

An SJT to predict expressions of subclinical personality disorders at work

Sheila K. Keener, Michael A. McDaniel, & Sven Kepes

Old Dominion University, Work Skills First, Virginia Commonwealth University

Background

- ▶ Cognitive ability is generally considered the best predictor of job performance
- ▶ Unfortunately though, not all individuals with high levels of cognitive ability end up having high job performance.
 - ▶ Personality also influences job performance.

Although cognitive ability is generally considered the best predictor of job performance (e.g., Schmidt & Hunter, 2004), some individuals who perform well on cognitive ability assessments may not ultimately have high job performance. Other factors, such as personality, also influence job performance.

Background

- ▶ Personality exists on a continuum.
 - ▶ Individuals may have symptoms associated with a personality disorder, but not a clinically diagnosable disorder.
 - ▶ Sub-clinical levels of personality disorders can still harm job performance.

For instance, there is evidence to suggest that expressions of certain personality disorders have a negative effect on job performance (e.g., Moscoso & Salagado, 2004). Yet, many psychologists now consider personality to exist on a continuum, such that individuals may have some symptoms associated with personality disorders (and decreased job performance), but not a clinically diagnosable disorder (e.g., De Fruyt & Salagado, 2003; Trull & Durrett, 2005). Therefore, as part of a selection process, it may be beneficial to identify individuals that are likely to exhibit behaviors associated with subclinical levels of personality disorders in the workplace.

Background

- ▶ Situational judgement tests (SJTs) are well-suited to assess personality.
 - ▶ SJTs use workplace-specific scenarios and response options.
 - ▶ SJTs with knowledge instructions are less susceptible to faking than traditional personality inventories.
 - ▶ SJTs provide incremental validity over cognitive ability.

Situational judgement tests (SJTs) are particularly well suited to assessing subclinical levels of personality as part of an employment selection process.

First, SJTs are low fidelity simulations (e.g., Motowidlo, Dunnette, & Carter, 1990) that use workplace-specific scenarios and response options. Second, SJTs with knowledge instructions (e.g., how *effective* is each behavior likely to be) are less susceptible to faking than typical personality inventories (Nguyen, Biderman, & McDaniel, 2005) and may, therefore, provide more accurate assessments than traditional personality inventories. Third, evidence suggests that SJTs provide incremental validity over cognitive ability assessments (McDaniel, Hartman, Whetzel, & Grubb, 2007). Thus, including them in test batteries is like to result in better overall prediction.

Study Goals

- ▶ Develop an SJT that assesses tendencies toward expression of sub-clinical levels of personality disorders in the workplace.
- ▶ Assess the construct validity of the SJT.

Therefore, in this paper, we develop an SJT whose scenarios and responses sought to assess tendencies toward expressions of personality disorders in the workplace. To assess the construct validity of the SJT, we also administered a personality disorder inventory and examined the correlations between predictor and criteria scales.

Method

- ▶ **Sample:** 1,651 undergraduate students and Mturkers
- ▶ **SJT Measure:** 23-scenario SJT with 162 scoreable response options
 - ▶ Four different scoring methods used
 - ▶ Raw consensus
 - ▶ Standardized consensus
 - ▶ Elastic net regression
 - ▶ Extreme gradient boosting

The sample consisted of 1,651 respondents, some of whom were undergraduates in a southeastern university and who completed the measures as part of a psychology department course requirement. The remaining respondents were drawn from adult U.S. residents through the Amazon Mechanical Turk Platform and were paid for their participation.

A 23-scenario SJT with 162 scorable response options was developed. The response options included behaviors that would be attractive to those with subclinical levels of personality disorders. The effectiveness of each of the 162 items were rated by respondents on a 6-point Likert scale. The SJT was scored with two consensus approaches (raw consensus and standardized consensus; McDaniel, Psotka, Legree, Yost, & Weekley, 2011) as well as elastic net regression (Friedman, Hastie, & Tibshirani, 2008) and extreme gradient boosting (Chen & Guestrin, 2006). For the two consensus methods, analyses were also performed with items with mid-range means dropped, as past research suggests validity is improved when these items are removed (McDaniel, Psotka, Legree, Yost, and Weekley, 2011)

Analyses involving elastic net regression and extreme gradient boosting require a training sample to develop the prediction model and a test sample to evaluate the

prediction model. Each PDQ-4 scale has its own elastic net model and its own extreme gradient boosting model. The training sample was 1,000 randomly drawn observations. The remaining 651 observations were the test sample. The raw and consensus scale development did not rely on knowledge of the PDQ-4 derived scales and did not require a training and test sample. However, for consistency across analyses, we developed the consensus scales in the training sample and reported their validity in the test sample. Following McDaniel et al. (2011), the two consensus scoring scales were calculated using all 162 items and again without the items whose means were in the mid-range of the Likert scale.

Method

- ▶ **Personality measure:** Personality Diagnostic Questionnaire
 - ▶ Assesses 12 aspects of personality (only 11 were included in this study)
 - ▶ Scoring methods:
 - ▶ Test publisher instructions
 - ▶ Scale scores were aggregated into higher-order personality factors
 - ▶ Scale scores were summed to create an overall personality disturbance scale.

Respondents were also administered the PDQ-4 (Hyler, 1994), which assesses standing on 12 personality disorders dimensions (Paranoid, Schizoid, Schizotypal, Negativistic, Borderline, Avoidant, Depressive, Dependent, Histrionic, Narcissistic, Obsessive-Compulsive, Antisocial). Due to IRB restrictions, only 11 of the personality disorder dimensions were assessed (Antisocial was not included in the study). The PDQ-4 was scored consistent with the instructions provided by the test publisher. Scale scores on the PDQ-4 were aggregated into higher-order factors and also summed to yield a personality disturbance scale.

Higher-order factors of the PDQ-4 were derived based on the three cluster-model of personality disorders (American Psychiatric Association, 2000, 2013) and confirmatory factor analyses (CFA). CFA models were developed in one random sample and evaluated in a confirmation sample.

All SJT scale correlations with the PDQ-4 derived scales were calculated on the test sample. The SJT scales and the PDQ-4 were scored such that high scores indicated possible subclinical personality disorders.

Results

Model	χ^2	df	CFI	RMSEA	SRMR
Model 1	460.152***	41	.876	.110 (.101, .120)	.062
Model 2	292.404***	32	.911	.099 (.088, .109)	.051
Model 3	177.323***	24	.939	.087 (.076, .100)	.043
Model 4 – Sample 1	109.008***	17	.959	.080 (.066, .095)	.032
Model 4 – Sample 2	134.804***	17	.950	.092 (.078, .107)	.041

We first conducted a series of CFAs to confirm the proposed personality clusters.

Note. *** indicates $p < .001$; Model 1: Cluster A (Paranoid, Schizoid, and Schizotypal personality disorders), Cluster B (Negativistic, Borderline, Histrionic, and Narcissistic personality disorders), Cluster C (Avoidant, Depressive, Dependent, and Obsessive-Compulsive personality disorders); Model 2: Cluster A (Paranoid, Schizoid, and Schizotypal personality disorders), Cluster B (Negativistic, Borderline, and Histrionic personality disorders), Cluster C (Avoidant, Depressive, Dependent, and Obsessive-Compulsive personality disorders); Model 3: Cluster A (Paranoid, Schizoid, and Schizotypal personality disorders), Cluster B (Negativistic and Borderline personality disorders), Cluster C (Avoidant, Depressive, Dependent, and Obsessive-Compulsive personality disorders); Model 4: Cluster A (Paranoid, Schizoid, and Schizotypal personality disorders), Cluster B (Negativistic and Borderline personality disorders), Cluster C (Avoidant, Depressive, and Dependent personality disorders)Table Note.

Note that the results of the CFAs found that three of the 11 personality disorders loaded on multiple factors (i.e., narcissistic, histrionic, and obsessive-compulsive personality disorders). Therefore, these were excluded from analyses at the cluster-level.

Results

Constructs	Raw Consensus		Standardized Consensus		Elastic Net	Extreme Gradient Boosting
	All items	Midrange items dropped	All items	Midrange items dropped	All items	All items
Cluster A “odd or eccentric”	.05	.12	.18	.20	.30	.30
Cluster A subscales						
Paranoid	.03	.08	.14	.14	.27	.25
Schizoid	.09	.11	.18	.15	.24	.23
Schizotypal	.00	.09	.11	.16	.29	.27

Correlations of test sample SJT scales (N = 651) with Cluster A personality disorder constructs by type of SJT scoring method.

Results

Constructs	Raw Consensus		Standardized Consensus		Elastic Net	Extreme Gradient Boosting
	All items	Midrange items dropped	All items	Midrange items dropped	All items	All items
Cluster B “dramatic, emotional, or erratic”	-.08	-.01	.02	.08	.22	.19
Cluster B subscales						
Negativistic	-.06	.00	.03	.07	.16	.13
Borderline	-.07	-.01	.02	.07	.19	.16

Correlations of test sample SJT scales (N = 651) with Cluster B personality disorder constructs by type of SJT scoring method.

Results

Constructs	Raw Consensus		Standardized Consensus		Elastic Net	Extreme Gradient Boosting
	All items	Midrange items dropped	All items	Midrange items dropped	All items	All items
Cluster C "anxious or fearful"	-.08	-.05	.00	.03	.13	.19
Cluster C subscales						
Avoidant	-.08	-.07	-.02	.00	.15	.18
Depressive	-.08	-.08	-.01	-.02	.11	.14
Dependent	-.03	.04	.05	.10	.17	.17

Correlations of test sample SJT scales (N = 651) with Cluster C personality disorder constructs by type of SJT scoring method.

Results

Constructs	Raw Consensus		Standardized Consensus		Elastic Net	Extreme Gradient Boosting
	All items	Midrange items dropped	All items	Midrange items dropped	All items	All items
Scales not included in clusters						
Histrionic ^a	.03	.11	.05	.11	.25	.27
Narcissistic ^b	.03	.15	.16	.23	.35	.34
Obsessive-Compulsive ^c	-.02	-.01	-.02	.00	.14	.10
Personality Disturbance ^d	-.03	.05	.09	.14	.25	.25

Correlations of test sample SJT scales (N = 651) with personality disorder constructs not included in clusters by type of SJT scoring method.

Discussion

Summary

- ▶ Raw and standardized consensus scales performed poorly compared to elastic net regression and extreme gradient boosting scales
- ▶ Scores on the SJT best predicted scores on:
 - ▶ Cluster A scales
 - ▶ Paranoid, Schizoid, Schizotypal, Histrionic, & Narcissistic subscales
 - ▶ Overall personality disturbance scale

Both raw and standardized consensus scales performed poorly relative to the elastic net regression and extreme gradient boosting scales. The SJT was best at predicting the Cluster A (“odd or eccentric”) scale and the Paranoid, Schizoid, Schizotypal, Histrionic, and Narcissistic personality disorder subscales, as well as the overall personality disturbance scale.

Therefore, this study shows that SJTs can be an effective way to identify individuals who likely to display behaviors associated with subclinical levels of personality disorders at work. Results of this study also indicated that the two most common methods of scoring SJTs (the raw consensus and standardized consensus methods) did not perform as well as elastic net regression and extreme gradient boosting. Thus, other SJTs may benefit from the use of these methods. Future research should build on this study by examining the predictive validity of this SJT, as well as its incremental validity over cognitive ability and traditional measures of the Big 5 personality traits.