LinkedIn-based Assessments of Applicant Personality, Cognitive Ability, and Likelihood of Organizational Citizenship Behaviors: Comparing Self-, Other-, and Language-based Automated Ratings

Nicolas Roulin & Rhea Stronach

Saint Mary's University

Article "in press" in *International Journal of Selection and Assessment*

Authors' note: Correspondence should be sent to Nicolas Roulin, Department of Psychology, Saint Mary's University, 923 Robie Street, Halifax, NS, Canada, B3H 3C3, E-mail: nicolas.roulin@smu.ca. The authors thank Khoi Anh Pham Le, Rene Arseneault, Mahboobe Sahebi, and Pilar Brooks for their help with data coding and material preparation. They also thank Mike Durland and Receptiviti for allowing them to use their Personality Insights platform.

Abstract

We compared self-reports or test-based assessments of personality, cognitive ability, and likelihood or tendencies to engage in organizational citizenship behaviors (OCB) from experienced workers (*targets*, *N* = 154) with one approach to rate these traits based on LinkedIn profiles using hiring professionals (*panel raters*, *N* = 200), graduate students in Industrial-Organizational Psychology (*I-O raters*, *N* = 6), and automated assessments with the language-based tool Receptiviti (for personality only). We also explored potential for adverse impact associated with this approach of LinkedIn profile assessments and how profile elements are associated with ratings. Results demonstrated that raters can reliably assess personality, cognitive ability, and OCB with one-item measures. LinkedIn showed little promise for valid assessments of personality (except some weak evidence for honesty-humility) and OCB tendencies for all data sources. And, we only found modest evidence of convergent validity for cognitive ability. Automated assessments of personality with Receptiviti were more consistent with raters' assessments than targets' self-reports. LinkedIn-based hiring recommendations did also *not* differ on the basis of gender, race, or age. Finally, in terms of profile content, longer LinkedIn profiles with more professional connections, more skills listed, or including a professional picture were viewed more positively by both types of raters. But these content elements were largely unrelated to targets' self-reports or test scores. Thus, organizations should be careful when relying on LinkedIn-based assessments of applicants' traits.

Keywords: LinkedIn; Social media; Automated personality assessments

Practitioners' points

- Cyber-vetting job applicants' social media profiles has become popular, but research is lagging behind practice.
- We examined LinkedIn-based assessments of personality, cognitive ability and likelihood of OCB.
- Human ratings were generally reliable and associated with limited adverse impact.
- Evidence of convergent validity was limited, and only visible for cognitive ability and honesty-humility.

LinkedIn-based Assessments of Applicant Personality, Cognitive Ability, and Likelihood of Organizational Citizenship Behaviors: Comparing Self-, Other-, and Language-based Automated Ratings

Over the last few years, assessments of job applicants based on their social media profiles, also known as cyber-vetting, has become an integral part of the personnel selection process in many organizations (Hartwell & Campion, 2020; Roth et al., 2016; Roulin & Bangerter, 2013; Zhang et al., 2020). A vast majority of the existing research has focused on personal social media platforms like Facebook. Initial evidence suggested that Facebook-based personality assessments could be reliable and valid (e.g., Kluemper & Rosen, 2009; Kluemper et al., 2012). Yet, subsequent investigations showed much lower levels of inter-rater agreements for assessments of applicant skills or hiring recommendations, and highlighted that Facebook-based judgments were largely unrelated to relevant workplace outcomes (e.g., job performance or turnover) and could lead to adverse impact (Van Iddekinge et al., 2016; Zhang et al., 2020). In addition, hiring professionals tend to focus on negative content (e.g., alcohol or drug use, sexual references, opinionatedness) to identify "red flags" when screening Facebook profiles (Hartwell & Campion, 2020; Tews et al., 2020). Applicants' reactions are also largely negative, and Facebook-based cyber-vetting is seen as low in face validity, unfair, and privacy invading (Cook et al., 2020; Stoughton et al., 2015). Taken together, the extant literature paints a grim picture and suggests that employers should generally refrain from using Facebook as a screening device.

However, this does not mean that employers and hiring professionals should necessarily discard social media altogether. Landers and Marin (2021) proposed that researchers examining technology should adopt a "technology-as-designed" paradigm, considering specific design characteristics or intended users. Applying this perspective to social media assessment, we suggest that professional social media, like LinkedIn, could represent a viable alternative to personal social media, like Facebook. Contrary to Facebook which is primarily used to communicate with friends and family, LinkedIn is typically used for professional networking and career development (Davis et al., 2020; Weidner et al., 2016). Both job seekers (Collmus et al., 2016; Johnson & Leo, 2020) and organizations (Guilfoyle et al., 2016; Nikolaou, 2014) are increasingly turning to LinkedIn in their job/applicant search process. As such, job seekers generally use LinkedIn as an ever-evolving online resume and post more job-related information (e.g., education, work experiences, skills, achievements) and less personal or legally-protected information (Zide et al., 2014). Importantly, preliminary evidence suggests that LinkedIn-based assessments demonstrate inter-rater agreement, temporal stability, *some* convergent and criterion-related validity, and likely limited adverse impact (Roulin & Levashina, 2019; van de Ven et al., 2017). Additionally, in comparison to personal social media platforms, applicants seem to react more positively to employers using LinkedIn to cyber-vet them, finding this platform fairer, more valid, and less privacy-invading (Cook et al., 2020; Stoughton, 2016). Yet, available evidence about the effectiveness of LinkedIn-based assessments is still limited. For instance, Roulin and Levashina (2019) relied on profiles of undergraduate students rated by MBA students. However, LinkedIn profiles of undergraduate students may not be representative of LinkedIn profiles in general, especially those of more experienced workers who might have more information available on their profiles. And, assessments or ratings provided by MBA students may not generalize to ratings from hiring manager or individuals more knowledgeable about ability or personality assessments. As such, more research is needed before we can confidently recommend organizations to use (or not use) LinkedIn in their selection process.

Moreover, existing research on LinkedIn-based assessments is largely limited to comparing self-reports to raters' judgments (Roulin & Levashina, 2019; van de Ven et al., 2017). This is certainly a valuable first step, as results can be compared to findings for self vs. other assessments of personality (Connolly et al., 2007). However, self-reports of personality and other traits or behaviors can be unreliable or biased by social desirability (Morgeson et al., 2007), and there are benefits to using multiple methods or sources of data to understand a phenomenon (e.g., Connolly et al., 2007; Denzin, 2012). A recent trend in personnel selection involves automated assessments of personality, for instance relying on technology to analyze patterns of language in text data and assess applicants' traits or qualifications. There is a growing body of work on the use of automated personality assessments based on social media content in the personality literature (e.g., Alexander et al., 2020; Bleidorn & Hopwood, 2019; Tay et al., 2020). Scholars have recently suggested how to use automated assessments in personnel selection too (e.g., Liem et al., 2018). There is also emerging work examining automated assessments of personality in video interviews based on both of-the-shelf tools which are increasingly used by organizations to save costs (Hickman et al., 2019) and new algorithms developed for the purpose of a study (Hickman et al., 2021). Yet, empirical research addressing the use of this technology to assess applicants' social media profiles (and especially professional platforms like LinkedIn) remains scarce.

Overall, the present research contributes to the growing literature on cyber-vetting and social media use in selection (e.g., Roulin & Levashina, 2019; Van Iddekinge et al., 2016; Zhang et al., 2020) in several ways. First, we replicate and expand recent efforts to assess the psychometric properties and potential adverse impact of raters' assessments of personality and cognitive ability based on LinkedIn profiles (Roulin & Levashina, 2019) using samples of experienced workers (as targets) and two independent groups of raters (an online panel of hiring professionals and graduate students in Industrial-Organizational Psychology). Second, while previous cyber-vetting research has focused on the Five Factor Model (FFM) of personality, we use the more comprehensive HEXACO model (Hough et al., 2015; Lee & Ashton, 2004). Third, in line with research suggesting that self-reports vs. observer-ratings of personality provide different but possibly complementary sources of information (Connolly et al., 2007), we examine relationships between self-reports or test scores, one approach to assess LinkedIn profiles with two different groups of raters, and automated assessments based on profiles using Receptiviti, an off-the-shelf language-based tool. Finally, we examine whether these LinkedIn-based assessments are associated with likelihood or tendencies to engage in OCBs, an understudied yet important behavioural outcome of employees in organizations (see Spitzmuller et al., 2008 for a review).

LinkedIn-based Assessments of Job Applicants

In line with the literature exploring social media assessments (Kluemper et al., 2012; Roulin & Levashina, 2019; Tay et al., 2020), we propose that a relevant theoretical framework to assess LinkedIn-based assessments is the realistic accuracy model (RAM; Funder, 1995, 2012). This model argues that the accuracy of personality judgments by a rater/perceiver depends on four (multiplicative) elements: (1) the relevant behavioural cues to the trait in the environment (i.e., does LinkedIn offer users the opportunity to display the trait of interest), (2) the availability of behavioral cues (i.e., is information about the trait available to the rater on a LinkedIn profile), (3) the extent to which cues are detected (i.e., is the rater capable of identifying the relevant information and motivated to use it), and (4) how the cues are used in the judgments (i.e., is the rater actually using the information obtained when assessing the trait).

As described by Roulin and Levashina (2019), the first two RAM elements are particularly relevant for examining the psychometric properties of LinkedIn-based assessments. Although the latter two elements provide valuable foundations for understanding rater behaviour, we primarily focus on the relevance and availability of trait cues in the subsequent sections outlining the psychometric properties of LinkedIn-based assessments.

LinkedIn-based Personality Assessments

The majority of research examining social media assessments of personality has been conducted using Facebook. Overall, studies have reported moderate-to-strong inter-rater reliability (i.e., ICCs ranging from .43 to .99) and low-to-moderate levels of convergent validity (i.e., correlations between self-reports and raters ranging from .08 to .44) for Facebook-based assessments of the Five-Factor Model of personality (FFM; Back et al., 2010; Kluemper & Rosen, 2009; Kluemper et al., 2012; Van Iddekinge et al., 2016). Among the “Big Five”, the highest self-rater correlations were found for extraversion and (to a lesser extent) openness. In an experiment, Schroeder et al. (2020) also showed that using a more structured approach to assess personality on Facebook was not associated with improvements in reliability or validity.

According to the RAM principles (Funder, 1995), one should expect somewhat lower reliability and validity coefficients with LinkedIn assessments of personality as compared to Facebook, due to fewer trait cues and limited availability. There might be a curvilinear relationship between the amount of information available in a profile and the accuracy of assessments (Roth et al., 2016), and the standardization of LinkedIn profiles might facilitate rating consistency. However, LinkedIn profiles offer comparatively fewer opportunities for users to provide information about most personality traits and is perceived by hiring managers to be less effective than Facebook in assessing personality (Hartwell & Campion, 2020). Among personality traits, extraversion is likely a trait that may be assessed on LinkedIn because it can be relatively easily demonstrated through the number of connections, teamwork activities, leadership roles, or volunteering. At their core, LinkedIn profiles are like extended online resumes. And, extraverted applicants tend to include more extracurricular activities, volunteering, elected office roles, and club memberships in their resumes (Cole et al., 2009). In addition, taking on leadership roles (at work or in school) is a central element of the social boldness facet of extraversion, whereas having more professional connections, being involved in volunteering activities or showcasing teamwork experiences or skills might signal sociability. For instance, extraversion is related to the number of connections one has on social media as extraverts are more sociable (Amichai-Hamburger & Vinitzky, 2010) and thus more frequently interact with others (Kluemper et al., 2012). The way one presents oneself on LinkedIn (e.g., the profile picture chosen, the language used in the “about” section, or how previous accomplishments are described) could also signal elements of social self-esteem or liveliness. In contrast, other traits like agreeableness or emotionality are arguably less visible on LinkedIn than on Facebook.

A recent study by Fernandez et al. (2021) confirms that extraversion may be the most highly visible trait on LinkedIn. The authors coded over 600 LinkedIn profiles for LinkedIn-based signals theoretically derived from the FFM (excluding neuroticism, as it is argued to be an in-person identifiable trait) and had profile owners complete a self-report personality inventory. Their results highlighted seven indicators of extraversion (i.e., number of connections, human interactions, sport activity, leadership roles, additional pictures, number of skills, and leadership skills), with number of connections being the strongest indicator of extraversion (and of any personality trait) in their study. In investigating what traits were best predicted from LinkedIn based on the indicators identified, they found extraversion to be most accurately signalled, followed closely by conscientiousness. However, it should be noted that various signals of conscientiousness were also significantly related to self-report extraversion, neuroticism, and agreeableness. Thus, conscientiousness may be a more challenging trait to assess from LinkedIn because it "cross-pollinates" with various other traits. Regarding openness and agreeableness, these traits were least accurately signaled from LinkedIn, perhaps as they are more dependent on job context. Thus, research by Fernandez et al. (2021) supports that extraversion, as a highly-visible trait, may be most accurately assessed from LinkedIn due to accessible user-generated signals from profile elements.

With regard to psychometric properties of assessing extraversion from LinkedIn, empirical research regarding the reliability and validity of LinkedIn-based assessments of personality is limited to two papers. First, van de Ven et al. (2017) examined the inter-rater reliability and convergent validity of FFM personality assessments based on the LinkedIn profiles in two studies. In the first study, ten psychology students assessed the profiles of 62 students and 116 employees. In a second study, 20 psychology students assessed the (full or partial) profiles of 97 employees. Overall, they reported moderate-to-strong levels of inter-rater reliability (ICCs ranging from .41 to .73 in Study 1 and .48 to .92 in Study 2) and only found modest evidence of convergent validity for extraversion (r ranging from .24 to .37). Roulin and Levashina (2019) also examined inter-rater reliability and convergent validity in two studies involving a total of 41 MBA raters assessing the profiles of 133 undergraduate business students. They found moderate levels of inter-rater reliability for personality (average ICCs ranging from .26 to .55 in Study 1 and .24 to .58 in Study 2) and only found modest evidence of convergent validity for extraversion in their two studies (r = .20 and .31).

Hypothesis 1. Raters' assessments of personality based on LinkedIn profiles demonstrate inter-rater reliability for all personality traits.

Hypothesis 2a. Raters' assessments of personality based on LinkedIn profiles correlate with targets' self-reports for extraversion.

While most of the existing research on personality assessment on social media relies on the FFM, we employ the HEXACO model of personality (Lee & Ashton, 2004). HEXACO includes six core traits (honesty-humility, emotionality, extraversion, agreeableness, conscientiousness, and openness to experience). Although related to the FFM, HEXACO provides a more comprehensive view of personality and its structure has been confirmed in various cultures and languages (Ashton & Lee, 2005; Hough et al., 2015; Lee & Ashton, 2014; Zettler et al., 2020). Importantly, the honesty-humility factor is not directly captured by the FFM, but is associated with the dark triad model of personality, integrity, and has been found to predict job performance, counterproductive work behaviors, and OCBs (Anglim et al., 2018; Lee et al., 2013; Lee et al., 2019; Wendler et al., 2018).

Although no research has directly explored whether honesty-humility might be accurately assessed on LinkedIn, van de Ven and colleagues (2017) found raters' assessments of trait self-presentation based on LinkedIn profiles to correlate .29 with self-reports of self-presentation. Trait self-presentation included the tendency to be self-confident, dominant, achievement-oriented, and ambitious, which overlaps with the low-end of some domains of honesty-humility (e.g., low modesty or greed avoidance). In addition, in line with the RAM principles (Funder, 1995), LinkedIn offers various opportunities to make honesty-humility visible. In particular, individuals low on modesty or greed avoidance might list more skills in their profile, seek more recommendations from colleagues or supervisors, and use more self-promotional language in their profile summary or when describing work experiences. This would be consistent with research showing that employees low on honesty-humility tend to engage in more impression management at work, including self-promotion (Bourdage et al., 2015). As such, we anticipate that honesty-humility might be a second highly visible HEXACO trait that can be accurately assessed via LinkedIn profiles.

Hypothesis 2b. Raters' assessments of personality based on LinkedIn profiles correlate with targets' self-reports for honesty-humility.

Automated Assessments of Personality Using Receptiviti

Extant personality research has highlighted the limitations associated with relying exclusively on self-reported measures, and the benefits of using multiple sources of data (e.g., observer- or peer-ratings; Connolly et al., 2007). For instance, self-reports of personality can be impacted by social desirability or faking, the inability to introspect accurately, or attempts to appear consistent (Morgeson et al., 2007). Personality is also regularly assessed in job interviews, although correlations between self-reports and interviewers' ratings are generally small (Hickman et al., 2019; Levashina et al., 2014). And, although more structured interviews can lead to more accurate ratings, applicant faking can substantially reduce validity (Van Iddekinge et al., 2005). This has led researchers to develop alternative approaches to assess personality, such as situational judgment tests (e.g., Oostrom et al., 2019). More recently, research has started to examine the use of technology to automatize the assessment of personality. Some authors have suggested that automated assessments of personality have the potential to be more consistent and accurate than human judgments, but argued more research is needed to examine potential biases or legal and ethical issues associated with fairness or privacy (Alexander et al., 2020; Tippins et al., 2021). It is also important to emphasize that automated assessments (for instance based on natural language processing techniques) can take a number of forms, and their reliability, validity, and fairness depends on how the assessment model was designed, developed, or tested (Landers & Behrend, 2022).

Social media has been a fertile ground for such assessments. For example, Tay et al. (2020) reviewed and meta-analyzed research on social media-based assessments of the FFM, including six samples utilizing language- (i.e., text-) based automated assessment. Tay et al. (2020, p. 826) define this approach as involving “computers analyzing the language used in text data to develop algorithms for predicting a person’s standing on self-reported personality traits” and even suggest that it can potentially be “a way of enhancing (or even substituting) human judgement in personality assessment”. Yet, their findings show that automated assessments of the FFM based on Facebook profiles only modestly correlate with self-reports, with mean convergent validities in the .20s range. In addition, only one study examined relationships between automated assessments and observers’ reports of the FFM (Park et al., 2015), and reported correlations ranging from .20 to .30. None of the automated assessment studies reviewed by Tay et al. (2020) relied on LinkedIn. And, to our knowledge, no other published research exists employing language-based automated personality assessments to LinkedIn.

Thus, in addition to comparing targets’ self-reports of personality and assessments of LinkedIn profiles by two samples of human raters, we propose to examine automated assessments based on the textual information garnered from LinkedIn profiles from a commercially-available product: Receptiviti Language-based Personality Insights. This is an “off-the-shelf” tool built on the extensive literature examining the relationships between language and personality (e.g., Pennebaker et al., 2003; Yarkoni, 2010). It is based on the Linguistic Inquiry and Word Count system (LIWC; Pennebaker et al., 2015), which uses an extensive dictionary to classify language elements into 94 psychologically-relevant categories. Past research has established that LIWC categories are related to the FFM. For instance, individuals using more words classified in the “anxiety” category are higher on neuroticism, whereas those using more words associated with the “friends” or “other references” categories are more extraverted (Yarkoni, 2010). The Receptiviti tool has been initially validated against personality self-reports of social media users (from Facebook and Twitter) from four different datasets (Golbeck, 2016). For instance, in the larger dataset (with 8,569 Facebook users), correlations between the Receptiviti-based FFM scores and self-reports ranged from .20 to .24. Obschonka et al. (2017) also replicated the relationships between LIWC categories and the FFM reported by Yarkoni (2010) using automated FFM scores from Receptiviti, and found correlations ranging from .40 to .78. Obschonka et al. (2017) then showed that Receptiviti-based FFM scores from Twitter posts can help differentiate successful managers from entrepreneurs.

The RAM principles and general arguments presented in the previous section about human ratings should also largely apply to Receptivi-based automated ratings. For instance, LinkedIn profiles of more extraverted users should include more words associated with LIWC categories representative of extraversion (e.g., references to others, social processes, communication). Although we would expect similar things for honesty-humility, Receptiviti only generates scores for the FFM and thus does not directly capture honesty-humility. Overall, we propose the following hypothesis:

Hypothesis 3. Automated language-based assessments of personality (using Receptiviti) based on LinkedIn profiles correlate with targets’ self-reports for extraversion only.

LinkedIn-based Cognitive Ability Assessments

According to the RAM principles (Funder, 1995), several cues of applicants' cognitive ability can be visible on LinkedIn profiles and should thus help achieve accurate assessments. For instance, resume research shows that education credentials and academic achievements (e.g., GPA, Dean's list, scholarships) or experience in complex jobs might signal higher levels of cognitive abilities (Cole et al., 2003). Not only do LinkedIn profiles offer applicants ample opportunities to display such information (e.g., provide detailed descriptions of education and work achievements), but these could be confirmed via recommendations or endorsements from colleagues or supervisors. In addition, hiring managers also perceive LinkedIn to be particularly effective to assess cognitive ability (Hartwell & Campion, 2020; Roulin & Bangerter, 2013).

Yet, research on assessments of cognitive ability based on social media profiles remains scarce. Research exploring cognitive ability assessments on Facebook has found evidence of interrater reliability (ICC = .98; Kluemper & Rosen, 2009) and only weak evidence of convergent validity (r = .23 using ACT scores; Van Iddekinge et al., 2016). With regard to LinkedIn, Roulin and Levashina (2019) also found evidence of inter-rater reliability (ICCs = .60 and .47) and modest evidence of convergent validity (r = .30 and .27 using Wonderlic scores) in both their studies. We thus expect to replicate these findings in the present study, and to expand them by using multiple sources of data, with targets' cognitive ability test scores and LinkedIn-based assessments using two groups of raters.

Hypothesis 4. Raters' assessments of cognitive ability demonstrate inter-rater reliability.

Hypothesis 5. Raters' assessments of cognitive ability correlate with targets' test scores.

LinkedIn-based Predictions of Organizational Citizenship Behaviors

Research examining the predictive validity of LinkedIn is extremely scarce. Roulin and Levashina (2019) found that LinkedIn-based hiring recommendations were positively correlated with some career success indicators collected two years later (i.e., the target individual obtaining a job aligned with their degree or getting promoted). In contrast, a recent study by Cubrich et al. (2021) utilized a field sample of 486 financial service professionals to examine the criterion-related validity of LinkedIn profiles content to predict sales performance metrics. Trained raters coded LinkedIn profile elements using pre-determined categories (e.g., profile picture present, summary section present, highest level of education achieved, recommendations received and given) to examine the relationships between these various profile elements and two objective sales performance metrics (i.e., bringing in new business and expanding current business). With minimal exceptions (e.g., presence of a summary section), Cubrich et al.'s (2021) results largely demonstrated that LinkedIn profile elements were not strongly related to task performance (at least in the financial services/sales industry). Overall, there is very limited evidence about the criterion-related validity of the content of LinkedIn profiles or assessments based on such profiles.

In addition to assessing personality and cognitive abilities (two valid predictors of in-role, task performance; e.g., Schmitt, 2014), social media profiles may be used by hiring professionals to assess the likelihood that job applicants may achieve higher levels of extra-role performance. For instance, OCBs involve behaviors that are not critical to employees' mains tasks or job, but help organizational functioning either at an individual or organizational level (Lee & Allen, 2002; Organ & Ryan, 1995). Research on social media and OCBs is generally scant (Roth et al., 2016), and preliminary evidence suggests OCBs may be difficult to accurately predict from Facebook profiles. For instance, Van Iddekinge et al. (2016) found no evidence that Facebook-based assessments of abilities (i.e., a composite of interpersonal skills, adaptability, and creativity) related to extra-role performance (r = -.07). Similarly, Zhang et al. (2020) reported near-zero or even negative correlations between both unstructured (r = -.01) and structured (r = -.22) Facebook ratings of applicants and extra-role performance. However, it is worth noting that neither study directly asked raters to predict the likelihood of the profile users performing OCBs, and thus may not reflect raters' ability to assess potential extra-role performance from Facebook profiles.

Congruently, no research to date has directly examined assessment of OCB tendencies based on LinkedIn profiles. However, in line with the RAM principles (Funder, 1995), LinkedIn may provide ample opportunities for users to explicitly outline behaviours that map onto OCBs. For instance, applicants can highlight how they contributed to the development and success of their organization or served on committees (e.g., union reps, health and safety) when describing previous work or volunteer experience. Colleagues' recommendations may also emphasize how the applicant assisted them with their work tasks or was a generally supportive colleague. In addition, Aguado et al. (2019) examined the criterion-related validity of 615 LinkedIn profiles from professionals in the information and communication technology sector. Their results demonstrated that working extra hours (including time devoted to activities related to improving their organization's effectiveness) was positively associated with three categories of LinkedIn profiles content: "social capital" (e.g., number of connections or recommendations received), "interest in updating knowledge" (e.g., education), and "non-professional information" (e.g., volunteering, interests). Overall, based on the RAM principles and the literature presented above, we propose that raters can accurately assess OCB tendencies. Again, using multiple data sources, we examine this with two samples of raters and relying on two measures of OCB (i.e., self-reports of OCBs in the past and a situational judgment test of OCB tendencies).

Hypothesis 6. Raters' assessments of OCB tendencies demonstrate inter-rater reliability.

Hypothesis 7. Raters' assessments of OCB tendencies correlate with (a) targets' self-reports; and (b) test scores.

LinkedIn-based Hiring Recommendations and Potential for Adverse Impact

A primary difference between personal (e.g., Facebook) and professional (e.g., LinkedIn) social media is the amount of personal and non-job-relevant information available to raters. Facebook has been described as a "weak" situation in which behavior is unconstrained (Hartwell & Campion, 2020). As a result, Facebook users regularly post pictures, updates, or comments providing large amounts of information about their demographic characteristics (e.g., age, gender, race, sexual orientation), religious or political preferences, or even behaviors potentially conceived as "red flags" such as profanity or drug use (e.g., Hartwell & Campion, 2020; Tews et al., 2020; Zhang et al., 2020). From a RAM perspective (Funder, 1995), such information could help achieve more accurate judgments of job applicants. However, it is also associated with more negative evaluations (Hartwell & Campion, 2020; Tews et al., 2020), adverse impact (Van Iddekinge et al., 2016), and raises ethical and legal concerns for hiring organizations (Schmidt & O'Connor, 2016).

In contrast, LinkedIn has been described as a somewhat "stronger" situation, much more similar to traditional selection procedures (Hartwell & Campion, 2020). As a result, content posted by users is more professional and work-oriented, and hiring managers search for more positive information (vs. "red flags") to make decisions. In addition, previous work on LinkedIn-based assessments found only small-to-medium differences in hiring recommendations for race (White and non-White applicants) or gender (Roulin & Levashina, 2019). However, this work was based on students, and thus characteristics like age have not yet been examined. Given the limited evidence available, we examine the potential for adverse impact with the following research question:

Research Question 1. Are LinkedIn-based hiring recommendations associated with group differences based on (a) gender, (b) race, or (c) age?

LinkedIn Profile Content and Ratings

Although we relied largely on the first two elements of the RAM (i.e., the relevance and availability of behavioural cues; Funder, 1995, 2012) when examining the psychometric properties of LinkedIn-based assessments, the fourth element (i.e., the use of available cues in making judgments) is also important to understand what profile content is associated with judgments made by raters on social media. Previous research suggests that hiring professionals rate Facebook profiles with information about education, training, or communication skills more positively, but also rely on job-irrelevant information such as relationship status, drug or alcohol use, or sexual behaviors (Zhang et al., 2020). On the contrary, LinkedIn offers fewer cues about such personal information, but a variety of work-related cues (e.g., detailed professional experiences, list of skills, skill endorsements, recommendations). For instance, Fernandez et al. (2021) found that students in hospitality management listed 15.8 skills and had 401 professional connections on average on their LinkedIn profiles. While LinkedIn raters seem to value all work-related elements, research suggests they view profiles of students that include a profile picture, are longer, and have more connections particularly positively (Roulin & Levashina, 2019). We thus propose to replicate these findings using a sample of more experienced workers (targets) and two samples of raters:

Hypothesis 8. LinkedIn-based hiring recommendations are positively associated with (a) profile length, (b) having a profile picture, and (c) the number of professional connections.

Method

Samples

Our study included three samples: a sample of experienced workers active on LinkedIn (i.e., the *targets*), a sample of professionals with hiring experience recruited via online panels such as Mechanical Turk and Prolific (i.e., the *online panel raters*), and a sample of graduate students in Industrial-Organizational Psychology (i.e., the *I-O raters*).

Targets. We recruited 154 workers and LinkedIn users, including 88 U.S. residents via Amazon Mechanical Turk, 47 U.S. residents via Prolific, and 19 Canada residents through a university alumni association. The mean age of targets was 40.97 (*SD* = 9.49). The sample was 55% male and 45% female, 54% White, 17% Black, 19% Asian, 6% Hispanic, and 4% Others, and largely university or college educated (88%). Participants had on average 18.37 years of work experience (*SD* = 10.16) and 90% were currently employed. All respondents successfully responded to attention check items randomly embedded in the survey ("I have travelled to Mars", "I can walk on water", and "I eat glass").

Online Panel Raters. We recruited 200 U.S. or Canada residents through Prolific (170) and Mechanical Turk (30), who were pre-screened for having experience hiring personnel. The mean age of raters was 41.59 (*SD* = 11.66). The sample was 56% male (44% female) and mostly White (79%, with 6% Black, 8% Asian, 4% Hispanic, and 3% Others). Participants had on average of 7.91 years of experience in human resources and/or hiring (*SD* = 7.92), 93% were currently employed, and 74% had a managerial role. Most raters where somewhat familiar (15%), familiar (53%), or very familiar (31%) with LinkedIn prior to the study, and most used LinkedIn as part of the hiring process in their organization (35% sometimes, 22% regularly, 9% all the time). All respondents successfully responded to one attention check item embedded in the demographics section (i.e., please select "somewhat agree").

I-O Raters. We recruited 6 graduate students in I-O Psychology (one recent PhD graduate, four PhD students, one senior MSc student; 2 females, 4 males) at the same North American university, with extensive knowledge and training in personnel selection and personality assessment.

Procedure

Targets were recruited on Mechanical Turk or Prolific (compensated USD$5 in total) or through social media posts by an alumni association (compensated with a USD$15 e-gift card). We screened participants to recruit only those who were 30 years or older and had prior work experience (at least five years of work experience for alumni, at least one year of work experience for MTurk or Prolific participants). This was done to ensure that participants had more content included on their LinkedIn profiles, as compared to inexperienced participants used in prior studies (e.g., Roulin & Levashina, 2019). We also ensured that all targets were already active on LinkedIn prior to the study (e.g., using screeners on Prolfic or on the CloudResearch platform; Litman et al., 2016). We then followed an approach largely similar to recent work on social media assessments (e.g., Roulin & Levashina, 2019; Van Iddekinge et al., 2016) to obtain self-reported data, access profiles, and get ratings. That is, targets were initially asked to complete an online questionnaire including a self-reported measure of personality, a cognitive ability test, a situational judgment test (SJT) of OCB, and self-reports of past OCB behaviors.[1] They also provided demographic information. Participants were asked to provide their full name as part of the survey and were also instructed to connect with a LinkedIn profile created specifically for the study. The sole purpose of this mock profile was to have all 154 "targets" in our list of "connections" and to access their full LinkedIn profile (so that their profile could be evaluated by raters, automated assessments, or content-coded). Note that the only way for a participant to connect with our study profile on LinkedIn was to be logged into their own profile and send us a connection request. We also ensured that the names on the LinkedIn profiles matched those provided in the surveys.[2] Targets' LinkedIn profiles were then saved as screenshots (so that all key elements - e.g., pictures, connections, endorsements, etc. - could be visible to raters) as well as pdf documents (for automated assessments based on textual data only).

We recruited two groups of online panel raters. Mechanical Turk raters were asked to view 10 profiles (randomly selected from all target profiles and randomly presented to avoid ordering and contrast effects) and assess the target's personality, cognitive ability, likelihood of

[1] We recruited our sample in various phases. An initial sample of 100 targets was recruited (alumni participants in March/April 2019 and MTurk participants in January 2020), with all measures collected at time 1 and the OCB SJT data collected in April 2020 in exchange for extra compensation (included in the total compensation reported above – with 80 returning respondents). Importantly, OCB SJT scores did not differ between those who completed the measures separately (M = 9.68) vs. together (M = 9.56), F = 0.146, p = .61. Because this initial sample included a somewhat small percentage of minority participants (and to properly test for group differences and potential adverse impact), we recruited an additional 54 targets, using MTurk/Prolific screeners to only recruit minority respondents in January 2021. For that sub-sample of targets, all measures were completed at the same time.

[2] All this was done to ensure that the LinkedIn profiles truly belonged to the participants (i.e., the targets). Although asking for participants' names (or reviewing their LinkedIn profile) can be considered as contrary to MTurk or Prolific terms of services, we made that very explicit in our study description on the platforms and in our informed consent form. We also included the instruction to provide a name and connect on LinkedIn at the very start of the study, so that participants uncomfortable with this could self-select out. This also explains the somewhat high attrition (or "return") rate, which was 42% for the MTurk sample and 29% for the Prolific sample. We also coded the work experience on a sub-sample of 33 LinkedIn profiles and the number of years of experience correlated .71 (p < .001) with self-reported years of experience, providing additional evidence that targets provided their real LinkedIn profiles.

OCB, and hiring recommendations (compensated USD$7.50), whereas Prolific raters were asked to view/rate 5 profiles and rated the same elements (compensated USD$4.50). Raters were instructed to act as a hiring manager for a large North-American company doing an initial screening of applicants for a general management position. We included a timer on each rating page, so that raters had to spend at least 2 to 3 minutes on each profile (depending on profile length) before moving to the next one. In total, each profile was assessed by a minimum of 6 raters (and we only use the data from the first 6 raters in our analyses for consistency).

The six I-O raters rated all 154 profiles in seven batches, with batches ranging from 20-28 profiles each. Like the groups of online panel raters, the profiles were also randomly presented to avoid ordering and contrast effects. They were instructed to rate no more than one batch per day and to take a break approximately every thirty minutes while assessing a batch (to avoid fatigue). The six I-O raters were compensated USD $250 each. Each LinkedIn profile was later coded for various content elements (see *Measures*) by a trained research assistant. Finally, we used the Receptiviti platform to obtain automated assessments of personality.

Measures

Personality Self-reports. Targets' personality was measured using the HEXACO-60 (Ashton & Lee, 2009), which captures six traits with ten items each (i.e., Honesty-Humility, Emotionality, Extraversion, Agreeableness, Conscientiousness, and Openness). Responses were provided on a 5-point Likert scale (1 = strongly disagree to 5 = strongly agree). Example items include "I rarely hold a grudge, even against people who have badly wronged me" (Agreeableness) or "I plan ahead and organize things, to avoid scrambling at the last minute" (Conscientiousness). Reliabilities for the HEXACO-60 are presented in Table 1, and are similar to those reported in previous work (Ashton & Lee, 2009).

Personality Ratings. All raters assessed the same six personality traits of the targets based on their LinkedIn profile (1 = very low to 5 = very high), with one item per trait (similar to the approach used by Roulin & Levashina, 2019). Each trait was presented alongside a few example adjectives from the HEXACO definitions. For instance, Honesty-Humility was presented together with “sincerity, fairness, greediness, modesty” and Extraversion with “social, optimistic, energetic”.

Automated Assessments of Personality. We obtained automated assessments of the FFM personality using the Receptiviti Language-based Personality Insights, which is based on the Linguistic Inquiry and Word Count system (LIWC; Pennebaker et al., 2015) and the extensive literature examining the relationships between language and personality (e.g., Pennebaker et al., 2003; Yarkoni, 2010).[3] Importantly, while the HEXACO and FFM are not completely equivalent, there is extensive overlap between measures of the HEXACO and the FFM of personality, especially when it comes to extraversion (see Ashton & Lee, 2019 for a review).

Cognitive Ability Scores. Targets’ cognitive ability was obtained using the 16-item version of the International Cognitive Ability Resource (ICAR) test (Condon & Revelle, 2014). The ICAR is a public measure of cognitive ability that has been used in over 70 studies since its creation (see Dworak et al., 2021 for a review). It demonstrated convergent validity with the WAIS-IV (Young & Keith, 2020) as well as various ability and achievement tests, such as the SAT, ACT, and GRE (Condon & Revelle, 2016). There is also evidence of measurement

[3] For more information, see: https://docs.receptiviti.com/the-receptiviti-api. We also obtained automated assessments from a second platform, IBM Watson Personality Insights. Yet, this platform only scores text inputs with minimum 100 words. Moreover, IBM decided to discontinue this service in the middle of our data collection and analysis. Because of this, we could only score 91 of the 154 LinkedIn profiles collected. We thus only present the data from Receptiviti in the manuscript, but include those with IBM Watson in our online supplement available on the open science framework page for this project. See:https://osf.io/5rzgd/?view_only=d1dc29f9c55749e3bd1900f0d9a46763

invariance for age and sex (Young et al., 2019). The ICAR includes four types of items (i.e., letter and number series, matrix reasoning, verbal reasoning, and three-dimensional rotation), with scores ranging from 0 to 16.

Cognitive Ability Ratings. Raters assessed targets' cognitive abilities from 1 = very low to 5 = very high using one item: "Based on your assessment of his/her LinkedIn profile, how would you rate the candidate's cognitive abilities or intelligence".

OCB – Reported Behaviors. Targets' self-reports of organizational citizenship behaviors (OCB) were measured using a 16-item OCB performance measure (Lee & Allen, 2002). This measure included both individual-focused (OCBI; e.g., "help others who have been absent") and organization-focused (OCBO; e.g., "express loyalty toward the organization") items, and responses were provided on a 7-point Likert scale about how often they perform each behavior (1 = never to 7 = always).

OCB Tendencies. To obtain a test-based measure of targets' tendencies to engage in OCB, we developed an OCB situational judgement test following best-practices in SJT development (Lievens & Motowidlo, 2016). Building on the content of items from Lee and Allen (2002), we created eight SJT questions (four OCBI, four OCBO). Each question involved a brief workplace-related situation followed by four possible behavioral responses that respondents ranked from best to worst. Both items and response options were presented in a randomized order. Each SJT item maximum score was 12 points (see Appendix A). Before using the OCB SJT, we performed two preliminary stages of content validation. First, eight Subject Matter Experts (SMEs; graduate students in Industrial-Organizational Psychology) were asked to review each item, rank the four response options from worst (i.e., lowest demonstration of OCB) to best (i.e., highest demonstration of OCB), and indicate whether each question was measuring OCBI or OCBO. SMEs also provided suggestions to clarify some items (e.g., reduce jargon, ensure gender neutrality). After making the necessary adjustments, the revised SJT was then pilot-tested with an independent sample of 52 Mechanical Turk participants, who completed both the 8-item SJT (α = .79) and the 16-item (α = .91) self-report OCB measure (Lee & Allen, 2002). The correlation between the two was r = .47, p = .001 (with r = .35, p = .01 for OCBI only and r = .57, p < .001 for OCBO only).

OCB Likelihood Ratings. Raters were provided with four OCB items to assess how likely they think the target is to engage in OCB. We selected the two items for OCBI (e.g., "give up their time to assist or help coworkers who have work or non-work problems.") and OCBO (e.g., "express loyalty towards the organization and defend its image, e.g., when problem arise, when people criticize it") with the highest factor loadings in Lee and Allen (2002). Raters were asked to indicate how much they agreed that the target would perform each behavior at work (1 = strongly disagree to 5 = strongly agree).

Hiring recommendations. Raters completed a 5-item measure of hiring recommendation (Roulin & Levashina, 2019). Items (e.g., "I would be willing to hire this applicant") were rated on a 5-point Likert scale (1 = strongly disagree to 5 = strongly agree).

Profile content. All LinkedIn profiles were coded by a trained research assistant for content elements, similar to Roulin and Levashina (2019). We coded profile length (i.e., number of words in the profile, after using the "save as pdf" LinkedIn function). We also coded whether the target had a professional profile picture (yes = 1 /no = 0), included a summary section (yes = 1 /no = 0), the number of professional connections (LinkedIn only shows values up to 500), the number of skills listed on the profile, the number of skill endorsements received, the number of recommendations received, the number of job experiences listed on the profile, and the level of details provided about experiences (0 = no information, 1 = only list job titles and organization, 2 = provided detailed information about the job profile, key responsibilities, and/or accomplishments).

Results

LinkedIn-based Personality Assessments

Inter-rater agreement was computed using intraclass correlation coefficients (ICCs), which are presented in Table 1, together with descriptive statistics and internal consistency reliabilities. As in Roulin and Levashina (2019) and suggested by Shrout and Fleiss (1979), ICCs for the hiring professionals were calculated utilizing one-way random with average ratings, ICC(1, 6). This is because the 6 raters for each profile were randomly drawn from a total of 200 potential raters. In contrast, the ICCs for the I-O raters were calculated utilizing two-way random approach, ICC(2, 6), because the same six raters assessed all profiles.[4] For personality, ICCs were generally average to high for both the hiring professionals (.63 to .78, with the exception of emotionality with ICC = .07) and the I-O raters (.67 to .87, with the exception of emotionality with ICC = .14), thus providing general support for Hypotheses 1.[5]

We examined convergent validity by correlating the various types of scores collected for personality, cognitive ability, and OCB tendencies: targets' self-reports or test scores, raters' LinkedIn-based assessments, and automated assessments using Receptiviti (for personality only). The main correlations between the various sources of data are presented in Table 2.[6] We found

---

[4] To facilitate comparisons between samples or with other studies, we also report the equivalent ICC(1,1) in Table 1.
[5] Several raters commented on the difficulty of interpreting the construct of emotionality, likely because of the emotionality label, the adjective presented (e.g., anxious, worried, easily stressed), and that fact that it was the only "negative" personality trait to be assessed.
[6] Given the large number of variables and sources examined, we did not include a table with correlations between all variables in the manuscript. But, for the sake of transparency, we provide one in a spreadsheet format on the OSF

some evidence for the convergent validity of LinkedIn-based ratings of honesty-humility for both the hiring professionals ($r = .17$, $p = .03$) and the I-O raters ($r = .16$, $p = .05$). Although consistent with Hypothesis 2b, we note that these validities are considered low in practice. In contrast, we found no support for Hypothesis 2a, with no relationship for extraversion for both the hiring professionals ($r = .03$, $p = .67$) and the I-O raters ($r = .08$, $p = .35$). As expected, there was no evidence of convergent validity for the other personality traits, with the exception of emotionality ($r = .37$, $p < .001$ for the hiring professionals; $r = .16$, $p = .04$ for the I-O raters).

The relationships between self-reports and automated assessments of personality using Receptiviti provided no support for Hypothesis 3. Targets' self-reports of extraversion were unrelated to automated assessments ($r = .08$, $p = .35$). As expected, there was no evidence of convergent validity for the other personality traits, with the exception of a weak relationship for conscientiousness ($r = .14$, $p = .10$).

Interestingly, we found strong correlations between the two groups of raters for all personality traits (r ranging from .62 for honesty-humility to .82 for conscientiousness, all $p <$ .001), with the exception of emotionality ($r = .12$, $p = .13$), which suggests that the online panel of hiring professionals and the graduate I-O students perceived the targets' personality similarly. Finally, we observed some relevant correlations between raters' personality assessments and the automated assessments using Receptiviti. For instance, we found significant positive correlations for both hiring professionals and I-O raters for extraversion ($r = .45$ and .39, both $p < .001$),

page for this project. We also included a simplified version of our dataset with all key variables. See:https://osf.io/5rzgd/?view_only=d1dc29f9c55749e3bd1900f0d9a46763 agreeableness ($r = .20$, $p = .01$; $r = .27$, $p = .001$), and conscientiousness ($r = .39$ and .41, both $p < .001$).[7]

LinkedIn-based Cognitive Ability Assessments

The ICCs for LinkedIn-based assessments of cognitive ability were high for both the hiring professionals (.79) and the I-O raters (.88), suggesting that cognitive ability can be reliably assessed on this platform and supporting Hypothesis 4. In addition, we found modest evidence of convergent validity for cognitive ability both with the hiring professionals ($r = .26$, $p < .01$) and the I-O raters ($r = .18$, $p = .03$), thus providing some support for Hypotheses 5. Cognitive ability assessments by the two groups of raters were also highly consistent ($r = .77$, $p < .01$).

LinkedIn-based Predictions of OCB

The ICC for LinkedIn-based assessments of the likelihood to engage in OCB were also high for both the hiring professionals (.74) and the I-O raters (.78), supporting Hypothesis 6. We also found a high ICC for hiring recommendations (.83 and .90). Finally, we found no relationship between raters' assessments of OCB likelihood with either the self-reported OCBs ($r = .06$, $p = .47$ for the hiring professionals; $r = .06$, $p = .48$ for the I-O raters), or the SJT OCB ($r = .12$, $p = .17$ for the hiring professionals; $r = .02$, $p = .86$ for the I-O raters). Together, these results provide no support for Hypotheses 7a or 7b. Interestingly, despite the very small correlations with self-reports or SJT scores, assessments of OCB tendencies by the two groups of raters were highly consistent ($r = .75$, $p < .001$). We also note that self-reports of past OCBs

[7] We also replicated all the analyses presented in Table 2 using partial correlations, controlling for profile length (since lower convergent validity could be due to some of the profile including limited information). We found the same pattern of results.

were associated with the SJT of OCB tendencies (r = .50, p < .001), thus providing additional support for the construct validity of our new SJT measure of OCB tendencies[8].

Group Difference and Potential Adverse Impact

Related to our Research Question, we examined group differences (and thus the potential for adverse impact) based on key demographic characteristics using the average hiring recommendations by both types of raters. We found no differences based on gender for the hiring professionals, with similar recommendations for male (M = 3.44, SD = .84) and female (M = 3.49, SD = .77) targets, F(1, 152) = 0.11, p = .74, d = .06. These results were confirmed with the I-O raters, with similar recommendations for male (M = 3.17, SD = .85) and female (M = 3.26, SD = .83) targets, F(1, 152) = 0.44, p = .51, d = .11. We also found no differences based on race for the hiring professionals, with similar recommendations for White (M = 3.44, SD = .78) and non-White (M = 3.47, SD = .84) targets, F(1, 152) = 0.05, p = .83, d = .03. This was also observed for the I-O raters, with similar recommendations for White (M = 3.25, SD = .80) and non-White (M = 3.17, SD = .88) targets, F(1, 152) = 0.31, p = .58, d = .09. We note that for both the hiring professionals and I-O raters, examination of hiring recommendations by race showed higher ratings for Asian (M = 3.95, SD = 0.57; M = 3.64, SD = 0.68), followed by White (M = 3.44, SD = .78; M = 3.25, SD = 0.80), Black (M = 3.19, SD = .78; M = 2.79, SD = .72), and Hispanic (M = 3.16, SD = 1.01; M = 3.10, SD = 1.19) targets. However, post-hoc tests showed that only the Asian-White (p < .01) and Asian-Black (p < .01) differences were significant for hiring professionals. Only the Asian-Black (p < .01) difference was significant for I-O raters.

---

[8] Although we designed our OCB SJT to include four items for OCB-I and four items for OCB-O, a confirmatory factor analysis suggested that a 1-factor structure (X^2 = 19.93, p = .46, RMSEA = .00, SRMR = .05, CFI = 1.00, TLI = 1.00) was a much better fit with our data than a 2-factor structure (X^2 = 42.22, p = .003, RMSEA = .10, SRMR = .15, CFI = .81, TLI = .72).

Importantly, these analyses are based on very small samples for each minority group (i.e., 26 Black, 10 Latino, and 29 Asian participants) and should thus be interpreted with caution. Finally, target's age was unrelated to hiring recommendations, with $r = .00$, $p = .99$ for hiring professionals and $r = -.06$, $p = .47$ for I-O raters.[9]

Although not directly related to adverse impact, we also examined potential differences for other target characteristics. We found that, for both hiring professionals and I-O raters, hiring recommendations were higher for college-educated ($M = 3.54$, $SD = .78$; $M = 3.28$, $SD = .83$) than non-college-educated ($M = 2.91$, $SD = .78$; $M = 2.73$, $SD = .74$) targets, $F(1, 152) =$ 10.23/7.02, $p = .002/.001$, $d = .80/.58$, yet unrelated to years of work experience ($r = -.05$, $p =$ .57; $r = -.06$, $p = .49$).

Profile Content and Ratings

In our final set of analyses, we examined correlations between various content elements of LinkedIn profiles and targets' self-reports or test scores, as well as LinkedIn-based human and automated ratings of personality, cognitive ability, and OCB. Results are presented in Table 3. When looking at LinkedIn-based ratings by hiring professionals, we found significant correlations between the vast majority of profile elements and ratings of personality (except for emotionality), cognitive ability, OCB, and hiring recommendations. For instance, targets with longer profiles were rated as higher on most personality traits ($r = .46$ to $.61$, except for emotionality $r = -.05$), cognitive ability ($r = .53$), OCB ($r = .50$), and received higher hiring recommendations ($r = .52$). When looking at correlates of hiring recommendations specifically,

[9] Following the suggestion from a reviewer, we also examined potential adverse impact based on race, gender, and age using the 4/5th rule in four different hiring scenarios (i.e., hiring targets with average recommendations higher than 3 or 4, as well as hiring targets with the top 25% or 10% recommendations, using ratings from both groups of raters). Results are presented in Appendix B. Out of the 24 scenarios, only 4 (one for race, 3 for age) showed some evidence of potential adverse impact against the minority group.

the number of connections (r = .55), detailed work experiences (r = .54), number of skills listed (r = .50), or having a professional picture (r = .49) emerged as the most important elements for hiring professionals, whereas relationships for skills endorsements (r = .26) or recommendations (r = .23) appeared to be somewhat less relevant. The same pattern emerged for I-O raters (see Table 3). Overall, these findings provide support for Hypothesis 8.

In sharp contrast, very few profile content elements were significantly associated with targets' self-reports or test scores. For instance, targets lower on self-reported agreeableness had fewer connections (r = -.19), and those lower on extraversion listed fewer skills (r = -.18). Targets who reported engaging in less OCB (r = -.33) or scoring lower on the OCB SJT (r = -.18) received more skill endorsements.

We also found several significant correlations between automated assessments of personality using Receptiviti and LinkedIn profile elements (see bottom of Table 3). For instance, the Receptiviti language-based algorithm scored longer profiles as more conscientious (r = .33), but less open (r = -.18). More generally, we found some substantial correlations with extraversion and conscientiousness (both positively associated with having a professional picture, a profile summary, more connections, or more skills listed in the profile).

Discussion

Main Findings and Theoretical Contributions

Our study contributes to the fast-growing body of research on social media assessments and cyber-vetting in a number of important ways. First, our findings confirm that LinkedIn might *not* be the ideal platform to assess applicants' personality, at least when using a rating approach like the one used here. Previous studies based on student raters and largely student targets reported moderate inter-rater agreement but very limited evidence of convergent validity, with significant (but small) correlations only for extraversion among the FFM (Roulin & Levashina, 2019; van de Ven et al., 2017). Our findings suggest that both experienced hiring professionals and I-O graduate students (i.e., with less hiring experience, but arguably a stronger understanding of personality assessment) can achieve high levels of agreement when rating the personality of experienced applicants (with ICCs using six raters ranging from .63 to .87, except for emotionality). Yet, we found very limited evidence of convergent validity for the six HEXACO traits, when ratings were compared to self-reports. Specifically, we found only some weak evidence for honesty-humility (r = .17; .16). We also observed somewhat stronger validity correlations for emotionality (r = .37; .16), but would consider those as most likely spurious (or practically less relevant) given the extremely low ICCs for that trait, and direct comments from several raters about the difficulty and confusion when trying to assess this trait.

Building on the first two principles of the RAM (Funder, 1995, 2012), we expected traits for which more information can be found on LinkedIn, like extraversion or honesty-humility, to be more accurately assessed. The very limited evidence found in our study suggests that cues to personality could be particularly difficult to access based on the profiles of experienced workers or applicants, whose content might be mostly about skills and work experiences. In contrast, the profiles of students or younger applicants might be more about academic achievements, volunteering, or extra-curricular activities, all of which are associated with personality (Cole et al., 2003). Another explanation could be that more experienced applicants engage in more impression management on their profile (Myers et al., 2021; Roulin & Levashina, 2016), and so the personality they depict on LinkedIn is less consistent with their true self. For instance, applicants might present themselves as more extraverted, agreeable, conscientious, or honest and humble than they truly are on their profile. This would explain the high ICCs and very strong correlations observed between the two groups of raters (with the clear exception of emotionality), but the lower correlations between raters and self-reports.

A detailed examination of correlations using aggregated ratings across all online panel raters or I-O raters (see the detailed correlation table in the online supplement) showed strong relationships between raters' assessments of different personality traits. For instance, when using averaged ratings across the six I-O raters (and excluding the unreliable emotionality ratings), correlations between aggregated assessments ranged from .57 (extraversion and conscientiousness) to .77 (honesty-humility and agreeableness). At first glance, this might suggest a "halo" effect or a lack of discriminant validity, a problem observed with other selection methods too (e.g., assessment centers; Lance, 2008). However, correlations within raters were generally smaller, although they differed by rater (e.g., the within-rater correlations between extraversion and conscientiousness ranged from -.10 and .51, and those for honesty-humility and agreeableness ranged from .30 and .66). In terms of RAM principles, this suggests that (many) LinkedIn profiles might not include enough detailed information. And, discussions with the I-O raters confirmed that they sometimes struggled to make precise and unique judgments about specific traits. As such, raters might be forced to form an overall (positive or negative) impression of each target, and used it to anchor their ratings of specific personality traits. This might reinforce inter-rater agreement, but negatively impact both discriminant and convergent validity. More generally, it is also possible that LinkedIn profiles offer opportunities to assess applicants' personality traits, but that the measurement approach used here was not optimal (see *Limitations* sections below).

Our study was also the first to explore automated personality assessments of LinkedIn profiles, an approach previously restricted to personal social media like Facebook or Twitter (Alexander et al., 2020; Tay et al., 2020). We found very weak correlations between language-based automated assessments using Receptiviti and targets' self-reports. As with human ratings, these findings could suggest that LinkedIn is likely not the ideal platform to assess personality automatically, differing from findings that Facebook shows somewhat more promising results (slightly higher correlations, and for most FFM traits; Tay et al., 2020). Importantly, our study compared various sources of data with automated assessments (self-reports or test scores, and two groups of human raters), whereas most of the existing research on Facebook has compared automated assessments with self-reports only (Tay et al., 2020). One exception is Park et al. (2015) who also examined informant reports (i.e., targets' friends). Interestingly, they found slightly smaller correlations between language-based, automated assessment and informant reports than automated and self-reports. In contrast, recent work on automated assessments of personality in video interviews (Hickman et al., 2021) found small and rather inconsistent correlations between automated ratings and self-reports (e.g., within-sample *r*s ranging from -.11 to .33 across three studies) or self-reports and human raters/interviewer (*r*s = .04 to .43), but much larger and consistent correlations between automated ratings and human raters (*r*s = .24 to .74). More generally, Hickman et al. concluded that machine learning models trained on interviewer-reports outperformed those trained on self-reports in terms of convergent validity, but also discriminant validity and generalizability. Our findings seem generally aligned with Hickman et al.'s, with moderate positive correlations between automated assessments using Receptiviti and raters' assessments of extraversion, agreeableness, conscientiousness based on LinkedIn profiles (e.g., *r*s = .20 to .45), but much smaller and inconsistent relationships when comparing automated assessments with self-reports (or raters' scores and self-reports). This suggests that there may be potential for automated assessments of LinkedIn profile to partly replicate human ratings. Yet, the very limited validity obtained with human ratings suggest that replicating such assessments should perhaps not be the objective. That said, as we discuss in more details in our *Limitations* section, the tool we used for automated assessment (Receptiviti Language-based Personality Insights) was validated using textual information from Facebook or Twitter, and another tool optimized for LinkedIn content might have achieved better results.

Our study further confirmed that cognitive ability can be reliably assessed with just one item based on LinkedIn profiles (ICC = .79; .88), and that assessments demonstrate modest convergent validity with test scores (r = .26 and r =.18). These findings are relatively consistent both with the RAM principle (Funder, 1995) given the visibility of cues of cognitive ability on LinkedIn profiles, and preliminary results with students (Roulin & Levashina, 2019). The slightly lower validity coefficients observed in our study compared to prior student data (Roulin & Levashina, 2019) might be because experienced workers could be more selective about the information they include on their LinkedIn profile, either to make a more positive impression on others or to match their self-views. Yet, an alternative explanation may be the use of different cognitive ability tests (ICAR vs. Wonderlic). Additional exploratory analyses showed that college-educated targets obtain higher scores on the cognitive ability tests, and both group of raters also rated college-educated targets as higher on cognitive abilities[10]. This suggest that

[10] College-educated targets (M = 7.50, SD = 3.41) obtained slightly higher ICAR test scores than the non-college-educated group (M = 5.89, SD = 3.68), $t(152) = 1.87$, p = .06. And, our panel of hiring professional raters evaluated college-educated targets (M = 3.69, SD = 0.66) as higher on cognitive abilities than the non-college-educated targets (M = 3.12, SD = 0.60), $t(152) = 3.46$, p < .001. The same was true for the I-O raters (M = 3.34, SD = 0.68 vs. M = 2.69, SD = 0.48), $t(152) = 3.94$, p <.001).

raters might (correctly) rely on information about educational attainment on targets' LinkedIn profiles as signal of their cognitive abilities.

In terms of OCB intentions, our findings show that LinkedIn-based ratings of the likelihood to engage in OCB demonstrated strong inter-rater agreement (ICC = .75). However, they were not associated with self-reports of past OCBs or with scores on our newly-developed SJT of OCB tendencies. This suggests that it might not be possible for hiring professional to assess OCB tendencies based on LinkedIn profiles, which is consistent with previous attempts using Facebook (Van Iddekinge et al., 2016; Zhang et al., 2020). That said, although prior studies and meta-analyses suggest that self-reports of OCB can be valid, they can also be associated with impression management or social desirability (Allen et al., 2000; Carpenter et al., 2014). Our SJT OCB might also be prone to impression management, and future research might want to explore alternative sources of OCB data, such as from coworkers or supervisors.

Another important finding of this study is the very low potential for adverse impact of LinkedIn-based hiring recommendations. Indeed, we found very small differences in hiring recommendations from both hiring professionals and I-O raters for gender (d = .06 and .11) and race (d = .03 and .09), as well as virtually no relationship with age (r = .00 and -.06). These results with two types of raters and experienced workers as targets supplement those obtained by Roulin and Levashina (2019) with students. They are also in sharp contrast with previous research highlighting likelihood of protected information being visible and the higher risks of adverse impact for Facebook-based assessments (Van Iddekinge et al., 2016; Zhang et al., 2020). The limited group differences and lower potential for adverse impact on LinkedIn as compared to Facebook is also aligned with applicants' more positive attitudes or reactions towards LinkedIn vs. Facebook (Cook et al., 2020; Stoughton, 2016; Stoughton et al., 2015). Of course, these promising findings are specific to the measurement of hiring recommendations used in the study, and it is possible that other approaches to assess LinkedIn profiles might lead to higher adverse impact.

Our study further confirms that some LinkedIn profile elements are more strongly associated with positive hiring recommendations by raters than others. Profiles that were longer, included more professional connections, more skills listed, or included a professional picture were viewed more positively by hiring professionals and I-O raters, whereas receiving more skill endorsements or recommendations appeared to be somewhat less important for raters. These findings confirm earlier results obtained with MBA student raters and profiles of undergraduate students (Roulin & Levashina, 2019). However, the profile elements most valued by raters were either not associated or mostly negatively associated with targets' actual personality traits, abilities, or behaviors (based on self-reports or test scores). More precisely, targets with more professional connections were less agreeable and those with more skills listed were less extraverted. No other relationships were found for the profile elements values by raters. This suggests that these elements might not represent valid indicators of applicants' qualities (or at least their personality, cognitive abilities, and OCB tendencies). Interestingly, some additional profile elements could represent more valid signals of applicant qualifications. For instance, consistent with past research (e.g., Berry et al., 2006), additional analyses showed that targets' self-reported level of educational attainment was slightly but positively associated with their cognitively ability test scores (r = .23, p < .01). It was also related to LinkedIn-based ratings of cognitive abilities from online panel raters (r = .27, p < .001) and I-O raters (r = .30, p < .001), confirming that our raters likely used education credentials on profiles to estimate targets' cognitive abilities.

Interestingly, we found that targets who had received more skill endorsements reported engaging in less OCB or OCB tendencies (both based on self-reports and SJT scores). This suggests that employees who engage in OCB and are motivated by prosocial values or organizational concerns ("good soldiers", see Bolino, 1999; Bourdage et al., 2012) might not pursue direct rewards for their actions, in the form of listing many skills on their profile and seeking endorsements on LinkedIn. In contrast, other employees who are more strategic or political in their actions, for instance those seeking skills endorsements to appear more impressive on LinkedIn, were likely focused on advancing their in-role performance rather than extra-role performance and might be engaging in less OCB, or perhaps doing it with impression management motives (i.e., “good actors”; Bolino, 1999; Bourdage et al., 2012). At the suggestion of a reviewer, we also examined the relationships between the number of recommendations LinkedIn users *provided* to others and OCB self-report, as well as ratings of OCB likelihood by hiring professionals and I-O raters. While the relationship between the number of recommendations given and OCB SJT scores was non-significant (r = -.08, p = .36), the relationship between the number of recommendations given negatively correlated to self-reports of OCBs (r = -.18, p = .02). In addition, it was significantly but positively correlated to I-O raters’ assessments of OCB likelihood (r = .32, p < .001). A similar pattern was observed for hiring professionals’ assessments, although it did not reach significance (r = .15, p = .06). Thus, those who gave more recommendations to others on LinkedIn may be “good actors” yet perceived by I-O raters (and to a lesser extent, hiring professionals) to be “good soldiers”. Overall, in terms of RAM (Funder, 1995, 2012), our findings suggest that some of the cues used by hiring professionals in their LinkedIn judgments (at least with the assessment approach used here) might not be relevant to key traits about applicants they are attempting to assess. And, in the case of assessing OCB tendencies via recommendations given to others, may mislead hiring professionals as to the intentions of applicants in performing OCB. Thus, hiring professionals might be better off ignoring this information altogether when assessing profiles of (experienced) job applicants.

Finally, although the automated assessments we used in this study (i.e., using Receptiviti Personality Insights) rely exclusively on language to assess personality traits, the relationships observed with LinkedIn profile elements might shed some light on the “black box” of automated assessments of personality (Alexander et al., 2020) using an example of an off-the-shelf tool. For instance, profiles that were longer, included a professional picture, a summary, and more detailed job experience were automatically assessed as higher on conscientiousness. While the language-based algorithm arguably does not directly capture those elements (and certainly could not consider pictures), such a description is consistent with important facets of conscientiousness (e.g., diligence or perfectionism; Lee & Ashton, 2004). And, it is possible that LinkedIn users who are truly more conscientious not only have a more detailed and comprehensive profile, but also use language that signals conscientiousness. Similarly, the automated scores for extraversion were positively correlated with elements such as the number of connections or having a clear profile summary, which would also be consistent with important facets of extraversion (e.g., sociable or social boldness; Lee & Ashton, 2004). However, the lack of relationship between these profile elements and targets’ self-reports or test scores suggest that more research comparing automated assessments and other sources of data is necessary (with LinkedIn and other social media platforms).

Practical Implications

Our findings, combined with the growing body of literature on social media assessments, have direct practical implications for hiring managers and organizations currently engaging in cyber-vetting or planning to do so. First, assessing LinkedIn profiles can represent a reliable source of information to make a quick assessment of applicants' cognitive ability, but validity was only moderate. Overall, asking hiring managers to estimate an applicant's cognitive level with one broad item using their LinkedIn profile might be seen as a practical, quick, and economical way to perform an initial screening. Yet, because the validity appears to be limited, organizations would be better off relying on traditional cognitive ability testing. Second, our study shows that high levels of inter-rater agreement can be achieved with LinkedIn-based assessments using six raters for all elements evaluated (and with different types of raters). While moderate levels can be achieved with as few as two raters (Roulin & Levashina, 2019), and there are certainly costs associated with asking several hiring managers to assess applicants' profiles, using more raters is likely more reliable, fair, and legally-defensible. Thus, we recommend if cyber-vetting, to employ several hiring managers to assess LinkedIn profiles to improve the reliability of ratings.

Third, LinkedIn is not the ideal source of personality assessments for experienced workers, except to some extent for honesty-humility, and organizations should be advised to rely on established personality assessments (or should develop a more structured and thus valid assessment approach than the one used in this study). While personal social media like Facebook could be a more relevant source of personality information (e.g., Kluemper et al., 2012), we caution that they may be associated with larger sub-group differences, higher potential for adverse impact, and likely more negative applicant reactions than LinkedIn (Cook et al., 2020; Van Iddekinge et al., 2016). Finally, with regard to automated assessments of personality, while our findings using Receptiviti identify relevant overlap with raters' assessment, the limited overlap with targets' self-reports suggests that more research is required before organizations can be recommended to rely on them to assess applicants' LinkedIn profiles. For instance, instead of using an off-the-shelf tool, new technologies could be designed and validated specifically to optimize assessments based on information available on LinkedIn profiles.

Lastly, the SJT for OCB developed in this study could represent a potentially relevant assessment tool to use in selection. SJTs have become popular and effective methods to assess a wide range of knowledge, traits, or attitudes (Lievens & Motowidlo, 2016). For instance, Motowidlo et al. (2016) recently proposed a measure to assess "prosocial implicit trait policy" by presenting participants with short scenarios and asking them to rate the effectiveness of the described behaviors. Our SJT relies on a format more akin to traditional SJTs, where test takers are presented with different response options and have to rank them. This approach might involve more complexity and thus be less prone to social desirability or faking. Yet, while our results are promising, a longer version of the test should be developed and validated, and more research is certainly needed (see below), before this SJT can be used for personnel selection.

Limitations and Future Research Directions

Overall, the findings of this study should be interpreted in light of a "technology-as-designed" paradigm (Landers & Marin, 2021). As such, it is important to acknowledge that our results and potential implications are bounded by the choices made in designing how raters reviewed and assessed LinkedIn profiles (i.e., traits assessed, instructions provided, measurement approach) or in terms of the technology used (i.e., Receptiviti as our automated assessment tool). We describe these limitations, and how they can be addressed in future research, below.

While previous work on LinkedIn-based assessments was limited to a small number of student raters (Roulin & Levashina, 2019; van de Ven et al., 2017), our study relied on two sources of raters: a large sample of raters with professional hiring experience and I-O graduate students with extensive knowledge in assessment. However, our hiring professionals were recruited via MTurk and Prolific. Although we tried to incentivise accuracy (with competitive compensation and by forcing raters to spend at least 3 minutes on each profile), we cannot exclude that (some) raters might only have had limited motivation to achieve high levels of accuracy in their ratings. In contrast, hiring professionals engaging in cyber-vetting as part of an actual selection process have an incentive to be as accurate as possible in their judgments to help their organization select the best applicants. While the similarities observed with the (arguably more motivated) I-O raters is reassuring, our findings for inter-rater reliability or convergent validity might still be underestimated (although the "halo" effects described above might also have counter-acted that). Future research could attempt to provide incentives to raters based on their level of accuracy. We also encourage field studies be conducted in high-stakes selection contexts.

Cyber-vetting is considered to be generally performed in an informal way (Berkelaar, 2017). Therefore, we chose to not provide training to our raters in order to increase generalizability. It has been suggested that structuring assessments (Roth et al., 2016) or providing raters with training (Schroeder et al., 2020) should help with reliability or validity. Yet, empirical evidence is rather limited. For instance, Zhang et al. (2020) found that providing raters with training and more structure did not increase inter-rater agreement or criterion-related validity in Facebook-based assessments. Roulin and Levashina (2019) found that LinkedIn-based itemized assessments led to higher inter-rater agreement than global assessments, but effects on group differences were mixed (reduced for race, increased for gender). It is also possible that, in practice, hiring managers use an even less structured approach. That is, they might not attempt to assess applicants' personality or OCB tendencies, but rather focus on job-relevant skills or simply make a global assessment of their fit with the job. As such, the task our raters performed might not completely replicate what actual managers do. In addition, the measurement approach differed between self-reports and LinkedIn-based ratings (e.g., 10 vs. 1 item per trait for the HEXACO). Although this approach is similar to past research (e.g., Kluemper et al., 2012; Roulin & Levashina, 2019; van de Ven et al., 2017), this might have contributed to the low convergent validities observed. Overall, more research is needed to examine the role of measurement, structure, and training in cyber-vetting.

Similarly, future research could examine LinkedIn-based hiring recommendations for specific positions or industries (e.g., using a job description with a list of requirements vs. a generic managerial role). Our analyses of the potential for adverse impact were based on a measure of hiring recommendations, mean differences, and effect sizes. While this is consistent with best practices used in past research (e.g., Hough et al., 2001), we did not ask our raters to make actual selection or hiring decisions. We thus could only explore other indicators used in practice (or in court), such as the 4/5th rule (Dunleavy et al., 2015), by transforming hiring recommendations scores into various cut-offs scenarios (see Appendix B). Future research could include an actual selection task to directly examine such indicators.

In addition, while we recruited a relatively large sample of targets with work experience (vs. smaller samples and largely students in previous studies), our participants were obtained via Mechanical Turk or Prolific and had somewhat short LinkedIn profiles on average (*M* = 333.40 words, *SD* = 332.38). According to the RAM (Funder, 1995), the limited amount of information available to raters might explain the lower-than-expected convergent validities for extraversion or honesty-humility. Roth et al. (2016) also proposed that the amount of information could have a curvilinear relationship with accuracy, with too little or too much information hurting assessment accuracy. Although we confirmed our findings when controlling for profile length, future research should replicate our findings with a sample of experienced targets who possess more detailed/longer LinkedIn profiles. For instance, researchers could screen participants based on the length and level of completeness of their profiles. Additionally, we relied on targets' self-reports of personality and OCB, which can be prone to bias (although we also used an SJT measure for OCB). Future studies could also attempt to obtain peer-ratings of personality (and OCB), for instance from coworkers, instead of (or in addition to) self-reports. It might also be that LinkedIn is a sub-optimal platform to assess the Big-Five or HEXACO personality traits but might be more relevant to assess other traits like achievement orientation or resilience. Future studies could explore whether higher convergent validity can be found for other traits.

We also specifically recruited a sample of targets who had an active LinkedIn profile, yet most of them (88%) were employed and it is likely that very few of these individuals were actively seeking new work opportunities (e.g., only three included the #OpenToWork green circle symbol in their profile). Future research could examine how those who are actively job seeking may include more behavioral indicators that may allow managers' to better assess the profile owner's personality. Conversely, future research could also examine hiring managers' reactions to applicants who are not at all active on LinkedIn. Such a behavior could be considered by some managers as either suspicious (i.e., hiding something) or refreshing (i.e., original), although this might also be job-dependent (Carr, 2016). Practically speaking, excluding applicants because they do not have a LinkedIn presence might also be problematic if such behavior is more prevalent for minority applicants.

The Receptiviti platform we used for automated assessment only provides personality scores (i.e., percentiles) using the FFM, which is only partly comparable to HEXACO (Ashton & Lee, 2019; Ludeke et al., 2019). Future studies could attempt to develop an algorithm to specifically capture the six personality traits (and 24 domains/facets) of the HEXACO model. Moreover, automated assessments require enough text data to provide a reliable assessment of personality, but many of the LinkedIn profiles used in this study were relatively short. As such, future research should try to replicate our findings with more detailed/longer LinkedIn profiles only. Or, it might also be that most LinkedIn profiles are generally short and thus alternative automated methods that do not require as much textual information could be explored. Because we only used automated assessments for personality (and not hiring recommendations), we did not directly examine the potential for adverse impact using this approach. Yet, more research on the fairness and potential biases associated with automated assessments is needed (Alexander et al., 2020). Thus, future research could explore whether recommendations based on automated ratings of LinkedIn profile differ based on applicants' characteristics (e.g., race, age, gender). Moreover, future research could explore whether platforms could be developed to automatically and accurately assess other traits of applicants, such as cognitive ability or the likelihood to engage in OCB, based on LinkedIn profiles.

Our study focused on the reliability and convergent validity of LinkedIn-based assessments. Yet, there are other important aspects of validity (i.e., construct, content, criterion-related) that future research should examine. For instance, although Roulin and Levashina (2019) found some evidence that LinkedIn assessments were associated with career-related outcomes and our study found no relationship with OCBs, future research could examine whether such assessment can predict job performance or counter-productive work behaviors.

The new SJT for OCB developed as part of this study may represent a valuable measure both for selection researchers and practitioners. Although we provide initial evidence of internal consistency, a clear 1-factor structure, and convergent validity with self-reports of OCB (both in our pilot study and our main study), further research is required to fully validate this measure. This might include administrating to a larger sample of employees or job applicants, and examining its relationships with other SJT assessing conceptually-similar constructs (e.g., prosocial implicit trait policy; Motowidlo et al., 2016) or with other workplace outcomes (e.g., task performance, counterproductive work behaviors, turnover). Future work could also explore its resistance to impression management or faking. For instance, Bourdage et al. (2012) have found that OCB motivated by impression management is negatively related to honesty-humility and conscientiousness, but OCB motivated by prosocial value or organizational concerns are positively related to conscientiousness, agreeableness, or openness. In our study we found positive correlations between our SJT for OCB and honesty-humility (r = .26, p = .003), conscientiousness (r = .27, p = .002), and openness (r = .22, p = .01), suggesting that our SJT measure might truly capture “good soldiers” rather than “good actors”. Future research could examine this more directly.

Conclusion

Cyber-vetting has become commonplace in personnel selection, but research is still largely lagging behind and extensively focused on personal social media. Our study examined one approach to LinkedIn-based assessments by human raters and one off-the-shelf automated assessment tool (Receptiviti). It highlighted that raters’ assessments of experienced applicants’ cognitive ability were reliable but only modestly valid, whereas assessments of personality and likelihood to engage in OCB were reliable but generally not valid. Automated assessments of personality were more strongly related with raters' assessments than self-reports. LinkedIn-based hiring recommendations were also associated with limited potential for adverse impact. Overall, we caution organizations to carefully consider the risks and benefits associated with using social media (e.g., LinkedIn) in their selection process.

References

Aguado, D., Andrés, J. C., García Izquierdo, A. L., & Rodríguez, J. (2019). LinkedIn" Big Four": job performance validation in the ICT sector. *Journal of Work and Organizational Psychology*, *35*, 53-64. https://doi.org/10.5093/jwop2019a7

Alexander, L. I., Mulfinger, E., & Oswald, F. L. (2020). Using Big Data and machine learning in personality measurement: Opportunities and challenges. *European Journal of Personality*. https://doi.org/10.1002/per.2305

Allen, T. D., Barnard, S., Rush, M. C., & Russell, J. E. A. (2000). Ratings of organizational citizenship behavior: Does the source make a difference? *Human Resource Management Review*, *10*(1), 97-114. https://doi.org/https://doi.org/10.1016/S1053-4822(99)00041-8

Amichai-Hamburger, Y., & Vinitzky, G. (2010). Social network use and personality. *Computers in Human Behavior*, *26*(6), 1289-1295. https://doi.org/https://doi.org/10.1016/j.chb.2010.03.018

Anglim, J., Lievens, F., Everton, L., Grant, S. L., & Marty, A. (2018). HEXACO personality predicts counterproductive work behavior and organizational citizenship behavior in low-stakes and job applicant contexts. *Journal of Research in Personality*, *77*, 11-20. https://doi.org/https://doi.org/10.1016/j.jrp.2018.09.003

Ashton, M. C., & Lee, K. (2005). Honesty-Humility, the Big Five, and the Five-Factor model. *Journal of Personality*, *73*(5), 1321-1354. https://doi.org/10.1111/j.1467-6494.2005.00351.x

Ashton, M. C., & Lee, K. (2009). The HEXACO–60: A short measure of the major dimensions of personality. *Journal of Personality Assessment*, *91*(4), 340-345. https://doi.org/10.1080/00223890902935878

Ashton, M. C., & Lee, K. (2019). How well do Big Five measures capture HEXACO scale variance? *Journal of Personality Assessment*, *101*(6), 567-573. https://doi.org/10.1080/00223891.2018.1448986

Back, M. D., Stopfer, J. M., Vazire, S., Gaddis, S., Schmukle, S. C., Egloff, B., & Gosling, S. D. (2010). Facebook profiles reflect actual personality, not self-idealization. *Psychological Science*, *21*(3), 372-374.

Berkelaar, B. L. (2017). Different ways new information technologies influence conventional organizational practices and employment relationships: The case of cybervetting for personnel selection. *Human Relations*, *70*(9), 1115-1140. https://doi.org/10.1177/0018726716686400

Berry, C. M., Gruys, M. L., & Sackett, P. R. (2006). Educational attainment as a proxy for cognitive ability in selection: Effects on levels of cognitive ability and adverse impact. *Journal of Applied Psychology*, *91*(3), 696-705. https://doi.org/https://doi.org/10.1037/0021-9010.91.3.696

Bleidorn, W., & Hopwood, C. J. (2019). Using machine learning to advance personality assessment and theory. *Personality and Social Psychology Review*, *23*(2), 190-203. https://doi.org/10.1177/1088868318772990

Bolino, M. C. (1999). Citizenship and impression management: Good soldiers or good actors? *The Academy of Management Review*, *24*(1), 82-98. https://doi.org/10.5465/AMR.1999.1580442

Bourdage, J. S., Lee, K., Lee, J., & Shin, H. (2012). Motives for organizational citizenship behavior: Personality correlates and co-worker ratings of OCB. *Human Performance*, *25*(3). https://doi.org/10.1080/08959285.2012.683904

Bourdage, J. S., Wiltshire, J., & Lee, K. (2015). Personality and impression management: Correlates and implications. *Journal of Applied Psychology*, *100*(2), 537-546. https://doi.org/10.1037/a0037942

Carpenter, N. C., Berry, C. M., & Houston, L. (2014). A meta-analytic comparison of self-reported and other-reported organizational citizenship behavior. *Journal of Organizational Behavior*, *35*(4), 547-574. https://doi.org/ https://doi.org/10.1002/job.1909

Carr, C. T. (2016). An uncertainty reduction approach to applicant information-seeking in social media: Effects on attributions and hiring. In N. R. Landers & B. G. Schmidt (Eds.), *Social Media in Employee Selection and Recruitment: Theory, Practice, and Current Challenges* (pp. 59-78). Springer International Publishing. https://doi.org/10.1007/978-3-319-29989-1_4

Cole, M. S., Feild, H. S., & Giles, W. F. (2003). Using recruiter assessments of applicants' resume content to predict applicant mental ability and Big Five personality dimensions. *International Journal of Selection and Assessment*, *11*, 78-88.

Cole, M. S., Feild, H. S., Giles, W. F., & Harris, S. G. (2009). Recruiters' inferences of applicant personality based on résumé screening: Do paper people have a personality? *Journal of Business and Psychology*, *24*(1), 5-18. https://doi.org/https://doi.org/10.1007/s10869-008-9086-9

Collmus, A. B., Armstrong, M. B., & Landers, R. N. (2016). Game-thinking within social media to recruit and select job candidates. In N. R. Landers & B. G. Schmidt (Eds.), *Social Media in Employee Selection and Recruitment: Theory, Practice, and Current Challenges* (pp. 103-124). Springer International Publishing. https://doi.org/10.1007/978-3-319-29989-1_6

Condon, D. M., & Revelle, W. (2014). The International Cognitive Ability Resource: Development and initial validation of a public-domain measure. *Intelligence*, *43*, 52-64. https://doi.org/10.1016/j.intell.2014.01.004

Condon, D. M., & Revelle, W. (2016). Selected ICAR data from the SAPA-Project: Development and initial validation of a public-domain measure. *Journal of Open Psychology Data*, *4*(1). https://doi.org/http://doi.org/10.5334/jopd.25

Connolly, J. J., Kavanagh, E. J., & Viswesvaran, C. (2007). The convergent validity between self and observer ratings of personality: A meta-analytic review. *International Journal of Selection and Assessment*, *15*(1), 110-117. https://doi.org/10.1111/j.1468-2389.2007.00371.x

Cook, R., Jones-Chick, R., Roulin, N., & O'Rourke, K. (2020). Job seekers' attitudes toward cybervetting: Scale development, validation, and platform comparison. *International Journal of Selection and Assessment*. https://doi.org/10.1111/ijsa.12300

Cubrich, M., King, R. T., Mracek, D. L., Strong, J. M. G., Hassenkamp, K., Vaughn, D., & Dudley, N. M. (2021). Examining the criterion-related validity evidence of LinkedIn profile elements in an applied sample. *Computers in Human Behavior*, *120*, 106742. https://doi.org/10.1016/j.chb.2021.106742

Davis, J., Wolff, H.-G., Forret, M. L., & Sullivan, S. E. (2020). Networking via LinkedIn: An examination of usage and career benefits. *Journal of Vocational Behavior*, *118*, 103396. https://doi.org/10.1016/j.jvb.2020.103396

Denzin, N. K. (2012). Triangulation 2.0. *Journal of Mixed Methods Research*, *6*(2), 80-88. https://doi.org/10.1177/1558689812437186

Dunleavy, E., Morris, S., & Howard, E. (2015). Measuring adverse impact in employee selection decisions. In *Practitioner's guide to legal issues in organizations* (pp. 1-26). Springer. https://doi.org/https://doi.org/10.1007/978-3-319-11143-8_1

Dworak, E. M., Revelle, W., Doebler, P., & Condon, D. M. (2021). Using the International Cognitive Ability Resource as an open source tool to explore individual differences in cognitive ability. *Personality and Individual Differences*, *169*, 109906. https://doi.org/https://doi.org/10.1016/j.paid.2020.109906

Fernandez, S., Stöcklin, M., Terrier, L., & Kim, S. (2021). Using available signals on LinkedIn for personality assessment. *Journal of Research in Personality*, *93*, 104122. https://doi.org/10.1016/j.jrp.2021.104122

Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, *102*(4), 652-670.

Funder, D. C. (2012). Accurate personality judgment. *Current Directions in Psychological Science*, *21*(3), 177-182. https://doi.org/10.1177/0963721412445309

Golbeck, J. A. (2016). Predicting Personality from Social Media Text. *AIS Transactions on Replication Research*, *2*(2), 1-10. https://doi.org/10.17705/1atrr.00009

Guilfoyle, S., Bergman, S. M., Hartwell, C., & Powers, J. (2016). Social Media, Big Data, and Employment Decisions: Mo' Data, Mo' Problems? In N. R. Landers & B. G. Schmidt (Eds.), *Social Media in Employee Selection and Recruitment: Theory, Practice, and Current Challenges* (pp. 127-155). Springer International Publishing. https://doi.org/10.1007/978-3-319-29989-1_7

Hartwell, C. J., & Campion, M. A. (2020). Getting social in selection: How social networking website content is perceived and used in hiring. *International Journal of Selection and Assessment*, *28*(1), 1-16. https://doi.org/10.1111/ijsa.12273

Hickman, L., Bosch, N., Ng, V., Saef, R., Tay, L., & Woo, S. E. (2021). Automated video interview personality assessments: Reliability, validity, and generalizability investigations. *Journal of Applied Psychology*.

Hickman, L., Tay, L., & Woo, S. E. (2019). Validity evidence for cff-the-shelf language-based personality assessment using video interviews: Convergent and discriminant relationships with self and observer ratings. *Personnel Assessment and Decisions*, *5*(3), 3. https://doi.org/https://doi.org/10.25035/pad.2019.03.003

Hough, L. M., Oswald, F. L., & Ock, J. (2015). Beyond the Big Five: New directions for personality research and practice in organizations. *Annual Review of Organizational Psychology and Organizational Behavior*, *2*(1), 183-209.

Hough, L. M., Oswald, F. L., & Ployhart, R. E. (2001). Determinants, detection and amelioration of adverse impact in personnel selection procedures: Issues, evidence and lessons learned. *International Journal of Selection and Assessment*, *9*(1-2), 152-194. https://doi.org/10.1111/1468-2389.00171

Johnson, M. A., & Leo, C. (2020). The inefficacy of LinkedIn? A latent change model and experimental test of using LinkedIn for job search. *Journal of Applied Psychology*. https://doi.org/https://doi.org/10.1037/apl0000491

Kluemper, D. H., & Rosen, P. A. (2009). Future employment selection methods: evaluating social networking web sites. *Journal of Managerial Psychology*, *24*(6), 567-580. https://doi.org/10.1108/02683940910974134

Kluemper, D. H., Rosen, P. A., & Mossholder, K. W. (2012). Social networking websites, personality ratings, and the organizational context: More than meets the eye? *Journal of Applied Social Psychology*, *42*(5), 1143-1172. https://doi.org/10.1111/j.1559-1816.2011.00881.x

Lance, C. E. (2008). Why assessment centers do not work the way they are supposed to. *Industrial and Organizational Psychology*, *1*(1), 84-97. https://doi.org/https://doi.org/10.1111/j.1754-9434.2007.00017.x

Landers, R. N., & Behrend, T. S. (2022). Auditing the AI auditors: A framework for evaluating fairness and bias in high stakes AI predictive models. *American Psychologist*, No Pagination Specified-No Pagination Specified. https://doi.org/10.1037/amp0000972

Landers, R. N., & Marin, S. (2021). Theory and technology in organizational psychology: A review of technology integration paradigms and their effects on the validity of theory. *Annual Review of Organizational Psychology and Organizational Behavior*, *8*, 235-258. https://doi.org/10.1146/annurev-orgpsych-012420-060843

Lee, K., & Allen, N. J. (2002). Organizational citizenship behavior and workplace deviance: The role of affect and cognitions. *Journal of Applied Psychology*, *87*(1), 131. https://doi.org/10.1037/0021-9010.87.1.131

Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO Personality Inventory. *Multivariate Behavioral Research*, *39*(2), 329-358. https://doi.org/10.1207/s15327906mbr3902_8

Lee, K., & Ashton, M. C. (2014). The dark triad, the big five, and the HEXACO model. *Personality and Individual Differences*, *67,* 2-5.

Lee, K., Ashton, M. C., Wiltshire, J., Bourdage, J. S., Visser, B. A., & Gallucci, A. (2013). Sex, Power, and Money: Prediction from the Dark Triad and Honesty–Humility. *European Journal of Personality*, *27*(2), 169-184. https://doi.org/10.1002/per.1860

Lee, Y., Berry, C. M., & Gonzalez-Mulé, E. (2019). The importance of being humble: A meta-analysis and incremental validity analysis of the relationship between honesty-humility and job performance. *Journal of Applied Psychology*, *104*(12), 1535-1546. https://doi.org/https://doi.org/10.1037/apl0000421

Levashina, J., Hartwell, C. J., Morgeson, F. P., & Campion, M. A. (2014). The structured employment interview: Narrative and quantitative review of the research literature. *Personnel Psychology*, *67*(1), 241-293. https://doi.org/10.1111/peps.12052

Liem, C. C. S., Langer, M., Demetriou, A., Hiemstra, A. M. F., Sukma Wicaksana, A., Born, M. P., & König, C. J. (2018). Psychology Meets Machine Learning: Interdisciplinary Perspectives on Algorithmic Job Candidate Screening. In H. J. Escalante, S. Escalera, I. Guyon, X. Baró, Y. Güçlütürk, U. Güçlü, & M. van Gerven (Eds.), *Explainable and Interpretable Models in Computer Vision and Machine Learning* (pp. 197-253). Springer International Publishing. https://doi.org/10.1007/978-3-319-98131-4_9

Lievens, F., & Motowidlo, S. J. (2016). Situational judgment tests: From measures of situational judgment to measures of general domain knowledge. *Industrial and Organizational Psychology*, *9*(1), 3-22. https://doi.org/https://doi.org/10.1017/iop.2015.71

Litman, L., Robinson, J., & Abberbock, T. (2016). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, 1-10. https://doi.org/10.3758/s13428-016-0727-z

Ludeke, S. G., Bainbridge, T. F., Liu, J., Zhao, K., Smillie, L. D., & Zettler, I. (2019). Using the Big Five Aspect Scales to translate between the HEXACO and Big Five personality models. *Journal of Personality*, *87*(5), 1025-1038.

Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K. R., & Schmitt, N. (2007). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology*, *60*, 683-729. https://doi.org/10.1111/j.1744-6570.2007.00089.x

Motowidlo, S. J., Ghosh, K., Mendoza, A. M., Buchanan, A. E., & Lerma, M. N. (2016). A context-independent situational judgment test to measure prosocial implicit trait policy. *Human Performance*, *29*(4), 331-346. https://doi.org/http://dx.doi.org/10.1080/08959285.2016.1165227

Myers, V., Price, J. P., Roulin, N., Duval, A., & Sobhani, S. (2021). Job seekers' impression management on Facebook: scale development, antecedents, and outcomes. *Personnel Assessment and Decisions*, *7*(1), 102-113. https://doi.org/https://doi.org/10.25035/pad.2021.01.010

Nikolaou, I. (2014). Social networking web sites in job search and employee recruitment. *International Journal of Selection and Assessment*, *22*(2), 179-189. https://doi.org/10.1111/ijsa.12067

Obschonka, M., Fisch, C., & Boyd, R. (2017). Using digital footprints in entrepreneurship research: A Twitter-based personality analysis of superstar entrepreneurs and managers. *Journal of Business Venturing Insights*, *8*, 13-23. https://doi.org/https://doi.org/10.1016/j.jbvi.2017.05.005

Oostrom, J. K., de Vries, R. E., & de Wit, M. (2019). Development and validation of a HEXACO situational judgment test. *Human Performance*, *32*(1), 1-29. https://doi.org/10.1080/08959285.2018.1539856

Organ, D. W., & Ryan, K. (1995). A meta-analytic review of attitudinal and dispositional predictors of organizational citizenship behavior. *Personnel Psychology*, *48*(4), 775-802. https://doi.org/10.1111/j.1744-6570.1995.tb01781.x

Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., Ungar, L. H., & Seligman, M. E. (2015). Automatic personality assessment through social media language. *Journal of Personality and Social Psychology*, *108*(6), 934. https://doi.org/http://dx.doi.org/10.1037/pspp0000020

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*.

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology*, *54*(1), 547-577. https://doi.org/10.1146/annurev.psych.54.101601.145041

Roth, P. L., Bobko, P., Van Iddekinge, C. H., & Thatcher, J. B. (2016). Social media in employee-selection-related decisions: A research agenda for uncharted territory. *Journal of Management*, *42*(1), 269-298. https://doi.org/10.1177/0149206313503018

Roulin, N., & Bangerter, A. (2013). Social networking websites in personnel selection: A signaling perspective on recruiters' and applicants' perceptions. *Journal of Personnel Psychology*, *12*(3), 143-151. https://doi.org/10.1027/1866-5888/a000094

Roulin, N., & Levashina, J. (2016). Impression management and social media profiles. In N. R. Landers & B. G. Schmidt (Eds.), *Social Media in Employee Selection and Recruitment: Theory, Practice, and Current Challenges* (pp. 223-248). Springer International Publishing. https://doi.org/10.1007/978-3-319-29989-1_11

Roulin, N., & Levashina, J. (2019). LinkedIn as a new selection method: Psychometric properties and assessment approach. *Personnel Psychology*, *72*(2), 187-211. https://doi.org/10.1111/peps.12296

Schmidt, G. B., & O'Connor, K. W. (2016). Legal concerns when considering social media data in selection. In N. R. Landers & B. G. Schmidt (Eds.), *Social Media in Employee Selection and Recruitment: Theory, Practice, and Current Challenges* (pp. 265-287). Springer International Publishing. https://doi.org/10.1007/978-3-319-29989-1_13

Schmitt, N. (2014). Personality and cognitive ability as predictors of effective performance at work. *Annual Review of Organizational Psychology and Organizational Behavior, 1*(1), 45-65.

Schroeder, A. N., Odd, K. R., & Whitaker, J. H. (2020). Agree to disagree: Examining the psychometrics of cybervetting. *Journal of Managerial Psychology*. https://doi.org/10.1108/JMP-09-2018-0420

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420-428. https://doi.org/10.1037/0033-2909.86.2.420

Spitzmuller, M., Van Dyne, L., & Ilies, R. (2008). Organizational citizenship behavior: A review and extension of its nomological network. In J. Barling & C. L. Cooper (Eds.), *The SAGE handbook of organizational behavior* (Vol. 1, pp. 106-123). SAGE Publications.

Stoughton, J. W. (2016). Applicant Reactions to Social Media in Selection: Early Returns and Future Directions. In N. R. Landers & B. G. Schmidt (Eds.), *Social Media in Employee Selection and Recruitment: Theory, Practice, and Current Challenges* (pp. 249-263). Springer International Publishing. https://doi.org/10.1007/978-3-319-29989-1_12

Stoughton, J. W., Thompson, L. F., & Meade, A. W. (2015). Examining applicant reactions to the use of social networking websites in pre-employment screening. *Journal of Business and Psychology*, *30*(1), 73-88. https://doi.org/10.1007/s10869-013-9333-6

Tay, L., Woo, S. E., Hickman, L., & Saef, R. M. (2020). Psychometric and validity issues in machine learning approaches to personality assessment: A focus on social media text mining. *European Journal of Personality*, *34*(5), 826-844. https://doi.org/https://doi.org/10.1002/per.2290

Tews, M. J., Stafford, K., & Kudler, E. P. (2020). The effects of negative content in social networking profiles on perceptions of employment suitability. *International Journal of Selection and Assessment*, *28*(1), 17-30. https://doi.org/https://doi.org/10.1111/ijsa.12277

Tippins, N. T., Oswald, F. L., & McPhail, S. M. (2021). Scientific, legal, and ethical concerns about AI-based personnel selection tools: a call to action. *Personnel Assessment and Decisions*, *7*(2), 1. https://doi.org/10.25035/pad.2021.02.001

van de Ven, N., Bogaert, A., Serlie, A., Brandt Mark, J., & Denissen Jaap, J. A. (2017). Personality perception based on LinkedIn profiles. *Journal of Managerial Psychology*, *32*(6), 418-429. https://doi.org/10.1108/JMP-07-2016-0220

Van Iddekinge, C. H., Lanivich, S. E., Roth, P. L., & Junco, E. (2016). Social media for selection? Validity and adverse impact potential of a Facebook-based assessment. *Journal of Management*, *42*(7), 1811-1835. https://doi.org/10.1177/0149206313515524

Van Iddekinge, C. H., Raymark, P. H., & Roth, P. L. (2005). Assessing personality with a structured employment interview: Construct-related validity and susceptibility to response inflation. *Journal of Applied Psychology*, *90*(3), 536-552. https://doi.org/10.1037/0021-9010.90.3.536

Weidner, N., O'Brien, K. E., & Wynne, K. T. (2016). Social Media Use: Antecedents and Outcomes of Sharing. In N. R. Landers & B. G. Schmidt (Eds.), *Social Media in Employee Selection and Recruitment: Theory, Practice, and Current Challenges* (pp. 79-101). Springer International Publishing. https://doi.org/10.1007/978-3-319-29989-1_5

Wendler, K., Liu, J., & Zettler, I. (2018). Honesty-Humility Interacts With Context Perception in Predicting Task Performance and Organizational Citizenship Behavior. *Journal of Personnel Psychology*, *17*(4), 161-171. https://doi.org/10.1027/1866-5888/a000203

Yarkoni, T. (2010). Personality in 100,000 Words: A large-scale analysis of personality and word use among bloggers. *Journal of Research in Personality*, *44*(3), 363-373. https://doi.org/https://doi.org/10.1016/j.jrp.2010.04.001

Young, S. R., & Keith, T. Z. (2020). An Examination of the Convergent Validity of the ICAR16 and WAIS-IV. *Journal of Psychoeducational Assessment*, *38*(8), 1052-1059. https://doi.org/10.1177/0734282920943455

Young, S. R., Keith, T. Z., & Bond, M. A. (2019). Age and sex invariance of the International Cognitive Ability Resource (ICAR). *Intelligence*, *77*, 101399. https://doi.org/https://doi.org/10.1016/j.intell.2019.101399

Zettler, I., Thielmann, I., Hilbig, B. E., & Moshagen, M. (2020). The nomological net of the HEXACO model of personality: A large-scale meta-analytic investigation. *Perspectives on Psychological Science*, *15*(3), 723-760. https://doi.org/10.1177/1745691619895036

Zhang, L., Van Iddekinge, C. H., Arnold, J. D., Roth, P. L., Lievens, F., Lanivich, S. E., & Jordan, S. L. (2020). What's on job seekers' social media sites? A content analysis and effects of structure on recruiter judgments and predictive validity. *Journal of Applied Psychology*. https://doi.org/https://doi.org/10.1037/apl0000490

Zide, J., Elman, B., & Shahani-Denning, C. (2014). LinkedIn and recruitment: how profiles differ across occupations. *Employee Relations*, *36*(5), 583-604. https://doi.org/10.1108/ER-07-2013-0086

Table 1

Descriptive Statistics and Reliability Indicators

	Means (SD)					**Reliability**			
		LinkedIn Ratings				LinkedIn Ratings Interrater agreement			
						Online Panel Raters		I-O Raters	
	Targets' self-reports and test scores	Online Panel Raters	I-O Raters	Automated Assessment	Self-reports internal consistency	ICC (1,6)[a]	ICC (1,1)[b]	ICC (2,6)[a]	ICC (1,1)[b]
Personality									
Honesty-Humility	3.57 (.67)	3.46 (.50)	3.28 (.49)	-	.79	.67	.24	.67	.23
Emotionality	3.10 (.67)	2.62 (.40)	2.76 (.28)	28.65 (3.93)	.80	.07	.01	.14	.00
Extraversion	3.46 (.63)	3.45 (.74)	3.23 (.65)	55.64 (5.70)	.82	.78	.37	.79	.36
Agreeableness	3.38 (.59)	3.51 (.52)	3.26 (.50)	49.31 (6.25)	.78	.63	.22	.70	.22
Conscientiousness	3.95 (.52)	3.59 (.72)	3.34 (.83)	48.64 (8.07)	.78	.78	.37	.87	.50
Openness	3.72 (.57)	3.48 (.68)	3.17 (.56)	41.82 (4.42)	.71	.76	.35	.71	.25
Cognitive ability	7.31 (3.47)	3.63 (.68)	3.27 (.69)	-	.76	.79	.39	.88	.52
OCB - likelihood	9.62 (1.34)	3.61 (.57)	3.33 (.54)	-	.72	.74	.33	.82	.39
OCB – self-reports	5.11 (.88)	-	-	-	.93	-	-	-	-
Hiring recommendations	-	3.46 (.81)	3.21 (.84)	-	-	.83	.44	.90	.56

Note. N = 154 (but N = 134 for OCB-SJT). OCB = Organizational Citizenship Behaviors. [a] The main inter-rater agreement (ICC) used to examine reliability are based on 6 raters, and using ICC(1) for the online panel raters, and ICC(2) for the I-O raters. [b] We also report equivalent ICC(1,1), to facilitate comparisons across samples and with other studies.

Table 2

Convergent Validities Using Different Sources

	(1) - (2)	(1) - (3)	(1) - (4)	(2) - (3)	(2) - (4)	(3) - (4)
Personality						
Honesty-Humility	.17*	.16*	-	.62**	-	
Emotionality	.37*	.16*	-.09	.12	-.03	-.01
Extraversion	.03	.08	.08	.76**	.45**	.39**
Agreeableness	-.14†	-.10	.01	.64**	.20*	.27**
Conscientiousness	.09	.13	.14†	.82**	.39**	.41**
Openness	-.03	-.03	.13	.66**	-.28**	-.14†
Cognitive ability	.26**	.18*	-	.77**	-	-
OCB – likelihood	.12	.02	-	.75**	-	-
OCB – reported behaviors	.06	.06	-	-	-	-
Hiring recommendations	-	-	-	.82**	-	-

Note. N = 151 to 154. OCB = Organizational Citizenship Behaviors. (1) Self-reports and test scores; (2) Online panel of hiring professionals/raters; (3) Industrial/Organizational Psychology graduate student raters; (4) Automated assessments using Receptiviti's LIWC/Language-based Personality Insights. ** p < .01; * p < .05, † p < .10.

Table 3

Correlations Between LinkedIn Profile Content and Scores from the Five Sources

	Profile length	Professional picture	Profile summary	Connec- tions	Skills listed	Endorse- ments	Recom- mendations	Jobs	Detailed experiences
Mean	333.40	0.63	0.56	201.21	18.27	64.57	0.54	4.14	1.43
SD	332.38	0.48	0.50	191.17	14.08	140.56	2.10	3.23	0.66
Targets' self-reports and test scores									
H	.04	-.03	.06	-.10	.00	.01	-.02	-.02	.09
E	.03	-.02	-.01	-.08	-.18*	.01	.00	.01	-.03
X	.03	.04	.04	.06	.01	-.07	.08	.05	-.02
A	-.13	-.12	-.15	-.19*	-.09	-.12	-.01	-.12	-.09
C	-.01	.09	.07	.04	-.06	-.07	-.04	.08	.08
O	-.05	-.06	-.08	-.11	.01	.01	.01	-.07	.05
CA	.02	.00	-.01	.12	-.04	.02	.02	.05	.10
OCB – SR	.02	.11	-.03	-.11	-.12	-.33**	-.06	.03	-.03
OCB - SJT	.08	.17	.06	-.03	-.06	-.18*	-.01	.09	.02
LinkedIn-based ratings – Online panel of hiring professionals									
H	.46**	.38**	.40**	.35**	.42**	.28**	.14	.43**	.59**
E	-.05	-.10	-.06	-.03	-.04	.03	-.04	-.02	.00
X	.53**	.59**	.55**	.52**	.50**	.25**	.23**	.51**	.61**
A	.50**	.44**	.42**	.38**	.39**	.18*	.17*	.51**	.55**
C	.57**	.49**	.44**	.50**	.55**	.28**	.26**	.48**	.59**
O	.61**	.50**	.53**	.49**	.57**	.29**	.22**	.53**	.55**
CA	.53**	.44**	.36**	.53**	.50**	.31**	.22**	.46**	.47**
OCB	.50**	.50**	.43**	.45**	.47**	.23**	.16*	.43**	.57**
Hiring rec.	.52**	.49**	.37**	.55**	.50**	.26**	.23**	.45**	.54**
LinkedIn-based ratings – I/O Psychology raters									
H	.38**	.49**	.40**	.37**	.39**	.14	.16*	.45**	.52**
E	-.16*	-.39**	-.14	-.22**	-.18*	.01	-.08	-.14	-.25**
X	.56**	.58**	.55**	.61**	.45**	.24**	.30**	.47**	.53**
A	.50**	.52**	.48**	.54**	.45**	.19*	.33**	.47**	.58**
C	.65**	.54**	.45**	.58**	.48**	.30**	.23**	.56**	.67**
O	.59**	.50**	.46**	.57**	.47**	.35**	.29**	.50**	.56**
CA	.59**	.47**	.33**	.56**	.40**	.29**	.21**	.52**	.51**
OCB	.64**	.58**	.50**	.63**	.52**	.27**	.30**	.54**	.67**
Hiring rec.	.67**	.63**	.48**	.65**	.52**	.25**	.26**	.59**	.69**
Automated LinkedIn-based ratings using Receptiviti									
E	.04	.03	-.01	-.07	-.06	.01	-.03	-.02	.03
X	.08	.22**	.25**	.33**	.22**	.11	.02	.03	.15
A	.08	.14	.18*	.13	.12	.08	.07	.02	.15
C	.33**	.44**	.46**	.39**	.37**	.10	.10	.36**	.43**
O	-.18*	-.18*	-.02	-.19*	-.18*	-.12	-.11	-.36**	-.10

Note. H = Honesty-humility; E = Emotionality; X = Extraversion; A = Agreeableness; C = Conscientiousness; O = Openness; CA = Cognitive ability; OCB = Organizational Citizenship Behaviors.
** $p < .01$; * $p < .05$, † $p < .10$.

Appendix A: OCB SJT

Instructions: For each scenario, please sort the response options **from best to worst.**

Item 1: It is 5pm on Friday and you just finished your 8-hour shift. You are about to leave the office to enjoy a well-deserved weekend, when your co-worker, Susan, comes to you. She explains that she is having issues with the new software program your company has installed. What would you do?

a) I would not take time to assist Susan.
b) I would remind Susan to use the manual distributed to all employees.
c) I would ask Susan what the problem is and give her a few pointers.
d) I would go with Susan to her computer and work through her problems with her.

Item 2: You have had a really long morning of conference calls, catching up on emails, and troubleshooting problems with clients. You are looking forward to a few moments to relax in the staff room at lunch when your co-worker Dana begins to tell you about the problems she is having with her landlord. She tells you her apartment really needs a paint job and that she has contacted the landlord multiple times about the issue, but the landlord has yet to fix it. She asks for your advice. What do you do?

a) I would not give any advice.
b) I would give her a brief piece of advice (e.g., why not paint it yourself).
c) I would help her brainstorm ways to get her apartment painted by her landlord or by someone else.
d) I would offer to come over after work hours and help her paint her apartment.

Item 3: There has been some bad press about your organization lately due to the environmental impact of the products manufactured by the company. Although this is a problem your organization is beginning to remedy, when at an industry conference, an attendee begins to talk about the bad press. What would you do?

a) I would say nothing.
b) I would say try to change the subject to something that does not look bad for the company.
c) I would say the company is making changes to address the problem.
d) I would explain the ways in which the organization is addressing the problem and that I support their initiative.

Item 4: You are extremely behind on your work and desperately need to catch up. You notice long-time co- worker Daniel, who sits in the cubicle next to you, is looking extremely stressed. What would you do?

a) I would ignore Daniel and continue with my work.
b) I would say hello to Daniel briefly.
c) I would ask Daniel how he is doing but return to my work as quickly as possible.
d) I would ask Daniel how is doing and take the time to hear what he is struggling with.

Item 5: You have been experiencing some stress due to concerns about your health. However, at work, your new co-worker Jason seems as though he is having a hard time settling into his new job. What would you do?

a) I would leave Jason to figure out his work on his own.
b) I would send Jason an email asking how he is settling in.
c) I would ask Jason at lunch how he is doing with his new work.
d) I would pop by Jason's desk during my work hours with a coffee for him to see how things are going.

Item 6: You are a service worker for a large chain retailer. When working at the cash register, a customer begins to complain to you about the company you work for and how the quality of merchandise is very poor. You are satisfied with your job and believe the merchandise quality is decent. What would you do?

a) I would ignore the customer's comments
b) I would say I disagree
c) I would say I am proud to work for the company.
d) I would say I am proud to work for the company and mention the high-quality products we do sell.

Item 7: Although outside your job description, you have an idea for improving the function of your organization by streamlining a component of the manufacturing process at your job. What would you do?

a) I would keep the idea to myself.
b) I would share it with a co-worker in passing, but not management.
c) I would wait until employees are asked for ideas by my manager at a team meeting.
d) I would ask for a meeting with my manager and share the idea with them.

Item 8: This Saturday evening, your company is hosting a special event to showcase a new product they are producing. Although not a requirement of your job, they have invited all employees to attend. The event venue is about a one-hour drive from your home and you already had plans to go out with friends. What would you do?

a) I would not attend.
b) I would attend only if my plans were cancelled.
c) I would make sure I could attend, but only stay briefly.
d) I would make sure I could attend and stay as long as possible.

Scoring

- *Ranking of responses options for items are as follows: d, c, b, a.*
- *If respondents ranked the response option in the correct position (e.g., 4th response option first) = 3 points; If they were off by one position (e.g., ranked 3rd instead of 4th) = 2 points; If they were off by two positions (e.g., 2nd instead of 4th) = 1 point; If they were off by three positions (e.g., 1st instead of 4th) = 0 points.*
- *A maximum of 12 points per question can be awarded. The scale score can also be computed by determining the mean score for OCBI (items 1-4) and OCBO (items 5-8).*

Appendix B: Adverse Impact based on Race, Gender, and Age for Various Hiring Scenarios

Raters	**"Hiring" Cut-off**	**% hired**		**Minority-Majority Ratio**	**Chi-Square**	**Averse impact (4/5th rule)**
Race		**White**	**Non-White**			
Online panel	score > 3	73.6	70.7	104.1	0.158	No
	score > 4	29.3	38.9	132.9	1.587	No
	Top 25%	24.4	27.8	113.9	0.229	No
	Top 10%	12.2	4.2	37.2	3.197†	Yes
I-O	score > 3	68.3	58.3	85.4	1.643	No
	score > 4	18.5	20.8	112.5	0.130	No
	Top 25%	23.2	27.8	119.9	0.430	No
	Top 10%	11.0	9.7	88.6	0.065	No
Gender		**Male**	**Female**			
Online panel	score > 3	69.0	57.7	109.7	0.843	No
	score > 4	35.7	31.4	88.0	0.314	No
	Top 25%	26.2	25.7	98.2	0.005	No
	Top 10%	6.0	11.4	192.0	1.481	No
I-O	score > 3	60.7	67.1	110.6	0.682	No
	score > 4	17.9	21.7	121.74	0.362	No
	Top 25%	25.0	25.7	102.9	0.010	No
	Top 10%	8.3	12.9	154.3	0.839	No
Age		**Under 40**	**Over 40**			
Online panel	score > 3	70.9	73.1	103.1	0.091	No
	score > 4	37.2	28.4	76.2	1.328	Yes
	Top 25%	27.9	22.4	80.2	0.604	No
	Top 10%	5.8	10.4	179.7	1.119	No
I-O	score > 3	67.4	58.2	86.3	1.384	No
	score > 4	21.2	16.4	77.5	0.550	Yes
	Top 25%	31.4	16.4	52.29	4.525*	Yes
	Top 10%	9.3	11.9	128.4	0.280	No

Note. "Hiring" cut-off scores are based on the average hiring recommendation score across all raters. For the Online panel raters, the top 25% cut-off score was 4.13 and the top 10% cut-off score was 4.37. For the I-O raters, these values were 3.93 and 4.23, respectively. Adverse impact happens when the minority-to-majority ratio is below 0.80 (i.e., 4/5th rule from the U.S. Equal Employment Opportunity Commission). Age was split into under/over 40 years old, based on the U.S. Age Discrimination in Employment Act. $^{**}p < .01$; $^{*}p < .05$, $^{\dagger}p < .10$.