generalizability analysis to observational studies. *Quality and Quantity*, *23*(2), 115–127. https://doi.org/10.1007/BF00151898

McCarthy, J. M., Bauer, T. N., Truxillo, D. M., Anderson, N. R., Costa, A. C., & Ahmed, S. M. (2017). Applicant Perspectives During Selection: A Review Addressing “So What?,” “What’s New?,” and “Where to Next?” *Journal of Management*, *43*(6), 1693–1725. https://doi.org/10.1177/0149206316681846

McColl, R., & Michelotti, M. (2019). Sorry, could you repeat the question? Exploring video-interview recruitment practice in HRM. *Human Resource Management Journal*, *29*(4), 637–656. https://doi.org/10.1111/1748-8583.12249

McKenna, K. Y. A., Green, A. S., & Gleason, M. E. J. (2002). Relationship formation on the internet: What’s the big attraction? *Journal of Social Issues*, *58*(1), 9–31. https://doi.org/10.1111/1540-4560.00246

Mennecke, B. E., Valacich, J. S., & Wheeler, B. C. (2000). The Effects of Media and Task on User Performance: A Test of the Task-Media Fit Hypothesis. *Group Decision and Negotiation*, *9*(6), 507–529. https://doi.org/10.1023/A:1008770106779

Middendorf, C. H., & Macan, T. H. (2002). Note-taking in the employment interview: Effects on recall and judgments. *Journal of Applied Psychology*, *87*(2), 293–303. https://doi.org/10.1037/0021-9010.87.2.293

Mischel, W., & Shoda, Y. (1995). A Cognitive-affective system theory of personality: reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review*, *102*(2), 246–268. https://doi.org/10.1037/0033-295X.102.2.246

Mulholland, T. M., Pellegrino, J. W., & Glaser, R. (1980). Components of geometric analogy solution. *Cognitive Psychology*, *12*(2), 252–284. https://doi.org/10.1016/0010-0285(80)90011-0

Nehls, K., Smith, B. D. B., & Schneider, H. A. (2015). Video-conferencing interviews in qualitative research. In *Enhancing Qualitative and Mixed Methods Research with Technology* (pp. 140–157). IGI Global. https://doi.org/10.4018/978-1-4666-6493-7.ch006

O’Leary-Kelly, S. W., & Vokurka, R. J. (1998). The empirical assessment of construct validity. *Journal of Operations Management*, *16*(4), 387–405. https://doi.org/10.1016/s0272-6963(98)00020-5

Palmer, M. T., & Simmons, K. B. (1995). Communicating Intentions Through Nonverbal Behaviors Conscious and Nonconscious Encoding of Liking. *Human Communication Research*, *22*(1), 128–160. https://doi.org/10.1111/j.1468-2958.1995.tb00364.x

Pasadhika, S., Altenbernd, T., Ober, R. R., Harvey, E. M., & Miller, J. M. (2012). Residency interview video conferencing. *Ophthalmology*, *119*(2). https://doi.org/10.1016/j.ophtha.2011.09.032

Potosky, D. (2008). A conceptual framework for the role of the administration medium in the personnel assessment process. *The Academy of Management Review*, *33*(3), 629–648. https://doi.org/10.5465/AMR.2008.32465704

Powell, D. M., Stanley, D. J., & Brown, K. N. (2018). Meta-analysis of the relation between interview anxiety and interview performance. *Canadian Journal of Behavioural Science*, *50*(4), 195–207. https://doi.org/10.1037/cbs0000108

Rabinowitz, J., Reichenberg, A., Weiser, M., Mark, M., Kaplan, Z., & Davidson, M. (2000). Cognitive and behavioural functioning in men with schizophrenia both before and shortly after first admission to hospital: : Cross-sectional analysis. *The British Journal of Psychiatry*, *177*(1), 26–32. https://doi.org/10.1192/bjp.177.1.26

Reeb, M. (1969). A structured interview for predicting military adjustment. *Occupational Psychology*, *43*, 193–199.

Ryan, A. M., Daum, D., Bauman, T., Grisez, M., Mattimore, K., Nalodka, T., & McCormuck, S. (1995). Direct, Indirect, and Controlled Observation and Ratind Accuracy. *Journal of Applied Psychology*, *80*(6), 664–670. http://www.google.co.in/search?hl=en&q=qm\_ch5&btnG=Search&meta=%5Cn&aq=f&aqi=&aql=&oq=&gs\_rfai

Sackett, P. R., & Harris, M. M. (1988). A further examination of the constructs underlying assessment center ratings. *Journal of Business and Psychology*, *3*(2), 214–229. https://doi.org/10.1007/BF01014490

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), 1435–1447. https://doi.org/10.1037/apl0000236

Sawilowsky, S. S. (2002). A quick distribution-free test for trend that contributes evidence of construct validity. *Measurement and Evaluation in Counseling and Development*, *35*(2), 78–88. https://doi.org/10.1080/07481756.2002.12069051

Schlebusch, S., & Roodt, G. (2008). Introduction to Assessment Centres. In S. Schlebusch & G. Roodt (Eds.), *Assessment Centres: Unlocking potential for growth* (pp. 1–20). Randburg: KnowRes Publishing. https://doi.org/10.1201/b13357-22

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*(2), 262–274. https://doi.org/10.1037//0033-2909.124.2.262

Schmitt, N., Gooding, R. Z., Noe, R. A., & Kirsch, M. (1984). Metaanalyses of validity studies published between 1964 and 1982 and the investigation of study characteristics‏. *Personnel Psychology*, *37*(3), 407–422. 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. https://doi.org/10.1108/MD-09-2012-0642

Shore, T. H., Thornton, G. C., & Shore, L. M. (1990). Construct validity of Two categories of assessment center dimension ratings. *Personnel Psychology*, *43*(1), 101–114. https://doi.org/10.1111/j.1744-6570.1990.tb02008.x

Spector, P. E., Vance, C. A., Schneider, J. R., & Hezlett, S. A. (2000). The relation of cognitive ability and personality traits to assessment center performance. *Journal of Applied Social Psychology*, *30*(7), 1474–1491. https://doi.org/10.1111/j.1559-1816.2000.tb02531.x

Stewart, D. W., & Shamdasani, P. (2017). Online Focus Groups. *Journal of Advertising*, *46*(1), 48–60. https://doi.org/10.1080/00913367.2016.1252288

Stone, D. L., Deadrick, D. L., Lukaszewski, K. M., & Johnson, R. (2015). The influence of technology on the future of human resource management. *Human Resource Management Review*, *25*(2), 216–231. https://doi.org/10.1016/j.hrmr.2015.01.002

Stone, D. L., Lukaszewski, K. M., Stone-Romero, E. F., & Johnson, T. L. (2013). Factors affecting the effectiveness and acceptance of electronic selection systems. *Human Resource Management Review*, *23*(1), 50–70. https://doi.org/10.1016/j.hrmr.2012.06.006

Straus, S. G., Miles, J. A., & Levesque, L. L. (2001). The effects of videoconference, telephone, and face-to-face media on interviewer and applicant judgments in employment interviews. *Journal of Management*, *27*(3), 363–381. https://doi.org/10.1177/014920630102700308

Taylor, S. E., & Fiske, S. T. (1975). Point of view and perceptions of causality. *Journal of Personality and Social Psychology*, *32*(3), 439–445. https://doi.org/10.1037/h0077095

Tenopyr, M. L. (1977). *Content-construct confusion*. *30*, 47–54.

Thornton, G. C., & Gibbons, A. M. (2009). Validity of assessment centers for personnel selection. *Human Resource Management Review*, *19*(3), 169–187. https://doi.org/10.1016/j.hrmr.2009.02.002

Tippins, N. T. (2015). Technology and Assessment in Selection. *Annual Review of Organizational Psychology and Organizational Behavior*, *2*, 551–582. https://doi.org/10.1146/annurev-orgpsych-031413-091317

Toldi, N. L. (2011). *Job Applicants Favor Video Interviewing in the Candidate-Selection Process*. Employment Relations Today; Wiley-Blackwell. https://doi.org/10.1002/ert

Tubiana, J. H., & Ben-shakhar, G. (1982). An objective group questionnaire as a substitute for a personal interview in the prediction of success in military training in Israel. *Personnel Psychology*, *35*(2), 349–957.

Tziner, A. E. (1988). Effects of Team Composition On Ranked Team Effectiveness: The Blocked Fractional Factorial Design. *Small Group Behavior*, *19*(3), 363–378.

Vadi, M. G., Malkin, M. R., Lenart, J., Stier, G. R., Gatling, J. W., & Applegate, R. L. (2016). Comparison of web-based and face-to-face interviews for application to an anesthesiology training program: a pilot study. *International Journal of Medical Education*, *7*, 102–108. https://doi.org/10.5116/ijme.56e5.491a

Valkenburg, P. M., & Peter, J. (2011). Online communication among adolescents: An integrated model of its attraction, opportunities, and risks. *Journal of Adolescent Health*, *48*(2), 121–127. https://doi.org/10.1016/j.jadohealth.2010.08.020

Walther, J. B. (2012). Interaction Through Technological Lenses: Computer-Mediated Communication and Language. *Journal of Language and Social Psychology*, *31*(4), 397–414. https://doi.org/10.1177/0261927X12446610

Wegener, D. T., & Petty, R. E. (1995). Flexible Correction Processes in Social Judgment: The Role of Naive Theories in Corrections for Perceived Bias. *Journal of Personality and Social Psychology*, *68*(1), 36–51. https://doi.org/10.1037/0022-3514.68.1.36

Wegge, J. (2006). Communication via videoconference: Emotional and cognitive consequences of affective personality dispositions, seeing one’s own picture, and disturbing events. *Human-Computer Interaction*, *21*(3), 273–318. https://doi.org/10.1207/s15327051hci2103\_1

Woods, S. A., Ahmed, S., Nikolaou, I., Costa, A. C., & Anderson, N. R. (2020). Personnel selection in the digital age: a review of validity and applicant reactions, and future research challenges. *European Journal of Work and Organizational Psychology*, *29*(1), 64–77. https://doi.org/10.1080/1359432X.2019.1681401

Wright, J. C., & Mischel, W. (1988). Conditional hedges and the intuitive psychology of traits. *Journal of Personality and Social Psychology*, *55*(3), 454–469. https://doi.org/10.1037/0022-3514.55.3.454

## Appendice*s*

Appendix 1- Distribution of assessed dimensions in a VAC and an FTF-AC

Graph 1: Distribution of Assessments in oral presentation ability in an FTF-AC Compared to a VAC (N=22,083)

Graph 2: Distribution of Assessments in leadership skills in an FTF-AC Compared to a VAC (N=21,829)

Graph 3: Distribution of Assessments in teamwork skills in an FTF-AC Compared to a VAC (N=21,829)

Graph 4: Distribution of Assessments in interpersonal sensitivity in an FTF-AC Compared to a VAC (N=22,083)