

2015

## Assessment of Personality through Behavioral Observations in Work Simulations

Andrew B. Speer

Neil Christiansen

Christopher Honts

Follow this and additional works at: <https://scholarworks.bgsu.edu/pad>



Part of the [Human Resources Management Commons](#), [Industrial and Organizational Psychology Commons](#), and the [Other Psychology Commons](#)

### Recommended Citation

Speer, Andrew B.; Christiansen, Neil; and Honts, Christopher (2015) "Assessment of Personality through Behavioral Observations in Work Simulations," *Personnel Assessment and Decisions*: Vol. 1 : Iss. 1 , Article 6.

DOI: <https://doi.org/10.25035/pad.2015.006>

Available at: <https://scholarworks.bgsu.edu/pad/vol1/iss1/6>

This Measurement and Measures is brought to you for free and open access by the Journals at ScholarWorks@BGSU. It has been accepted for inclusion in Personnel Assessment and Decisions by an authorized editor of ScholarWorks@BGSU.

# ASSESSMENT OF PERSONALITY THROUGH BEHAVIORAL OBSERVATIONS IN WORK SIMULATIONS

Andrew B. Speer<sup>1</sup>, Neil Christiansen<sup>2</sup>, and Christopher Honts<sup>2</sup>

1. American Family Insurance & Central Michigan University

2. Central Michigan University

## ABSTRACT

### KEYWORDS

assessment  
centers, personality,  
behavioral  
observation, trait  
activation potential,  
work simulations

This study outlines the development of a rating scale designed to measure personality-related behavior in the context of work simulations. The tool, labeled the Work Simulation Personality Rating Scale (WSPRS), was validated in an assessment center by rating the personality of 123 assessment center participants. Scores from the WSPRS were correlated with corresponding traits from a self-reported personality inventory, and a trait activation potential (TAP) framework was adopted to predict which traits would display best convergence based on assessment center observations. Correlations between the WSPRS dimensions and self-report trait scales ranged from .11 (Neuroticism) to .31 (Extraversion), with the rank-order of convergence exactly reproducing the ordering of the TAP judgments made by the experts.

Personality constructs are accepted as being important for understanding work behavior, but despite the prevalence of personality testing in applied settings, this method of assessment is not without its critics. Concerns exist regarding the validity of these measures (Morgeson et al., 2007), the ease by which personality tests are faked (Tett & Christiansen, 2007), and unfavorable applicant reactions to personality measures (Hausknecht, Day, & Thomas, 2004). To a large extent, these issues are not necessarily reflective of the underlying personality constructs but rather the overreliance on self-reports to measure them. A potential complement to self-report inventories is the assessment of personality traits via behavioral observation.

Research on self-other convergence of personality traits has found these relationships can reach moderate to high levels, and ratings from sources other than the self can account for unique variance in the prediction of important outcomes such as daily behaviors, academic performance, and job performance (Connelly & Ones, 2010; Kolar, Funder & Colvin, 1996; Oh, Wang, & Mount, 2011; Small & Diefendorff, 2006; Vazire & Mehl, 2008). Thus, there is potential utility in personality-based behavioral observation. However, regarding use within a selection context, questions arise as to what sources should provide ratings, where they should occur, and how trait inferences should be made. This study addressed the latter of these issues by developing and validating a scale for coding personality-rel-

evant behavior specifically for work simulations. Work simulations (e.g., assessments centers) provide opportunity for behaviorally based observation in a standardized context. Although this lends to favorable conditions to observe personality-related behavior, there is no personality rating scale developed, validated, and made available for contexts such as these. As such, the purpose of this study was to develop and validate a scale for coding personality-relevant behavior specifically for work simulations.

### Behavioral Observation, Work Simulations, and Assessment Centers

There are several reasons why behavioral observations of personality may be useful as a compliment to self-report inventories. First, although self-reports can more directly access mental states and intimate thoughts, these measures are also subject to biases in self-perception and various response tendencies (e.g., Baumeister, Vohs, & Funder, 2007). Even when respondents choose to be completely candid, personality inventories are inevitably limited by self-insight, and there is ample evidence that applicants may not always be honest in their responses (Tett & Christiansen, 2007). On the other hand, observer ratings rely on the ob-

#### Corresponding author:

Andrew B. Speer  
American Family Insurance  
826 Twin Pines Drive, Wisconsin, 53704  
Email: speerworking@gmail.com  
Phone: 989-255-3630

servation of explicit behaviors, describing a person as they are perceived in the “social world” (Funder, 1999; Hogan & Shelton, 1998). As individuals are not always aware of the impact of their behavior on other people, observers (even if not actually involved in the social interaction) may actually be in a better position to gauge this effect (Vazire & Mehl, 2008). In addition, response distortion is less likely an issue for observer ratings (Van Iddekinge, Raymark, & Roth, 2005), as engaging in effective impression management during dynamic interpersonal situations is more challenging than misrepresenting oneself on an inventory in order to obtain a more favorable score.

Perhaps the largest difficulty in using observer ratings of personality is obtaining these ratings in a standardized, relevant, and applicable context. That said, work simulations such as assessments centers (AC) offer a context that meet these criteria. ACs provide a means for measuring competencies across multiple work situations, otherwise known as exercises. Within exercises, assessors directly observe candidates’ behaviors and then rate them according to a set of behavioral dimensions (International Task Force on Assessment Center Guidelines, 2009). Given the standardized procedures involved in administration and scoring, ACs have emerged as a valuable, if costly, way to assess individuals for selection and development purposes.

It is within work simulations such as ACs that we believe other-oriented personality ratings might prove fruitful in selection contexts. Although generally well received in the business and legal communities, ACs are also not without their critics in the research literature. The most notable condemnations involve issues related to the construct validity of dimension ratings (e.g., Jackson, Barney, Stillman, & Kirkley, 2007; Lance, 2008). Without delving too deeply into the vast body of research documenting this phenomenon, a salient concern involves the haphazard way that dimensions are often developed and used (Woehr & Arthur, 2003; Arthur, Day, & Woehr, 2008). For example, dimensions such as inner work standards, personal breadth, inspiring trust, and social flexibility are generally not well-defined and their behavioral domains not well-understood. This calls into question the judgments made by raters on such dimensions and what constructs are actually being assessed.

Because personality dimensions tend to have more well-researched and better understood behavioral domains (Christiansen, Hoffman, Lievens, & Speer, 2013), it may be worthwhile to assess applicant personality alongside or as an alternative to traditionally measured AC dimensions. Many ACs already include self-report personality tests as complimentary measures, and these scores are usually incorporated as part of the feedback candidates receive. From the perspective of the applicant, results based on a combination of behavioral observations of personality made in the AC, along with scores from a self-report inventory,

may illicit better reception while simultaneously achieving similar or greater criterion-related validity (Oh et al., 2011). To our knowledge, there is no readily available and validated assessment for assessing applicant personality in these contexts, and as such the current study attempted to develop a viable method for coding personality-relevant behavior specifically within work simulations such as those found in ACs.

### Development of the Work Simulation Personality Rating Scale

The broad traits of the five-factor model (FFM) were used to organize the behaviors to be rated. The FFM has been successfully used to classify numerous narrow scales from a wide range of inventories, and at this point the construct domains are better understood than any other framework in personality psychology. A number of resources were utilized to identify behaviors for each FFM dimension, with a focus on their applicability in work simulations. First, we examined existing literature for behavioral checklists of AC dimensions that had a high degree of overlap with the FFM traits (see Lievens, Chasteen, Day, & Christiansen, 2006). Second, we inspected past research in personality psychology that has employed behavioral coding systems, most notably the Riverside Behavioral Q-sort (RBQ) developed by Funder and his colleagues (Funder, Furr, & Colvin, 2000). Finally, items from common personality tests were considered with regard to whether the behaviors described might be observed in a work simulation (Goldberg, et al., 2006). We wrote items at an intermediate level of specificity (in between molecular behaviors such as “smiles frequently” and molar behavioral descriptions), so items apply to a wide range of work situations. The developed scale is labeled the Work Simulation Personality Rating Scale (WSPRS), and specifics are discussed in more detail within the Methods section.

After development, the WSPRS was validated within a developmental AC. AC participants were observed across five exercises and then rated using the WSPRS. The resulting WSPRS dimension scores were correlated with self-rated trait scores, with the expectation that behavioral observations would converge with self-report scores of the same trait.

*Hypothesis 1: Work Simulation Personality Rating Scale scores will be positively correlated with self-report scores of the same trait.*

It is commonly accepted that personality judgments will be more accurate when targets are observed in situations that adequately allow expressions of trait-related behavior. Trait activation theory (TAT, Tett & Burnett, 2003) and Funder’s realistic accuracy model (RAM; Funder, 1999) act as useful frameworks to understand this requirement. The

major principle of TAT is that traits will only manifest into behaviors when there are situational cues (i.e. presses) relevant to the trait in question. For instance, a holiday party allows many cues for the trait of sociability to be expressed, whereas being alone in a room allows very little. The opportunity to observe differences in trait-related behaviors can be described by a situation's trait activation potential (TAP), which is conceptually similar to the relevance and availability components of the RAM model. When TAP is high for a given trait, inferences of that personality trait should be more accurate because more behaviors related to that trait will be expressed and with a higher degree of variability across targets<sup>1</sup>.

*Hypothesis 2: Self–other convergence will be stronger for those traits rated as having greater trait activation potential within the AC.*

## METHOD

### Participants and Procedure

Study participants underwent a developmental AC as part of an undergraduate capstone course where only through participation in the AC did students meet requirements for graduation. This AC was designed to reflect managerial job demands that one would encounter in an upper-level management position and had been established as a longstanding simulation for this purpose for 12 years prior to this study. Because of the evaluative nature of the AC and the effect of that evaluation on students' standing in the program, participants had a good degree of motivation to perform within it. Data were collected over a 2-year period, with a total of 123 undergraduate business students undergoing the AC. Participants were videotaped during five interactive AC exercises, and the recorded video segments were later viewed by raters for coding.

The developmental AC consisted of five exercises all designed to reflect managerial job demands: (a) a leaderless group discussion where participants were instructed to take the position of a department manager in a hypothetical organization and come to conclusions concerning a range of administrative issues; (b) a client meeting roleplay during which participants met with a disgruntled client about a complaint concerning one of the participants' employees; (c) a case analysis presentation that involved analyzing a critical issue facing the company, developing the outline of a strategic solution, and then presenting the solution to the board

of the company; (d) a subordinate coaching roleplay where participants met with an employee to discuss their progress and address their poor performance; and (e) a subordinate planning meeting in which candidate met with an employee who had been tasked to come up with detailed plans for a large annual fundraiser. These exercises were relatively brief and could be completed over a half-day period, resulting in videotapes of approximately 2 hours in length.

### Measures

*Work Simulation Personality Rating Scale.* Using the guidelines discussed above, a behavioral observation inventory for assessing personality during work simulations was developed and labeled the Work Simulation Personality Rating Scale. The WSPRS consists of eight behavioral items for each FFM dimension, yielding a total scale of 40 items. Items were designed to indicate how characteristic a particular behavior is of a person using a three-point scale ranging from 1 = *not characteristic* at all to 3 = *very characteristic*<sup>2</sup>. Examples of behaviors rated are “Behaves in an influential and persuasive manner” for Extraversion, “Makes supportive comments” for Agreeableness, “Emphasizes goals and accomplishments” for Conscientiousness, “Says interesting things” for Openness to Experience, and “Acts irritated or annoyed” for Neuroticism (see [Appendix](#) for the entire set of WSPRS items).

Participants were observed across all five exercises and then rated using the WSPRS by a pool of 14 assessors who all had experience making traditional AC ratings. All 14 assessors were I-O psychology graduate students and prior to making WSPRS ratings underwent a 1-day frame-of-reference training session regarding scale use. Following training and several practice sessions of ratings using real AC videos, assessors then viewed participants' behavior across all five videotaped AC exercises, making ratings on each of the 40 behaviors encompassed in the WSPRS after all exercises had been reviewed<sup>3</sup>. Each AC candidate was rated by three randomly assigned raters who were blind to the scores on the self-report inventory. Interrater agreement (ICC 1, 3) for dimension-level composites ranged from .62 to .87 ( $Mn=.76$ ). In terms of the WSPRS instrument itself, when collapsed across raters the internal consistency estimates for the eight-item scales were Extraversion (.94), Agreeableness (.85), Conscientiousness (.91), Openness to Experience (.93), and Neuroticism (.76).

*Self-report personality.* Self-reported personality was

1 Note that this statement will only be true if situational strength (Tett & Burnett, 2003) is not excessively high. Strong situations contain unambiguous behavioral demands, and the outcomes of behavior are clearly understood. Because of this, very strong situations limit behavioral variance and therefore prevent accurate judgments. That being said, it is unlikely a work simulation would exert such strong cues to completely restrict behavioral variance, and if such were the case, it would be a very poor work simulation.

2 Use of 5 or 7 point scales is also encouraged to increase scale variance.

3 Although within exercise ratings may have benefited the AC literature in terms of AC construct validity, the focus of this study was on behavioral observations of personality, of which decisions would be made based on the entire set of participant observations. This, coupled with resource constraints regarding the number of raters needed to make such ratings, was the reasoning behind making ratings only after viewing participants perform across all exercises.

measured using the 20-item mini-IPIP form (Donnellan, Oswald, Baird, & Lucas, 2006). Participants were asked to complete the self-report inventory online 3 to 5 days prior to their AC session. Participants responded to items using a five-point Likert scale ranging from 1 = *strongly disagree* to 5 = *strongly agree*. Internal consistency estimates of reliability of the four-item scales were: Extraversion (.75), Agreeableness (.61), Conscientiousness (.67), Openness to Experience (.47), and Neuroticism (.68).

*Trait activation potential of exercises.* Eight trained assessors made TAP (Tett & Burnett, 2003) ratings for each FFM dimension. The expert assessors consisted of I-O psychology graduate students with previous experience working as raters in multiple ACs, including the AC in this study. TAP gauges the psychological demands of a situation and has been used to describe AC exercises in the past (Haaland & Christiansen, 2002; Lievens et al., 2006; Speer, Christiansen, Goffin, & Goff, 2014; Speer, Christiansen, Melchers, König, & Kleinmann, 2014). In the current study, it was used to predict which traits would be most observable in the AC and therefore which traits would be expected to be most accurately judged. The TAP ratings were made using a process and measure similar to that of Haaland and Christiansen (2002), where assessors rated the relevance of each FFM dimension across all five of the AC exercises. Ratings from all five exercises were then aggregated for each FFM trait to reflect overall TAP scores for the entire AC. Interrater agreement estimates for the TAP scores ranged from .73 for Conscientiousness to .96 for Agreeableness. TAP scores for each FFM dimension are provided in Table 1. As can be seen, expert judgments indicated that Extraversion and Agreeableness had the highest TAP, whereas Conscientiousness and Neuroticism had the lowest.

*Assessment center performance.* In addition to the WSPRS ratings, within each exercise participants were rated by independent, trained raters on a set of traditional AC dimensions. The raters for this task were completely separate from the raters who made WSPRS ratings and underwent extensive training before making ratings. For the sake

of this study, all dimension ratings were aggregated within and across exercises to arrive to an overall assessment center rating (OAR).

## RESULTS

### Convergence Between Work Simulation Personality Rating Scale and Self-Report Personality

Correlations between the dimension scores on the WSPRS and the self-report trait scores are summarized in Table 1, and the full correlation matrix of behavioral ratings, self-ratings, and the OAR can be seen in Table 2. Results indicate the convergence between the WSPRS and self-report scores was moderate. Uncorrected correlations between the two methods of personality assessment were .31 ( $p < .01$ ) for Extraversion, .24 ( $p < .01$ ) for Agreeableness, .22 for Openness ( $p < .01$ ), .18 ( $p < .05$ ) for Conscientiousness, and .11 ( $p = ns$ ) for Neuroticism. Thus, for four of the five dimensions, the uncorrected correlations were positive and significant at the .05 level. Overall, results provide support for Hypothesis 1 in that the WSPRS was related to self-report trait scores for same-trait dimensions.

These correlations can also be considered in the context of the multitrait-multimethod (MTMM) perspective on construct validity. From this perspective, support is found if the monotrait-heteromethod correlations (the convergent validity coefficients described in the previous paragraph) tend to be stronger than the heterotrait-monomethod and heterotrait-heteromethod correlations. To consider average values of these correlations, those for Neuroticism were reversed so that they did not cancel out with other values. In general, only some of the expectations were met. The average convergent validity of .22 was larger than the average heterotrait-heteromethod correlation of .13 and similar to the average heterotrait-monomethod estimate for the self-report method (.26). However, it did not approach the heterotrait-monomethod for the behavioral observation method (.51).

This is a common finding in research using behavioral ratings in personnel assessment such as employment inter-

**TABLE 1.**

*Mean Trait Activation Potential Ratings and Correlations Between Composite Behavioral Observation FFM Ratings and Self-Reported Personality*

	Mean TAP ratings	<i>r</i>	$\rho$
Extraversion	3.88	.31**	.36**
Agreeableness	3.41	.24**	.31**
Openness	3.02	.22**	.30**
Conscientiousness	2.94	.18*	.22*
Neuroticism	2.80	.11	.13

*Notes.*  $N = 123$ . Work Simulation Personality Rating Scale (WSPRS).  $\rho$  is the corrected correlation after correcting for self-report unreliability. TAP = trait activation potential. TAP ratings were on a 1-5 scale ranging from 1 = *low TAP* to 5 = *high TAP*. \*  $p < .05$ . \*\*  $p < .01$ . Significance levels determined based on uncorrected correlations.

**TABLE 2.***Correlations Among Behavioral Observation Ratings, Self-Ratings, and Overall Assessment Center Ratings*

	1	2	3	4	5	6	7	8	9	10
1. BEH_E										
2. BEH_A	.06									
3. BEH_C	.70	.24								
4. BEH_N	-.65	-.30	-.64							
5. BEH_O	.71	.29	.90	-.70						
6. SELF_E	.31	.00	.25	-.16	.21					
7. SELF_A	.08	.24	.13	-.06	.07	.22				
8. SELF_C	-.01	.31	.18	-.13	.14	.11	.44			
9. SELF_N	-.06	-.06	-.12	.11	-.17	-.19	-.08	-.37		
10. SELF_O	.16	.19	.22	-.19	.22	.24	.37	.35	-.26	
11. OAR	.62	.21	.63	-.54	.63	.14	.06	.01	-.03	.07

*Note.* BEH prefixes indicate behavioral ratings of personality, and SELF indicates self-ratings. E = extraversion, A = agreeableness, C = conscientiousness, N = neuroticism, O = openness to experience. OAR = overall assessment center rating.

views and assessment centers where observations are made from a common sampling of behavior that is limited and discrete (Lievens & Christiansen, 2012). Moreover, findings of inflated heterotrait–monomethod correlations extend to many areas of assessment wherein different constructs are measured from common behavioral observations, including studies of patient-management problems with physicians (e. g., Julian & Schumacher, 1988), military examinations (e. g., Shavelson, Mayberry, Li, & Webb, 1990), hands-on science tasks (e. g., Baxter, Shavelson, Goldman, & Pine, 1992), bar examinations (e. g., Klein, 1992), and direct writing assessments (e.g., Dunbar, Koretz, & Hoover, 1991). In general, the more limited the scope of the behavioral sampling, the larger the heterotrait–monomethod correlations become—a finding also found in research on the construct validity of observer ratings of personality (Connelly & Ones, 2010). Expanding the number and types of simulations to allow a broader sampling of behavior across a wider range of situations would be expected to reduce the correlations across the dimensions of the WSPRS.

### Trait Activation Potential and Convergence

As can be seen in Table 1, the rank-orders of the convergent validity estimates, and the TAP ratings of opportunity to observe trait-relevant behaviors were identical. Using the trait as the unit of analysis, a vector correlation was computed between the TAP ratings and the uncorrected convergent validity estimates. The resulting vector correlation was both positive and significant ( $r = .93$ ,  $p < .01$ ), providing support for the hypothesis. Interestingly, the same trend existed for interrater reliability estimates. Extraversion was rated as the trait with the greatest opportunity to observe trait-related behaviors within the AC (highest TAP), and Extraversion was also the dimension with the highest agreement. Likewise, Neuroticism had the lowest

TAP rating and in turn had the poorest agreement. Thus, the degree of convergence between self-report and behaviorally observed personality appears to be dependent on the degree to which there is opportunity to observe trait-related behavior.

### Additional Analyses

Additional item-level psychometric information regarding the WSPRS is provided in Table 3. Included are descriptive statistics, corrected item-total correlations (CITCs), interrater agreement estimates, and correlations pertaining to convergent and discriminant validity. Overall, most items correlated highly with other items in their scales and had high IRR estimates. However, noticeable exceptions did occur with one Agreeableness item (“Argues their opinion or point”) and several of the Neuroticism items (e.g., “Openly emotional and/or volatile”). Although it is possible that these items did not adequately assess their intended constructs, the low CITCs could also be due to a lack of opportunity to observe these specific behaviors in the given situations.

### Localized Scale Revision

Given this was the initial evaluation of the WSPRS, it is possible that improvement can be made to the existing 40-item version of the scale. As such, we considered whether the WSPRS might be improved by identifying poorly functioning items. In doing this, we would be remiss to ignore the role of situational demands and their effects on trait activation. Items with undesirable psychometric properties in the present context may perform well in other settings where there are more cues for activation of the relevant behavior. In essence, it is difficult to separate whether an item is poor in measuring the intended construct or if

**TABLE 3.** *Descriptive Statistics, Reliability, Convergent, and Discriminant Validity of Behavior Observation Ratings of Personality*

Composite/item	Descriptive statistics		Reliability		Convergent validity		Discriminant validity	
	M	SD	IRA	CITC	r	p	r	r
<b>Extraversion</b>								
Behaves in an influential and persuasive manner	17.27	4.04	.87		.31	.36	.04	
Exhibits high enthusiasm and energy	1.92	0.59	.70	.80	.32	.37	.08	
Is talkative	1.79	0.63	.83	.83	.31	.36	.05	
Interacts confidently with others	2.06	0.62	.81	.82	.28	.33	.00	
Expressive with voice, face, and/or gestures	2.23	0.59	.77	.75	.28	.32	.01	
Seems detached from the situation (r)	1.97	0.63	.75	.79	.26	.30	.05	
Appears passive (r)	2.49	0.54	.77	.74	.25	.29	-.11	
Behaves timidly (r) <sup>e</sup>	2.34	0.63	.80	.80	.24	.27	-.03	
	2.44	0.56	.75	.78	.21	.24	-.01	
<b>Agreeableness</b>								
Makes supportive comments	17.93	2.83	.71		.24	.31	.11	
Expresses agreement or support	2.07	0.58	.54	.70	.27	.35	.12	
Acts in a polite manner towards others	2.12	0.46	.57	.70	.27	.35	.11	
Displays concern for others	1.98	0.61	.63	.69	.24	.31	.11	
Supports others' decisions	2.01	0.52	.62	.67	.19	.25	.07	
Behaves dismissively towards others (r)	2.21	0.46	.56	.67	.16	.20	.06	
Behaves in a rude or abrupt manner (r)	2.56	0.46	.57	.66	.12	.15	-.10	
Argues their opinion or point (r) <sup>e</sup>	2.65	0.44	.56	.65	.11	.14	-.11	
	1.93	0.57	.65	.16	.01	.01	.03	
<b>Openness to Experience</b>								
Says interesting things	16.91	3.50	.83		.22	.30	.06	
Discusses multiple aspects of ideas and topics	1.89	0.57	.67	.81	.25	.37	.06	
Considers both pros and cons	2.01	0.52	.55	.80	.25	.36	.09	
Exhibits a high degree of intelligence	1.99	0.57	.66	.82	.22	.32	.07	
Integrates others ideas and suggestions	2.02	0.58	.72	.88	.18	.26	.04	
Contributes new and creative ideas	2.11	0.52	.55	.74	.17	.25	.02	
Makes nonintellectual statements (r) <sup>e</sup>	1.97	0.58	.73	.72	.14	.21	.04	
Unconcerned with different thoughts and ideas (r) <sup>e</sup>	2.46	0.48	.50	.65	.14	.20	-.04	
	2.46	0.48	.57	.60	.09	.13	-.05	
<b>Conscientiousness</b>								
Emphasizes goals and accomplishments	17.39	3.48	.78		.18	.22	.12	
Keeps self or group organized	2.07	0.58	.72	.74	.22	.27	.12	
Consistently stays on task	1.88	0.61	.70	.77	.17	.21	.10	
Does not behave professionally (r)	1.87	0.57	.67	.80	.16	.20	.06	
Prioritizes or plans activities	2.48	0.52	.70	.67	.16	.19	-.04	
	2.07	0.58	.72	.80	.15	.18	.14	

(table continues)

Composite/item	Descriptive statistics		Reliability		Convergent validity		Discriminant validity	
	M	SD	IRA	CITC	r	$\rho$	r	$\rho$
Demonstrates initiative in carrying out tasks <sup>e</sup>	2.46	0.56	.76	.43	.13	.15	.07	.07
Easily distracted and does not follow through (r)	2.56	0.48	.70	.72	.07	.09	-.05	-.05
Considers all options and is thorough <sup>e</sup>	2.00	0.57	.66	.67	.03	.04	.12	.12
<b>Neuroticism</b>	11.79	2.21	.62		.11	.13	-.14	-.14
Appears calm and relaxed (r)	1.60	0.46	.40	.44	.12	.14	.10	.10
Interacts poorly or awkwardly	1.57	0.52	.59	.73	.08	.10	-.11	-.11
Interested in others and tasks (r)	1.77	0.52	.68	.48	.08	.10	.19	.19
Behaves in a nonnormative manner	1.36	0.41	.48	.61	.08	.09	-.08	-.08
Acts irritated or annoyed <sup>e</sup>	1.37	0.42	.55	.25	.08	.09	-.10	-.10
Openly emotional and/or volatile <sup>e</sup>	1.27	0.37	.57	.19	.07	.08	-.05	-.05
Seeks reassurance from others <sup>e</sup>	1.49	0.48	.56	.45	.02	.02	.00	.00
Displays low opinion of self <sup>e</sup>	1.36	0.43	.58	.46	.01	.02	-.02	-.02

Notes.  $N = 123$ . Composite level statistics are listed for each personality trait followed by statistics for each behavior linked to that trait. (r) indicates a trait behavior that was recorded. M = mean. SD = standard deviation. IRA = the interrater agreement for a trait composite or trait behavior across three independent raters. CITC = the corrected item total correlation. r is the uncorrected validity correlation between trait composites/behaviors and self-reported personality scores.  $\rho$  is the validity correlation between trait composites/behaviors and self-report corrected for self-report unreliability. Convergent validity is the correlation between behavioral observations and self-report personality for the same dimension. Discriminant validity is the average off-construct correlation between behavioral observations and self-reported personality. Items with a superscript e were identified and removed for localized scale revision, as discussed in the manuscript text.

that behavior simply couldn't be observed will in the given context. Thus, this effort was only one of maximizing scale functionality within this particular set of situations, making it solely a localized scale revision. For future scale use we recommended use of the entire 40-item version until additional data can be collected investigating scale properties across settings with different situational demands.

We therefore identified items adversely affecting scale reliability and convergent validity with self-reported personality scores, limiting removal to one to four items per trait. These items are identified by superscript "e" in Table 3. Upon removing these items, scale statistics were re-computed. Given the strong intercorrelations among items within scales, edits had minimal impact<sup>4</sup>. After reviewing a couple of the worst performing items in each scale, there were slight improvements in convergent validity and reliability, though the differences were not so great to warrant justification of permanent removal of these items; this is particularly true considering it is difficult to separate whether item-related behaviors are just not relevant to the trait construct or if the specified behaviors were just difficult to observe in the current set of work simulations.

### Relationships With AC Performance

Finally, and as seen in Table 2, the correlations between WSPRS ratings and the OAR were large, ranging from .63 ( $p < .01$ ) for Openness to Experience and Conscientiousness to .21 ( $p < .05$ ) for Agreeableness<sup>5</sup>. This is hardly surprising given that traditional AC ratings correlate with personality (Meriac, Hoffman, Woehr, & Fleisher, 2008), and ratings were made on the same set of observed behaviors. When the OAR was regressed on all five trait scores from the WSPRS, the multiple  $R$  was .69,  $R^2 = .48$ ,  $F(5, 103) = 18.73$ ,  $p < .01$ . Taking a closer examination, Extraversion emerged as the best predictor of the OAR ( $\beta = .32$ ), followed by Conscientiousness ( $\beta = .23$ ), Openness to Experience ( $\beta =$

4 Removal of the worst Extraversion item resulted in similar scale internal consistency (.93), similar interrater reliability (.87), and a minor increase in convergence with self-ratings ( $r = .32$ ,  $p < .01$ ). Removal of the worst Agreeableness item increased internal consistency from .85 to .89, slightly increased interrater reliability (.72), but only increased convergence with self-ratings from .24 ( $p < .01$ ) to .26 ( $p < .01$ ). Two items were removed for Openness, but the change had minimal impact on internal consistency (.93), interrater reliability (.83) and only an increase from .22 ( $p < .01$ ) to .24 ( $p < .01$ ) for convergence with self-ratings. One item was removed for Conscientiousness, and this resulted in similar internal consistency (.90), an increase from .78 to .84 for interrater reliability, and a small gain in self-rating convergence ( $r = .20$ ,  $p < .01$ ). Finally, Neuroticism displayed the worst psychometric qualities, but once again, the TAP for this trait in this particular AC was low. We removed the four most poorly performing items, cutting the scale in half. Upon doing this, we find near identical results for internal consistency (.75), interrater reliability (.63), and self-rating convergence ( $r = .12$ ,  $p = ns$ ).

5 Of the sample of 123 participants who were rated according to the WSPRS, only 109 had ratings on the traditional AC dimensions, making the sample size for these analyses  $N = 109$ .

.12), Neuroticism ( $\beta = -.08$ ), and Agreeableness ( $\beta = .07$ ).

### DISCUSSION

The goal of this study was to develop and evaluate a coding system that can be used to assess personality relevant behaviors within work situations. The resulting tool, the *Work Simulation Personality Rating Scale*, was used to evaluate behavior across five AC exercises. Initial validation evidence was promising. The WSPRS exhibited convergence with self-ratings and especially for those traits with higher trait activation potential in the AC.

In regard to these findings, the absolute magnitude of the correlations between WSPRS and self-report traits were modest. However, magnitude of convergence directly coincided with the TAP scores associated with each of these traits in the AC. These findings are consistent with research by Haaland and Christiansen (2002), who showed self-reported personality better converged with conceptually related AC dimensions when TAP was high. Essentially, judgments will be more accurate when behavior is observed in situations that are relevant to the trait and where interindividual variation in behavior is expected. As a result, care should be taken when using this tool to ensure that all FFM traits can be accurately observed within a work simulation. Otherwise, some trait estimates will have weaker reliability and attenuated validity. The WSPRS focuses raters on detecting and utilizing trait relevant information to make inferences of ratee personality, but ultimately accurate judgments can only be made if relevant ratee information is capable of being observed (Funder, 1999).

It is also worthwhile to consider the degree to which behavioral ratings should correlate with self-ratings. A moderate to large correlation indicates convergence with a known-measure of the same underlying construct and thus demonstrates a degree of validity. However, if behavioral ratings capture trait information not directly gleaned by self-report personality inventories (Connelly & Ones, 2010; Kolar et al., 1996), an extremely strong correlation would not actually be expected. Ratings based on observed behaviors capture unique information; they gauge how individuals *express* traits, and they are capable of capturing information that respondents may not be consciously aware of or choose not to divulge (Baumeister et al., 2007; Vazire & Mehl, 2008). Thus, if observer reports capture unique variance, only a moderate correlation would actually be expected between WSPRS scores and self-reported personality. Larger correlations might be expected, however, with other observer ratings of personality and especially observer ratings from a work context.

### Practical Implications

In applied settings it is very common for personality

inventories to be administered alongside AC exercises to achieve a broader assessment of dispositional tendencies. Utilizing behaviorally based ratings (such as the WSPRS) offers an alternative to this. Beyond potential gains in capturing unique and meaningful variance, acceptance of personality testing may also be higher when scores are based on judgments of actual behavior as opposed to responses on a paper-and-pencil inventory. The application of behaviorally based ratings is certainly not restricted to an AC setting either, as the instrument could be used in various work simulations, panel interviews, and potentially actual work contexts. However, in all these cases it is vital that practitioners consider whether the context allows for expression of trait-related behaviors. Using TAP ratings is a useful method for doing this, and only when a trait is relevant to a given context should behaviorally based ratings be made on that trait.

Of course, from a practical standpoint it is unclear whether feedback on the WSPRS traits would be well received by clients and job candidates, especially in comparison to traditional AC dimensions. Specifically, the traits don't resemble the sort of competencies typically found to be important to business stakeholders, and by nature personality traits are not value laden like AC dimensions. Specific behavioral examples could be provided to support the ratings and might help alleviate some of these concerns. To be clear also, we are not advocating dismissing traditional AC dimensions but rather that WSPRS ratings may be advantageous in combination with these dimensions if an ancillary purpose of a work simulation is to capture personality.

### Limitations and Future Directions

This study had several potential limitations and areas for future research. First, although the WSPRS trait scores were correlated with self-reported personality, we did not look at how they might relate to other criteria such as job performance. Future research should examine how applicants' WSPRS scores relate to job performance over and above those of self-report measures. We also believe it would be worthwhile to incorporate observer ratings of personality from other contexts as an additional criterion measure.

Second, this was only an initial audit of the reliability and validity of the WSPRS; there is certainly room for improvement in terms of item discrimination, interrater agreement, and construct validity. The psychometric statistics found in Table 3 suggest some items could be dropped or revised, and indeed we created a locally revised WSPRS that did just this with minimal improvement. That said, with only a single sample and only one set of work simulations, we are reluctant to make such changes as the results may be a function of the situational demands as much as

the targeted behaviors. For example, an item for the trait of Agreeableness "Argues opinion or point" may have functioned poorly because of lack of situational cues to elicit observable differences between candidates, or it may simply have not measured its intended construct well enough. Future research is needed on this topic. Researchers and practitioners are encouraged to select those items that suit the simulations in use.

Third, although the five factor model is the dominating framework of personality, it is also a broad taxonomy that can dilute more nuanced trait variance. There are many useful frameworks of narrower traits (e.g., NEO-PI-R, Costa & McCrae, 1992 or DeYoung, Quilty, & Peterson's 10-facet taxonomy, 2007), and these could potentially be advantageous from a behavioral observation standpoint. For example, it may be useful to separate enthusiasm and assertiveness instead of utilizing the broader construct of extraversion. In addition, when considering things like situational trait relevance, a more nuanced framework might help understand exactly what types of behaviors are capable of being observed in a given context. Indeed, it is likely FFM traits with lower TAP in our study may have had facets with higher opportunities to observe trait-related behaviors. Although we did not explore a hierarchical trait structure in this study, this may be an area for future WSPRS research. For instance, the eight items in each superordinate scale could potentially be broken into the DeYoung et al. (2007) facets to provide a narrower understanding of individual behavior, and this could be compared to prediction using the broader FFM.

It would also be interesting to apply the WSPRS to *in vivo* settings where trained raters observe people in their jobs and then correlate scores with outcomes of behavior. For example, observers could watch customer service interactions at a sales counter, rate the Extraversion and Agreeableness of the salespersons using the WSPRS, and then survey satisfaction after customers leave a retail store. We suggest that for a given setting assessors might only be able to realistically consider a subset of the traits unless observing across a wide range of contexts. Of course, limiting the range of situations may be necessary due to resource constraints or simply to standardize the situational demands across individuals. In any event, very careful analysis of the trait activation potential of the situation will be necessary.

### CONCLUSIONS

Furr and Funder (2007) recently posited that one reason why so few studies in the personality field utilize observations of actual behavior is that few accepted coding systems have been developed. We suspect that a similar state of affairs exists in the assessment and selection literature in our field. Undoubtedly, different assessment methods have advantages and disadvantages. Although self-report

personality measures are easier and less costly to utilize, it is also important to consider the advantages of utilizing multiple methods to obtain more reliable and accurate estimates of trait elevations. The results of this study show that inferences about personality *can* be drawn from observing simulated work behavior and that these observations can be both reliable and valid. The Work Simulation Personality Rating Scale developed here can serve as a tool to directly assess applicants' personality via behavioral observation in work simulations and aid to further our understanding of how personality affects work behavior.

## REFERENCES

- Arthur, W., Jr., Day, E. A., & Woehr, D. J. (2008). Mend it, don't end it: An alternate view of assessment center construct-related validity evidence. *Industrial and Organizational Psychology, 1*, 105-111.
- Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science, 2*, 396-403.
- Baxter, G. P., Shavelson, R. J., Goldman, S. R., & Pine, J. (1992). Evaluation of procedure-based scoring for hands-on science assessment. *Journal of Educational Measurement, 29*, 1-17.
- Christiansen, N.D., Hoffman, B.J., Lievens, F., & Speer, A.B. (2013). Assessment centers and the measurement of personality. In R.P. Tett, & N. D. Christiansen, (Eds.), *The handbook of personality at work* (pp 477-497). New York, NY: Taylor & Francis/Psychology Press.
- Connelly, B.S., & Ones, D.S. (2010). An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin, 136*, 1092-1122.
- Costa, P. T., Jr., & McCrae, R. R. (1992). NEO PI-R professional manual. Odessa, FL: Psychological Assessment Resources.
- DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology, 93*, 880-896.
- Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R.E. (2006). The mini-IPIP scales: Tiny-yet-effective measures of the Big Five Factors of personality. *Psychological Assessment, 18*, 192-203.
- Dunbar, S. B., Koretz, D. M., & Hoover, H. D. (1991). Quality control in the development and use of performance assessments. *Applied Measurement in Education, 4*, 289-303.
- Funder, D. C. (1999). *Personality judgment: A realistic approach to person perception*. San Diego, CA: Academic Press.
- Funder, D. C., Furr, R. M., & Colvin, C. R. (2000). The Riverside behavioral Q-sort: A tool for the description of social behavior. *Journal of Personality, 68*, 450-489.
- Furr, M. R., & Funder, D. C. (2007). Behavior observation. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 273-291). New York, NY: Guilford Press.
- Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. C. (2006). The International Personality Item Pool and the future of public-domain personality measures. *Journal of Research in Personality, 40*, 84-96.
- Haaland, S., & Christiansen, N. D. (2002). Implications of trait-activation theory for evaluating the construct validity of assessment center ratings. *Personnel Psychology, 55*, 137-163.
- Hausknecht, J. P., Day, D. V., & Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology, 57*, 639-683.
- Hogan, R., & Shelton, D. (1998). A socioanalytic perspective on job performance. *Human Performance, 11*, 129-144.
- International Task Force on Assessment Center Guidelines. (2009). Guidelines and ethical considerations for assessment center operations: International task force on assessment center guidelines. *International Journal of Selection and Assessment, 17*, 243-253.
- Jackson, D. J., Barney, A. R., Stillman, J. A., & Kirkley, W. (2007). When traits are behaviors: The relationship between behavioral responses and trait-based overall assessment center ratings. *Human Performance, 20*, 415-432.
- Julian, E. R., & Schumacher, C. F. (1988, March). *CBT pilot examination: Results and characteristics of CBX*. Paper presented at the conference of the National Board of Medical Examiners on Computer-Based Testing in Medical Education and Evaluation, Philadelphia, PA.
- Klein, S. P. (1992, April). *The effect of content area and test type on bar exam scores*. Paper presented at the National Conference of Bar Examiners, Chicago, IL.
- Kolar, D. W., Funder, D. C., & Colvin, C. R. (1996). Comparing the accuracy of personality judgments by the self and knowledgeable others. *Journal of Personality, 64*, 311-337.
- Lance, C.E. (2008). Where have we been, how did we get there, and where shall we go? *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1*, 140-146.
- Lievens, F., Chasteen, C. S., Day, E. D., & Christiansen, N. D. (2006). Large-scale investigation of the role of trait activation theory for understanding assessment center convergent and discriminant validity. *Journal of Applied Psychology, 91*, 247-258.
- Lievens, F. & Christiansen, N. D. (2012). Core debates

- in assessment center research Dimensions "versus" exercises. In D. Jackson, C.E. Lance, & B. Hoffman (Eds.), *The psychology of assessment centers* (pp 68-94). New York, NY: Routledge.
- Meriac, J. P., Hoffman, B. J., Woehr, D. J., & Fleisher, M. S. (2008). Further evidence for the validity of assessment center dimensions: A meta-analysis of the incremental criterion-related validity of dimension ratings. *Journal of Applied Psychology, 93*, 1042-1052.
- Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology, 60*, 683-729.
- Oh, I., Wang, G., & Mount, M. K. (2011). Validity of observer ratings of the five-factor model of personality traits: A meta-analysis. *Journal of Applied Psychology, 96*, 762-773.
- Shavelson, R. J., Mayberry, P., Li, W., & Webb, N. (1990). Generalizability of job performance measurements: Marine Corps rifleman. *Military Psychology, 2*, 129-144.
- Small, E. E., & Diefendorff, J. M. (2006). The impact of contextual self-ratings and observer ratings of personality on the personality-performance relationship. *Journal of Applied Social Psychology, 36*, 297-320.
- Speer, A. B., Christiansen, N. D., Goffin, R. D., & Goff, M. (2014). Situational bandwidth and the criterion-related validity of assessment center ratings: Is cross-exercise convergence always desirable. *Journal of Applied Psychology, 99*, 282-295.
- Speer, A. B., Christiansen, N. D., Melchers, K. G., König, C. J., & Kleinmann, M. (2014). Establishing the cross-situational convergence of the ability to identify criteria: Consistency and prediction across similar and dissimilar assessment center exercises. *Human Performance, 27*, 44-60.
- Tett, R. P., & Burnett, D. D. (2003). A personality trait-based interactionist model of job performance. *Journal of Applied Psychology, 88*, 500-517.
- Tett, R. P., & Christiansen, N. D. (2007). Personality tests at the crossroads: A response to Morgeson, Campion, Dipboye, Hollenbeck, Murphy, and Schmitt. *Personnel Psychology, 60*, 967-993.
- 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*, 536-552.
- Vazire, S., & Mehl, M. R. (2008). Knowing me, knowing you: The accuracy and unique predictive validity of self-ratings and other-ratings of daily behavior. *Journal of Personality and Social Psychology, 95*, 1202-1216.
- Woehr, D. J., & Arthur, W., Jr. (2003). The construct-related validity of assessment center ratings: A review and meta-analysis of the role of methodological factors. *Journal of Management, 29*, 231-258.

RECEIVED 2/8/2015 ACCEPTED 8/29/2015

## Appendix A. Work Simulation Personality Rating Scale (WSPRS)

Note. Five- and seven-point rating scales may also be used.

<b>Instructions:</b>				
After viewing the participant's behavior during a situation, please mark how characteristic each behavior was of the participant during the situation. Ratings range from (1 = not at all characteristic) to (3 = strongly characteristic). If a rating for a behavior cannot be inferred, mark it as not ratable (NR).				
<b>Extraversion</b>	Not ratable*	Not characteristic		Very characteristic
Exhibits high enthusiasm and energy	<input type="checkbox"/> NR	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
Talkative	<input type="checkbox"/> NR	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
Expressive with voice, face, and/or gestures	<input type="checkbox"/> NR	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
Behaves in an influential and persuasive manner	<input type="checkbox"/> NR	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
Interacts confidently with others	<input type="checkbox"/> NR	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
Behaves timidly	<input type="checkbox"/> NR	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
Seems detached from the situation	<input type="checkbox"/> NR	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
Appears passive	<input type="checkbox"/> NR	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
<b>Notes:</b>				
<b>Agreeableness</b>	Not ratable*	Not characteristic		Very characteristic
Expresses agreement or support	<input type="checkbox"/> NR	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
Supports others' decisions	<input type="checkbox"/> NR	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
Argues their opinion or point	<input type="checkbox"/> NR	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
Acts in a polite manner toward others	<input type="checkbox"/> NR	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
Displays concern for others	<input type="checkbox"/> NR	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
Makes supportive comments	<input type="checkbox"/> NR	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
Behaves dismissively towards others	<input type="checkbox"/> NR	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
Behaves in a rude or abrupt manner	<input type="checkbox"/> NR	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
<b>Notes:</b>				
<b>Conscientiousness</b>	Not ratable*	Not at All characteristic		Very characteristic

(table continues)

Emphasizes goals and accomplishments	<input type="checkbox"/> NR	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
Prioritizes or plans activities	<input type="checkbox"/> NR	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
Considers all options and is thorough	<input type="checkbox"/> NR	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
Keeps self or group organized	<input type="checkbox"/> NR	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
Consistently stays on task	<input type="checkbox"/> NR	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
Easily distracted and does not follow through	<input type="checkbox"/> NR	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
Does not behave professionally	<input type="checkbox"/> NR	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
Demonstrates initiative in carrying out tasks	<input type="checkbox"/> NR	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
<b>Notes:</b>				
<b>Neuroticism</b>	<b>Not ratable*</b>	<b>Not characteristic</b>	<b>Very characteristic</b>	
Seeks reassurance from others	<input type="checkbox"/> NR	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
Acts irritated or annoyed	<input type="checkbox"/> NR	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
Interested in others and tasks	<input type="checkbox"/> NR	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
Appears calm and relaxed	<input type="checkbox"/> NR	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
Interacts poorly or awkwardly	<input type="checkbox"/> NR	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
Openly emotional and/or volatile	<input type="checkbox"/> NR	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
Displays low opinion of self	<input type="checkbox"/> NR	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
Behaves in a nonnormative manner	<input type="checkbox"/> NR	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
<b>Notes:</b>				
<b>Openness</b>	<b>Not ratable*</b>	<b>Not characteristic</b>	<b>Very characteristic</b>	
Says interesting things	<input type="checkbox"/> NR	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
Exhibits a high degree of intelligence	<input type="checkbox"/> NR	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
Integrates others ideas and suggestions	<input type="checkbox"/> NR	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
Contributes new and creative ideas	<input type="checkbox"/> NR	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
Considers both pros and cons	<input type="checkbox"/> NR	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3

(table continues)

Makes nonintellectual statements	<input type="checkbox"/> NR	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
Unconcerned with different thoughts and ideas	<input type="checkbox"/> NR	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
Discusses multiple aspects of ideas and topics	<input type="checkbox"/> NR	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3