Agree to disagree: Examining the psychometrics of cybervetting

Amber N. Schroeder
Department of Psychology, The University of Texas at Arlington, Arlington, Texas, USA
Kaleena R. Odd
University of Nebraska at Omaha, Omaha, Nebraska, USA, and
Julia H. Whitaker
Department of Psychology, The University of Texas at Arlington, Arlington, Texas, USA

Received 26 September 2018
Revised 26 March 2019
31 December 2019
Accepted 8 May 2020

Abstract

Purpose – Due to the paucity of research on web-based job applicant screening (i.e. cybervetting), the purpose of the current study was to examine the psychometric properties of cybervetting, including an examination of the impact of adding structure to the rating process.

Design/methodology/approach – Using a mixed-factorial design, 122 supervisors conducted cybervetting evaluations of applicant personality, cognitive ability, written communication skills, professionalism, and overall suitability. Cross-method agreement (i.e. the degree of similarity between cybervetting ratings and other assessment methods), as well as interrater reliability and agreement were examined, and unstructured versus structured cybervetting rating formats were compared.

Findings – Cybervetting assessments demonstrated high interrater reliability and interrater agreement, but only limited evidence of cross-method agreement was provided. In addition, adding structure to the cybervetting process did not enhance the psychometric properties of this assessment technique.

Practical implications – This study highlighted that whereas cybervetting raters demonstrated a high degree of consensus in cybervetting-based attributions, there may be concerns regarding assessment accuracy, as cybervetting-based ratings generally differed from applicant test scores and self-assessment ratings. Thus, employers should use caution when utilizing this pre-employment screening technique.

Originality/value – Whereas previous research has suggested that cybervetting ratings demonstrate convergence with other traditional assessments (albeit with relatively small effects), these correlational links do not provide information regarding cross-method agreement or method interchangeability. Thus, this study bridges a crucial gap in the literature by examining cross-method agreement for a variety of job-relevant constructs, as well as empirically testing the impact of adding structure to the cybervetting rating process.

Keywords Selection, Human resource management, Decision-making, Scale development, Psychometrics

Paper type Research paper

A recent Society for Human Resource Management (SHRM, 2016) survey found that 43% of organizations screen job applicants by examining information provided on social networking sites (SNSs) via a process known as cybervetting (Berkelaar, 2017). As noted by Roth *et al.* (2016), little empirical work has investigated the cybervetting process, and employers who engage in this method are likely doing so without a structured or standardized process. This unstructured evaluation process could produce adverse effects, including discrimination based on job-irrelevant information, legal and privacy issues, and information overload (Davison *et al.*, 2011; Kluemper and Rosen, 2009; Roth *et al.*, 2016). Thus, in an effort to further our understanding of the psychometric properties of cybervetting techniques, this study examined the impact of increasing structure during a cybervetting evaluation.

This project was supported by the National Science Foundation [Award #1460942].

Journal of Managerial Psychology
Vol. 35 No. 5, 2020
pp. 435-450

0268-3946
DOI 10.1108/JMP-09-2018-0420

Organizational use of cybervetting assessments

Organizations use SNSs in a variety of employment practices, including recruitment, selection, and termination (Davison *et al.*, 2011). Many organizations cybervet as a means to gather pre-employment information about job applicants, including to confirm previously provided data about the applicant (Brown and Vaughn, 2011; SHRM, 2016), assess applicant fit (Chauhan *et al.*, 2013), "weed out" unwanted applicants (Berkelaar, 2017, p. 1,126), and reduce the number of applicants in the applicant pool (Byrnside, 2008). Yet, there may be unintended consequences to cybervetting, as organizational use of cybervetting evaluations has been linked to increased applicant perceptions of privacy invasion (Stoughton *et al.*, 2013), which may lead applicants to remove themselves from a selection process (Schneider *et al.*, 2015). Notably, SHRM (2016) reported that 76% of organizations are concerned about the liability risks associated with uncovering information about protected classes from SNSs, and this concern is a commonly reported reason for why organizations are discouraged from using cybervetting techniques (see, e.g. Davison *et al.*, 2016).

Nevertheless, because of the prevalence of cybervetting in applied settings, research has been conducted to increase our understanding of the strengths and weaknesses of this assessment technique. Namely, several studies have provided evidence of convergent validity across self-reported and cybervetting-based ratings of personality traits, such as the Big Five (Back *et al.*, 2010; Kluemper *et al.*, 2012) and narcissism (Buffardi and Campbell, 2008). Likewise, Park *et al.* (2015) and Gosling *et al.* (2011) demonstrated personality measurement convergence across self-assessments and SNS profile content analysis techniques. Notably, convergence studies in this area have largely focused on personality assessment and have typically excluded other relevant constructs. For example, as highlighted by Roth *et al.* (2016), selection assessments often assess cognitive ability and written communication, but little is known about the ability to assess these constructs via cybervetting techniques. Likewise, others have suggested that job-relevant attributes, such as cognitive ability, written communication and professionalism could be assessed via SNSs (Brown and Vaughn, 2011; Davison *et al.*, 2011). Thus, the current study bridges an important gap in the literature by partaking in a comprehensive comparison of traditional versus cybervetting-based assessments of personality, cognitive ability, written communication skills and professionalism.

Whereas demonstrating convergent validity is an important step in providing evidence of construct validity (Campbell and Fiske, 1959), the magnitude of cross-method correlations is a particularly important consideration. Namely, "high convergent validity means that [measures] may be used interchangeably for theory testing and administrative purposes" (p. 812), and when convergence is low, the assessment method may serve as a moderator between the construct of interest and relevant criteria such that criterion-related validity coefficients differ based on the measurement technique used to assess the predictor (Heneman, 1986). Notably, in cybervetting convergent validity examinations, small to moderate effect sizes have typically been demonstrated. For instance, Kluemper *et al.* (2012) reported a mean correlation of 0.29 between self-reported and cybervetting-based personality scores across two studies, and Van Iddekinge *et al.* (2016) found a 0.23 correlation between cybervetting estimates of cognitive ability and standardized test scores. Thus, this suggests that cybervetting techniques may not be an appropriate substitute for traditional measures of some job-relevant constructs.

Relatedly, discriminant validity reflects the extent to which measures are different or distinguishable from one another (Campbell and Fiske, 1959), such that differential prediction of outcomes by predictor measures of the same construct is problematic in terms of establishing construct validity. Stated differently, if two predictor measures of the same construct have different criterion-related validity coefficients with an outcome measure, this is evidence that the predictor measures differ in a meaningful way. Notably, two of the only studies to date comparing criterion-related validity coefficients across cybervetting and more traditional selection assessments have revealed contradictory findings. Namely,

Van Iddekinge *et al.* (2016) found that cybervetting assessments did not predict additional variance in job performance, turnover intentions, or actual turnover after accounting for variance predicted by more traditional assessments. In contrast, Kluemper *et al.* (2012) found that cybervetting assessments did account for incremental variance in the prediction of job performance above self-reported personality scores. Thus, because of the relatively small cross-method correlations and mixed findings regarding how cybervetting techniques compare to more traditional assessment methods in terms of outcome prediction, it remains unclear as to whether cybervetting techniques should be considered an alternate yet comparable means of assessing job-relevant constructs.

Yet another important psychometric consideration is whether cybervetting methods demonstrate acceptable levels of reliability. As with convergent validity estimates, reliability (i.e. rating consistency, such as in rank ordering; Tinsley and Weiss, 1975) is typically assessed using correlation-based analyses (Fleenor *et al.*, 1996). However, such approaches fail to provide information regarding the accuracy of measurements in comparison to a set standard and are not sufficient evidence of construct validity (Nunnally, 1978). Interrater agreement, on the other hand, reflects the extent to which raters make identical judgments (i.e. they provide the exact same ratings; Tinsley and Weiss, 1975). Thus, with high interrater agreement, raters become interchangeable. Several studies (e.g. Back *et al.*, 2010; Kluemper and Rosen, 2009) have demonstrated moderately high or better interrater agreement in a cybervetting context, thereby illustrating that raters are often coming to relatively similar conclusions in cybervetting assessments. However, as organizations often use cybervetting as a means to screen out job candidates (Chauhan *et al.*, 2013), it is important that organizations realize that just as reliability and agreement have different meanings, cross-method convergence is not the same as cross-method agreement. Stated differently, a low to moderate correlation between two methods does not imply interchangeability. Further, as research has been mixed regarding the incremental value of cybervetting approaches over that of traditional assessments (see, e.g. Kluemper *et al.*, 2012; Van Iddekinge *et al.*, 2016), it is important to know whether cybervetting methods have high cross-method agreement with other assessments of the same constructs so that organizations can understand the implications of utilizing cybervetting as an applicant screening technique. As such, the current study investigated score equivalency across traditional versus cybervetting-based assessments by examining cross-method agreement in order to provide valuable information regarding the interchangeability of various assessment techniques.

Potential sources of error in cybervetting-based assessment

Cybervetting assessments may be likely to contain a large degree of measurement error given the highly unstructured and unstandardized nature of this technique (e.g. Roth *et al.*, 2016). Contributing factors may include the abundance of job-irrelevant information, the absence of a social media presence for some job applicants, the use of strict privacy settings that may limit access to profile information (Davison *et al.*, 2012; Roth *et al.*, 2016), the manipulation of SNS content by applicants to convey a more positive image (Schroeder and Cavanaugh, 2018) and a lack of organizational protocols for SNS screening (SHRM, 2016). Further, SNSs often contain a large amount of content posted by friends of the applicant, which could distract or bias cybervetters (Davison *et al.*, 2012). Thus, researchers have called for increased structure in cybervetting techniques (Roth *et al.*, 2016).

Although empirical work has yet to investigate the role of structure in a cybervetting context, researchers have highlighted a parallel between job interviews and cybervetting, as both techniques involve the assessment of large amounts of qualitative information (Roth *et al.*, 2016). Notably, meta-analytic findings have revealed that structured interviews have greater reliability and validity than unstructured interviews (McDaniel *et al.*, 1994), and a variety of recommendations for improving the psychometric properties of interviews have

been set forth, such as designing behaviorally anchored rating scales and using standardized questions across applicants (Campion *et al.*, 1997). Thus, given the similarity between cybervetting and interviewing, as well as the psychometric value that standardization adds to interviewing methodologies, it seems likely that adding structure to cybervetting assessments may result in psychometric improvements to cybervetting methods.

Taken together, the current study expands on previous research comparing cybervetting techniques to traditional assessment formats by also considering cognitive ability, written communication skills, and professionalism, in addition to personality traits. Because of the relatively small effect sizes found in previous convergent validity examinations in this context, as well as the many potential sources of error present on SNSs, low cross-method agreement is expected. Further, it is expected that adding structure to cybervetting assessments will strengthen the psychometric properties of this technique.

H1. Applicant self-report and test scores will differ from cybervetting-based ratings of the same constructs.

H2. Adding structure to the cybervetting process will increase (a) cross-method agreement, (b) interrater agreement and (c) interrater reliability.

Additionally, researchers (e.g. Van Iddekinge *et al.*, 2006) have suggested that rater individual differences in an applicant interviewing context (e.g. interviewer experience, demographic characteristics, cognitive ability, personality) may influence interviewer decisions and affect the validity of the interview process. Lippa and Dietz (2000) found that rater intelligence was positively linked to trait accuracy (i.e. the correlation between self-reported and rater scores) for extraversion, and rater openness to experience was negatively related to trait accuracy for neuroticism. In a cybervetting context, Kluemper and Rosen (2009) found that judges who were more intelligent and emotionally stable were able to make better assessments of personality and cognitive ability. Thus, on an exploratory basis, we examined the influence of cybervetter characteristics on cross-method agreement.

RQ. Is cross-method agreement related to rater individual difference factors, including cognitive ability, personality, work experience, applicant assessment experience and familiarity with and length of time on Facebook?

Method

Participants

Participants included 122 US.-based individuals (mean age = 35.2 years, *SD* = 10.4; 58.2% female; 82.8% White/Caucasian) currently employed in supervisory positions recruited via Amazon Mechanical Turk (MTurk), an online crowdsourcing platform.

Materials

Given that there are 1.52 billion daily active users on Facebook (as of December 2018; Facebook, 2019), we chose this platform for the current study. Materials included Facebook profiles and assessment data from three individuals, a survey measuring cybervetter attributes and cybervetting rating forms used to assess Facebook profiles.

Facebook profiles and applicant-sourced data. To identify individuals whose Facebook profiles would be used in the current study, nine individuals from the researchers' personal Facebook networks completed measures of personality, cognitive ability, written communication skills and professionalism. To assess personality, potential profilees completed the 44-item Big Five Inventory (BFI; John *et al.*, 1991). Professionalism was examined using a modified version of the 15-item Professionalism—Documentation of Competence (ProDOC; Hershberger *et al.*, 2010) in which several items were revised to

broaden the context (e.g. "patient/coworker" was replaced with "others"). Both personality and professionalism items were rated using a five-point Likert scale with responses ranging from 1 (*disagree strongly*) to 5 (*agree strongly*).

Cognitive ability was assessed using Raven's Advanced Progressive Matrices Short Form (APM; Bors and Stokes, 1998; Raven *et al.*, 1988), in which participants performed 12 pattern completion tasks, and sample items from the Public Service Commission of Canada's Written Communication Proficiency Test (PSC, 2011) and the Educational Testing Service (ETS) HEIghten Written Communication Assessment (ETS, n.d.) were used to assess written communication skills. For the written communication skills test, participants read two passages and answered 13 multiple-choice questions assessing their knowledge of social and rhetorical situations, conceptual strategies and language use and conventions. For both the cognitive ability and written communication assessments, the percentage of correct responses was multiplied by five to transform composite scores to a five-point scale.

After examining scores on these eight constructs, the three individuals who maximized the range of construct scores (e.g. extraversion ranged from 1.75 to 4.13 on a five-point scale) were selected for the current study. To reduce the likelihood of confounding effects due to individual demographics, individuals were selected so that there would be minimal demographic diversity (i.e. ages ranged from 21 to 23; hours worked ranged from 20 to 40 per week; all individuals were White/Caucasian females). Facebook profiles for these three individuals were downloaded, and personal identifiers were removed. Each downloaded profile contained approximately 20 pages (based on scrolling view) of content from the following categories: timeline posts, about page information, friends list, photos, notes, events, groups, likes, check-ins and hobbies/interests (i.e. sports, music, books, TV shows).

Cybervetter attribute survey. Raters provided information regarding their sex, age, years of work experience, years of applicant screening experience, Facebook tenure in years and Facebook familiarity (i.e. on a five-point Likert scale, 1 = *not familiar* to 5 = *very familiar*). Additionally, raters completed the 12-item RAVEN and 44-item BFI described in the previous section to assess cognitive ability and personality, respectively.

Cybervetting-based rating forms. Cybervetting rating forms were provided to the raters based on the assigned experimental condition (see Table 1). In all conditions, raters assessed each Facebook profile using a four-item applicant suitability scale (i.e. an adaptation of Cable and Judge (1997) and Cole *et al.* (2004); average $\alpha = 0.91$ across profiles and conditions). Three items measured interview, hire, and job success likelihood using a five-point Likert scale with responses ranging from 1 (*very unlikely)* to 5 (*very likely*), and a final overall evaluation item was included using a five-point Likert scale with responses ranging from 1 (*very negative*) to 5 (*very positive*).

In each of the structured conditions, definitions of cognitive ability, extraversion, agreeableness, conscientiousness, neuroticism, openness to experience, written communication skills and professionalism were provided. Definitions were consistent with

	Unstructured condition	Structured condition 1	Structured condition 2	Structured condition 3
Rated applicant suitability	✓	✓	✓	✓
Were given definitions of key constructs		✓	✓	✓
Rated each applicant on each key construct			✓	✓
Based construct judgments on specific profile features				✓

Table 1.
Description of study conditions

or based on each of the corresponding measures administered to Facebook profilees described previously. In the second and third structured conditions, raters also evaluated each Facebook profilee's personality, cognitive ability, written communication skills and professionalism after reading definitions of each construct. Each one-item measure was rated on a five-point Likert scale with responses ranging from 1 (*strongly disagree*) to 5 (*strongly agree*) for personality, 1 (*very poor*) to 5 (*very good*) for cognitive ability and written communication skills, and 1 (*very unprofessional*) to 5 (*very professional*) for professionalism. Rating scale brevity was considered important in the current study to maintain a sense of realism (i.e. it is likely that organizational cybervetters generally make more holistic judgments about profilees rather than complete multi-item construct assessments). Relatedly, research has demonstrated that single-item measures can demonstrate acceptable (although not ideal) psychometric properties (Gosling *et al.*, 2003). In the third structured condition, raters were also given instructions to focus on specific profile features (e.g. for extraversion, base your judgment on the total number of friends). These recommendations were based on suggestions by Gosling *et al.* (2011), as well as on a logical basis.

Procedure

Using a mixed factorial design, participants assumed the role of hiring managers and evaluated three applicant Facebook profiles for a retail sales supervisor position. A mixed factorial design was utilized to examine the impact of increasing rating structure both within raters (i.e. a repeated measures design), as well as across raters (i.e. a between-subjects design). Half of the sample (i.e. 61 participants) completed an unstructured cybervetting evaluation and was then assigned to one of three structured cybervetting conditions. To ensure that completing an unstructured evaluation prior to employing a structured cybervetting approach did not result in confounding memory effects, the remaining half of the sample was randomly assigned to only one of the three structured cybervetting conditions (i.e. they did not complete an unstructured evaluation; see Table 1 for condition descriptions). The rater attributes survey was completed between the unstructured and structured cybervetting evaluations for the repeated measures design group and after the structured evaluation for the remaining participants.

Results

To examine *H1*, a series of one-sample *t*-tests were conducted to determine if the cybervetting-based ratings provided by the raters differed from the test scores and self-reported ratings of the applicant depicted in each profile. The results indicated that rater evaluations were significantly different from applicant test scores and self-reported ratings in 20 of 24 comparisons (i.e. there were eight constructs rated for each of the three profiles), with generally moderate to large effect sizes (see Table 2). In 14 of these comparisons (i.e. 58.3%), cybervetting-based ratings were higher than applicant test scores and self-reported ratings, and in six cases (i.e. 25.0%), cybervetting-based ratings were significantly lower than applicant test scores and self-reported ratings. Notably, there were no patterns of agreement by rated construct (i.e. no ratings of a specific construct demonstrated agreement with applicant-sourced data across all three profiles). Thus, *H1* was partially supported.

In assessing *H2*, first, independent samples *t*-tests were conducted to determine whether cross-method agreement (i.e. the degree to which cybervetting-based ratings differed from applicant test scores and self-reported ratings) differed by condition. Looking across profiles, there were no significant differences for any of the eight construct ratings by condition (see Figure 1). Therefore, as increasing rating scale structure did not produce higher cross-method agreement, *H2a* was not supported. Likewise, paired samples *t*-test results indicated that for

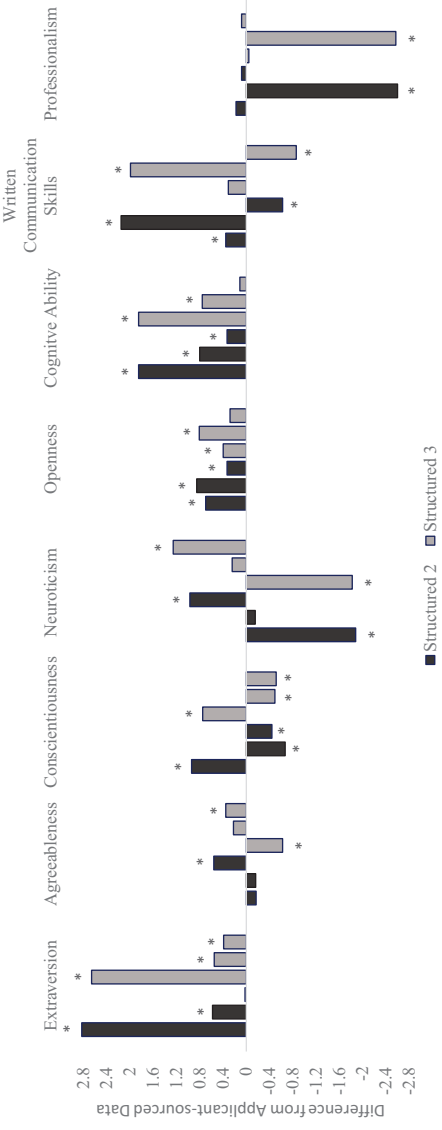
	Mean difference	*t*	*p*	*d*
Extraversion				
Profile 1	2.738*	35.066	<0.001	3.872
Profile 2	0.561*	6.877	<0.001	0.759
Profile 3	0.211*	2.254	0.027	0.249
Agreeableness				
Profile 1	−0.402*	−4.536	<0.001	0.501
Profile 2	0.037	0.377	0.707	0.042
Profile 3	0.451*	4.987	<0.001	0.551
Conscientiousness				
Profile 1	0.841*	8.206	<0.001	0.906
Profile 2	−0.573*	−4.675	<0.001	0.516
Profile 3	−0.477*	−4.549	<0.001	0.502
Neuroticism				
Profile 1	−1.850*	−14.885	<0.001	1.644
Profile 2	0.049	0.352	0.726	0.039
Profile 3	1.120*	7.813	<0.001	0.863
Openness to experience				
Profile 1	0.544*	5.865	<0.001	0.648
Profile 2	0.827*	10.152	<0.001	1.121
Profile 3	0.300*	3.093	0.003	0.342
Cognitive ability				
Profile 1	1.847*	22.805	<0.001	2.518
Profile 2	0.780*	9.530	<0.001	1.052
Profile 3	0.213*	2.540	0.013	0.280
Written communication skills				
Profile 1	0.187*	2.221	0.029	0.245
Profile 2	2.070*	25.413	<0.001	2.806
Profile 3	−0.742*	−8.323	<0.001	0.919
Professionalism				
Profile 1	0.069	0.734	0.465	0.081
Profile 2	−2.585*	−21.979	<0.001	2.427
Profile 3	0.081	0.886	0.378	0.098

Note(s): Analyses were run across conditions
*$p < 0.05$. †$p < 0.10$. $N = 82$

Table 2.
Results of one-sample *t*-tests examining cross-method agreement

raters who completed both unstructured and structured rating tasks, in eight out of nine cases (i.e. three profiles per structured condition), overall suitability ratings did not change from the unstructured to the structured condition, which further highlights that adding structure to the rating process had only minimal effects on cybervetting evaluations. Next, interrater agreement and reliability were examined by calculating intraclass correlations (ICCs). As illustrated in Table 3, interrater agreement and reliability only increased for neuroticism ratings when rating scale structure increased. As such, minimal support was garnered for *H2b* and *H2c*.

Further, to examine whether high interrater reliability and agreement across conditions were due to meaningful patterns in rater evaluations versus raters' inability to make distinctions between applicants across profiles (e.g. due to inattention issues), applicant suitability ratings for the three profiles were compared within each condition using a repeated measures ANOVA. Notably, the results indicated that ratings *did* differ across profiles ($F_{\text{unstructured}}(2,124) = 67.76$, $p < 0.01$; $F_{\text{structured1}}(2,78) = 45.34$, $p < 0.01$;



Note(s): Positive (negative) values reflect higher (lower) scores for the cybervetting-based assessment. The three bars in each condition denote profiles 1, 2, and 3, respectively. * indicates $p < 0.05$

Figure 1.
Differences between cybervetting ratings and applicant test score and self-reported ratings for each profile within the structured 2 and structured 3 conditions

Rating construct	Reliability				Agreement			
	US	*S*1	S2	*S*3	US	*S*1	*S*2	*S*3
Extraversion	–	–	0.731	0.648[a]	–	–	0.724	0.558[a]
Agreeableness	–	–	0.964	0.843[b]	–	–	0.945	0.711[b]
Conscientiousness	–	–	0.882	0.768	–	–	0.876	0.734
Neuroticism	–	–	0.775	0.879	–	–	0.691	0.838
Openness to experience	–	–	0.688[a]	0.656	–	–	0.634[a]	0.617
Cognitive ability	–	–	0.948	0.930	–	–	0.926	0.897
Written communication skills	–	–	0.950	0.937	–	–	0.933	0.917
Professionalism	–	–	0.982	0.973	–	–	0.981	0.966
Overall suitability	0.985	0.978	0.987	0.977	0.975	0.962	0.983	0.965

Note(s): US = unstructured condition; *S*1 = structured condition 1; *S*2 = structured condition 2; *S*3 = structured condition 3. Reliability could not be calculated for all variables in the US and *S*1 conditions, as raters provided only overall suitability ratings in these conditions

[a] Three raters with atypical rating patterns were removed

[b] Two raters with atypical rating patterns were removed

Table 3.
Interrater reliability and agreement coefficients

$F_{structured2}(2,78) = 76.55$, $p < 0.01$ and $F_{structured3}(2,82) = 43.14$, $p < 0.01$). More specifically, post hoc tests on overall perceptions of applicant suitability indicated that 10 of 12 paired profile comparisons (i.e. three profile comparisons per condition) were significant ($p < 0.05$), which demonstrates that high interrater reliability and agreement were not due to raters simply providing the same ratings to all three profiles.

In regard to the *RQ*, correlations between rater attributes and the degree to which cybervetting-based ratings differed from applicant test scores and self-reported ratings were examined. As illustrated in Table 4, cross-method agreement was higher (i.e. differences between cybervetting-based ratings and applicant-sourced data were smaller) for neuroticism and professionalism assessments when raters reported greater familiarity with Facebook, and for openness to experience measures as raters increased in neuroticism. Conversely, cross-method agreement was lower (i.e. differences between cybervetting-based ratings and applicant-sourced data were larger) for written communication assessments as raters increased in cognitive ability, for openness to experience assessments as rater extraversion increased, and for measures of both extraversion and agreeableness for non-White or multi-racial raters as compared to White/Caucasian raters. Additionally, females had marginally higher cross-method agreement than males for extraversion, raters higher in openness to experience had marginally higher cross-method agreement for neuroticism, and more conscientious raters had lower cross-method agreement for openness to experience.

Lastly, because of the generally low cross-method agreement demonstrated in this study, an exploratory analysis was conducted to examine the rating strategy used by cybervetters. Namely, three multiple regressions were conducted in which perceptions of applicant suitability was regressed on each of the eight construct ratings for each of the three Facebook profiles. Results indicated that professionalism carried significant weight in all three overall suitability evaluations, cognitive ability emerged as a significant predictor for two profiles, and openness to experience and written communication significantly impacted overall suitability perceptions in one of the three profiles (see Table 5).

Discussion

The results of this study provide valuable insight into the psychometric properties of the cybervetting process. Consistent with our predictions, rater evaluations generally did *not* align with the self-reported ratings and test scores of the applicants, thereby highlighting potential concerns regarding rating accuracy. According to Brunswik's (1956) lens model, trait attributions may occur based on behavioral cues that are not valid indicators of

Table 4.
Descriptive statistics and intercorrelations among rater attributes and cross-method agreement indices

Variable	*M* (SD)	1	2	3	4	5	6	7	8	9
Rater Attributes										
1. Sex	1.370(0.485)									
2. Age (years)	35.232(10.279)	−0.250*								
3. Race	0.207(0.408)	0.049	−0.182							
4. Work experience (years)	14.714(11.051)	−0.235*	0.859*	−0.200†						
5. Applicant screening experience (years)	4.908(4.814)	0.031	0.504*	−0.160	0.547*					
6. Facebook tenure (years)	7.003(2.528)	−0.165	−0.095	0.099	−0.006	0.069				
7. Facebook familiarity	4.610(0.698)	−0.011	−0.235*	0.071	−0.152	0.048	0.238*			
8. Cognitive ability	4.207(2.448)	−0.054	0.012	−0.019	0.030	0.081	0.284*	0.026		
9. Extraversion	3.965(0.908)	−0.227*	0.253*	−0.068	0.178	0.145	−0.035	0.061	−0.061	
10. Agreeableness	4.686(0.779)	−0.128	0.029	−0.038	−0.088	−0.221*	−0.089	0.152	−0.124	0.312*
11. Conscientiousness	5.080(0.646)	−0.033	0.198†	0.009	0.199†	0.181	0.087	0.161	−0.004	0.232*
12. Openness to experience	4.027(0.768)	0.149	−0.059	0.112	−0.099	−0.105	−0.109	0.103	−0.061	0.200†
13. Neuroticism	3.145(0.863)	−0.029	−0.120	0.093	−0.077	0.017	−0.009	−0.120	0.112	−0.391*
Differences from applicant data[a]										
14. Extraversion	1.399(0.357)	0.097†	−0.165	0.191†	−0.155	−0.079	0.089	0.054	−0.105	0.014
15. Agreeableness	0.699(0.379)	0.043	−0.069	0.281*	−0.166	−0.131	−0.112	0.080	−0.089	0.013
16. Conscientiousness	0.946(0.428)	−0.171	0.031	0.020	−0.078	0.069	0.053	−0.045	0.081	0.104
17. Openness to experience	0.871(0.322)	−0.142	−0.035	0.172	−0.076	−0.109	0.039	−0.040	−0.089	0.275*
18. Neuroticism	1.425(0.451)	−0.055	0.035	−0.142	0.061	0.008	0.041	−0.259*	0.047	−0.013

(*continued*)

Table 4.

Variable	*M* (SD)	1	2	3	4	5	6	7	8	9
19. Cognitive Ability	1.126(0.389)	−0.079	0.063	−0.023	−0.014	0.032	0.115	0.122	−0.059	0.197†
20. Written Communication	1.201(0.316)	0.067	0.024	−0.003	0.060	0.086	0.158	0.123	0.390*	0.103
21. Professionalism	1.290(0.469)	0.004	0.112	−0.016	0.104	−0.018	0.026	−0.226*	−0.098	0.012

Variable	10	11	12	13	14	15	16	17	18	19	20
Rater attributes											
11. Conscientiousness	0.382*										
12. Openness to experience	0.397*	0.228*									
13. Neuroticism	−0.492*	−0.500*	−0.086								
Differences from applicant scores[a]											
14. Extraversion	−0.053	0.066	0.068	0.093							
15. Agreeableness	0.065	0.018	0.033	−0.052	0.245*						
16. Conscientiousness	−0.042	0.103	0.102	−0.077	0.047	0.110					
17. Openness to experience	0.143	0.206†	0.060	−0.247*	0.307*	0.130	0.132				
18. Neuroticism	−0.041	0.040	−0.194†	−0.036	−0.099	−0.035	0.028	0.192†			
19. Cognitive ability	0.147	0.133	0.063	−0.081	0.020	0.130	0.061	0.250*	−0.059		
20. Written communication	−0.064	0.107	0.054	0.018	0.057	−0.107	0.045	−0.010	−0.164	0.208	
21. Professionalism	−0.069	−0.077	−0.004	−0.018	0.063	0.095	0.208†	−0.003	0.144	−0.215†	−0.223*

Note(s): $N = 82$ due to only structured conditions two and three providing construct ratings for each profile. Sex was coded: 0 = female, 1 = male. Race was coded: 0 = White/Caucasian, 1 = non-White/Caucasian or multi-race. Cognitive ability was scored from 0 to 12. Written communication was scored from 0 to 13. All other variables were rated on a 1 to 5 scale unless otherwise noted

[a] Absolute value of the difference between rater estimates and applicant self-report or test scores

*$p < 0.05$. †$p < 0.10$

Table 5.
Results of multiple regressions examining rating strategies

	Profile 1			Profile 2			Profile 3		
	β	*t*	*p*	*β*	*t*	*p*	*β*	*t*	*p*
Extraversion	0.046	0.664	0.509	0.049	0.732	0.467	0.071	0.790	0.432
Agreeableness	0.078	1.034	0.305	−0.032	−0.519	0.605	0.034	0.355	0.723
Conscientiousness	0.046	0.591	0.557	0.095	1.369	0.175	0.033	0.360	0.720
Neuroticism	−0.036	−0.517	0.607	−0.031	−0.521	0.604	−0.094	−1.111	0.270
Openness to experience	0.032	0.434	0.666	0.030	0.476	0.635	0.236*	2.410	0.018
Cognitive ability	0.140	1.601	0.114	0.326*	3.475	0.001	0.309*	3.044	0.003
Written communication skills	0.218*	2.203	0.031	0.041	0.398	0.692	−0.011	−0.104	0.917
Professionalism	0.502*	5.483	<0.001	0.587*	8.165	<0.001	0.386*	4.229	<0.001
	Adj. $R^2 = 0.663$*			*Adj.* $R^2 = 0.756$*			*Adj.* $R^2 = 0.552$*		

Note(s): *$p < 0.05$. †$p < 0.10$

underlying attributes. Thus, these results may suggest that raters relied on cues that were not diagnostic of applicant attributes. Further, contrary to our expectations, increasing structure generally did not enhance the psychometric properties demonstrated in cybervetting assessments. Rather, all study conditions demonstrated a relatively high degree of interrater reliability and interrater agreement regardless of the amount of structure in the rating process. Notably, unlike in interviews in which standardization may increase content stability across applicants, with cybervetting, no matter how structured the rating process is, the content on a given profile will not change. Thus, lack of effects regarding increased structure may be, in part, due to the static nature of content in cybervetting assessments. As structure can be introduced in a variety of ways (Campion *et al.*, 1997), perhaps other methods of structuring cybervetting-based assessments would be more effective in improving the psychometric properties of this technique. For example, a recent meta-analysis identified links between profile owner attributes and social media activity (Settanni *et al.*, 2018), which could be used to develop anchored rating scales for future cybervetting assessments. Additionally, cybervetting training may aid raters in more accurately identifying relevant behavioral cues.

Additionally, exploratory analyses indicated that there were some instances in which cross-method agreement was positively related to rater familiarity with Facebook and rater neuroticism and inversely associated rater extraversion and rater cognitive ability. Particularly interesting is that rater cognitive ability was inversely related to cross-method agreement in assessing written communication. As numerous accounts have demonstrated a positive link between cognitive ability and facets of written communication (see, e.g. Caemmerer *et al.*, 2018), perhaps this finding can be attributed to the informal nature of Facebook communication in which normative standards for communication differ from other contexts (e.g. the appropriateness of texting language like LOL). In addition, White/Caucasian raters demonstrated higher cross-method agreement than non-White or multi-racial raters. This may be attributable to an outgroup homogeneity effect in which individuals perceive more variability for ingroup than outgroup members (Rubin and Badea, 2012), as all three applicants in this study were White/Caucasian.

Interestingly, exploratory analyses revealed that rater assessments of professionalism, cognitive ability, openness to experience, and written communication carried the most weight in influencing perceptions of applicant suitability. However, it is important to highlight that even though rater perceptions of these constructs impacted their overall perceptions, cybervetting assessments generally lacked agreement with applicant-sourced data, which suggests that there may be data accuracy concerns in cybervetting-based assessment. Thus, despite generally high rater consensus, cybervetting rater strategies may be based on misconstrued conceptualizations of applicant attributes. This is particularly noteworthy, as these findings suggest that cybervetting may not be an appropriate substitute for more traditional selection assessments. And as research has suggested that cybervetting-based ratings do not demonstrate incremental validity in the prediction of job performance (Van Iddekinge *et al.*, 2016), this begs the question of whether cybervetting-based assessment has utility as an applicant screening tool. Thus, we suggest that organizations use caution when determining whether to cybervet job applicants until more research is conducted on this topic.

Study limitations and future research directions

Although study Facebook profiles were carefully chosen, future work should examine additional SNS profiles to ensure the observed effects were not related to specific features of the profiles used in this study. It would also be of interest to investigate whether similar effects are demonstrated on other SNS platforms. Further, as the current study controlled for demographic differences by selecting a small sample of White/Caucasian female applicants in their early 20s,

future research should also examine whether demographic attributes impact cross-method agreement, as gender bias has been demonstrated in other evaluation contexts (Koch *et al.*, 2015). Future work should also examine cybervetting as part of the larger screening process (e.g. in conjunction with other assessment data) to better understand the implications of its utilization in real-world employment contexts. Moreover, as our study design precluded an examination of traditional indices of convergence (i.e. cross-method correlations) due to sample size limitations, we encourage more research examining both cross-method agreement and correlational trends in this context. In addition, although we based the structured condition three rating form on previous research and logical deduction, more work is needed to better understand how SNS user activity is linked to applicant attributes.

Because study results revealed variability in cross-method agreement across profiles, we conducted an exploratory analysis to determine which profile attributes contributed to the greatest lack of agreement across methods. Profiles in which there was a difference of more than two points (on a five-point Likert scale) across methods tended to have applicant scores near a scale end point. More research on this phenomenon is warranted, as this may suggest that raters engaged in central tendency rating errors or that extreme levels of certain traits are more difficult to assess via cybervetting methods. In addition, as study results indicated that raters had similar perceptions of applicants even when limited rating guidance was provided, more research is needed examining cybervetter judgment and decision-making processes, as well as what types of SNS content are diagnostic of various applicant attributes. Moreover, investigations of other strategies for structuring cybervetting assessments that focus on SNS content (e.g. using computational software to standardize the type of content viewed by raters) would be of interest. Taken together, there are numerous avenues for research that would contribute to our understanding of the utility of social media evaluation in pre-employment screening.

References

Back, M.D., Stopfer, J.M., Vazire, S., Gaddis, S., Schmukle, S.C., Egloff, B. and Gosling, S.D. (2010), "Facebook profiles reflect actual personality, not self-idealization", *Psychological Science,* Vol. 21, pp. 372-374, doi: 10.1177/0956797609360756.

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,* Vol. 70, pp. 1115-1140, doi: 10.1177/0018726716686400.

Bors, D.A. and Stokes, T.L. (1998), "Raven's Advanced Progressive Matrices: norms for first-year university students and the development of a short-form", *Educational and Psychological Measurement,* Vol. 58, pp. 382-398, doi: 10.1177/0013164498058003002.

Brown, V.R. and Vaughn, E.D. (2011), "The writing on the (Facebook) wall: the use of social networking sites in hiring decisions", *Journal of Business and Psychology,* Vol. 26, pp. 219-225, doi: 10.1007/s10869-011-9221-x.

Brunswik, E. (1956), *Perception and the Representative Design of Psychological Experiments*, University of California Press, Berkeley, CA.

Buffardi, L.E. and Campbell, W.K. (2008), "Narcissism and social networking web sites", *Personality and Social Psychology Bulletin,* Vol. 34, pp. 1303-1314, doi: 10.1111/jopy.12305.

Byrnside, I. (2008), "Six clicks of separation: the legal ramifications of employers using social networking sites to research applicants", *Vanderbilt Journal of Entertainment and Technology Law,* Vol. 10, pp. 445-477.

Cable, D.M. and Judge, T.A. (1997), "Interviewers' perceptions of person-organization fit and organizational selection decisions", *Journal of Applied Psychology,* Vol. 82, pp. 546-561, doi: 10.1037/0021-9010.82.4.546.

Caemmerer, J.M., Maddocks, D.L., Keith, T.Z. and Reynolds, M.R. (2018), "Effects of cognitive abilities on child and youth academic achievement: evidence from the WISC-V and WIAT-III", *Intelligence*, Vol. 68, pp. 6-20.

Campbell, D.T. and Fiske, D.W. (1959), "Convergent and discriminant validation by the multitrait-multimethod matrix", *Psychological Bulletin*, Vol. 56, pp. 81-105.

Campion, M.A., Palmer, D.K. and Campion, J.E. (1997), "A review of structure in the selection interview", *Personnel Psychology,* Vol. 50, pp. 655-702, doi: 10.1111/j.1744-6570.1997.tb00709.x.

Chauhan, R.S., Buckley, M.R. and Harvey, M.G. (2013), "Facebook and personnel selection: what's the big deal?", *Organizational Dynamics,* Vol. 42, pp. 126-134, doi: 10.1016/j.orgdyn.2013.03.006.

Cole, M.S., Feild, H.S., Giles, W.F. and Harris, S.G. (2004), "Job type and recruiters' inferences of applicant personality drawn from resume biodata: their relationships with hiring recommendations", *International Journal of Selection and Assessment,* Vol. 12, pp. 363-367, doi: 10.1111/j.0965-075X.2004.00291.x.

Davison, H.K., Maraist, C. and Bing, M.N. (2011), "Friend or foe? The promise and pitfalls of using social networking sites for HR decisions", *Journal of Business and Psychology,* Vol. 26, pp. 153-159, doi: 10.1007/s10869-011-9215-8.

Davison, H.K., Maraist, C.C., Hamilton, R.H. and Bing, M.N. (2012), "To screen or not to screen? Using the internet for selection decisions", *Employee Responsibility and Rights Journal,* Vol. 24, pp. 1-21, doi: 10.1007/s10672-011-9178-y.

Davison, H.K., Bing, M.N., Kluemper, D.H. and Roth, P.L. (2016), "Social media as a personnel selection and hiring resource: reservations and recommendations", in Landers, N. and Schmidt, G.B. (Eds), *Social Media in Employee Selection and Recruitment: Theory, Practice, and Current Challenges*, Springer International Publishing, Switzerland, pp. 15-42, doi: 10.1007/978-3-319-29989-1_2.

Educational Testing Service (n d), HEIghten written communication sample items, available at: https://www.ets.org/s/heighten/pdf/written-communication-sample-questions.pdf.

Facebook (2019), "Facebook reports fourth quarter and full year 2018 results", available at: https://investor.fb.com/investor-news/press-release-details/2019/Facebook-Reports-Fourth-Quarter-and-Full-Year-2018-Results/default.aspx (accessed 10 June 2020).

Fleenor, J.W., Fleenor, J.B. and Grossnickle, W.F. (1996), "Interrater reliability and agreement of performance ratings: a methodological comparison", *Journal of Business and Psychology*, Vol. 10 No. 3, pp. 367-380.

Gosling, S.D., Rentfrow, P.J. and Swann, W.B. Jr (2003), "A very brief measure of the Big-Five personality domains", *Journal of Research in Personality,* Vol. 37 No. 6, pp. 504-528.

Gosling, S.D., Augustine, A.A., Vazire, S., Holtzman, N. and Gaddis, S. (2011), "Manifestations of personality in online social networks: self-reported Facebok-related behaviors and observable profile information", *Cyberpsychology, Behavior, and Social Networking,* Vol. 14, pp. 483-488, doi: 10.1089/cyber.2010.0087.

Heneman, R.L. (1986), "The relationship between supervisory rating and results-oriented measure of performance: a meta-analysis", *Personnel Psychology,* Vol. 39, pp. 811-826.

Hershberger, P.J., Zyrd, T.W., Rodes, M.B. and Stolfi, A. (2010), "Professionalism: self-control matters", *Medical Teacher,* Vol. 32, pp. e36-e41, doi: 10.3109/01421590903199676.

John, O.P., Donahue, E.M. and Kentle, R.L. (1991), *The Big Five Inventory–Versions 4a and 54*, University of California, Berkeley, Institute of Personality and Social Research, Berkeley, CA.

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

Kluemper, D.H., Rosen, P.A. and Mossholder, K.W. (2012), "Social networking websites, personality ratings, and the organizational context: more than meets the eye?", *Journal of Applied Social Psychology*, Vol. 42 No. 5, pp. 1143-1172, doi: 10.1111/j.1559-1816.2011.00881.x.

Koch, A.J., D'Mello, S.D. and Sackett, P.R. (2015), "A meta-analysis of gender stereotypes and bias in experimental simulations of employment decision making", *Journal of Applied Psychology,* Vol. 100, pp. 128-161, doi: 10.1037/a0036734.

Lippa, R.A. and Dietz, J.K. (2000), "The relation of gender, personality, and intelligence to judges' accuracy in judging strangers' personality from brief video segments", *Journal of Nonverbal Behavior,* Vol. 24, pp. 25-43.

McDaniel, M.A., Whetzel, D.L., Schmidt, F.L. and Maurer, S.D. (1994), "The validity of employment interviews: a comprehensive review and meta-analysis", *Journal of Applied Psychology,* Vol. 79, pp. 599-616, doi: 10.1037/0021-9010.79.4.599.

Nunnally, J.C. (1978), *Psychometric Theory*, McGraw-Hill Book Company, New York.

Park, G., Schwartz, H.A., Eichstaedt, J.C., Kern, M.L., Kosinski, M., Stillwell, D.J. and Seligman, M.E. (2015), "Automatic personality assessment through social media language", *Journal of Personality and Social Psychology*, Vol. 108, pp. 934-952.

Public Service Commission of Canada (2011), "Written communication proficiency test", available at: http://www.psc-cfp.gc.ca/ppc-cpp/psc-tests-cfp/wcpt-351-ecce-353-q-eng.htm.

Raven, J.C., Court, J.H. and Raven, J. (1988), *Manual for Raven's Progressive Matrices and Vocabulary Scales (Section 4)*, H. K. Lewis, London.

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

Rubin, M. and Badea, C. (2012), "They're all the same!... but for several different reasons: a review of the multicausal nature of perceived group variability", *Current Directions in Psychological Science*, Vol. 21, pp. 367-372, doi: 10.1177/0963721412457363.

Schneider, T.J., Goffin, R.D. and Daljeet, K.N. (2015), ""Give us your social networking site passwords": implications for personnel selection and personality", *Personality and Individual Differences*, Vol. 73, pp. 78-83.

Schroeder, A.N. and Cavanaugh, J. (2018), "Fake it till you make it: examining faking ability on social media pages", *Computers in Human Behavior,* Vol. 84, pp. 29-35, doi: 10.1016/j.chb.2018.02.011.

Settanni, M., Azucar, D. and Marengo, D. (2018), "Predicting individual characteristics from digital traces on social media: a meta-analysis", *Cyberpsychology, Behavior, and Social Networking*, Vol. 21, pp. 217-228, doi: 10.1089/cyber.2017.0384.

Society for Human Resource Management (2016), "SHRM survey findings: using social media for talent acquisition—recruitment and screening", available at: https://www.shrm.org/research/surveyfindings/pages/social-media-recruiting-screening-2015.aspx.

Stoughton, J.W., Thompson, L.F. and Meade, A.W. (2013), "Examining applicant reactions to the use of social networking websites in pre-employment screening", *Journal of Business and Psychology,* Vol. 30 No. 1, pp. 73-88.

Tinsley, H.E. and Weiss, D.J. (1975), "Interrater reliability and agreement of subjective judgments", *Journal of Counseling Psychology*, Vol. 22 No. 4, pp. 358-376.

Van Iddekinge, C.H., Sager, C.E., Burnfield, J.L. and Heffner, T.S. (2006), "The variability of criterion-related validity estimates among interviewers and interview panels", *International Journal of Selection and Assessment,* Vol. 14, pp. 193-205, doi: 10.1111/j.1468-2389.2006.00352.x.

Van Iddekinge, C.H., Lanivich, S.E., Roth, P.L. and Junco, E. (2016), "Social media for selection? Validity and adverse impact potential of a Facebook-based assessment", *Academy of Management Proceedings,* Vol. 42 No. 7, pp. 1811-1835.

Corresponding author
Amber N. Schroeder can be contacted at: amber.schroeder@uta.edu