

Executive Summary

A retail clothing company asked us to evaluate a set of assessment items they were considering for hiring sales associates. The company had been hiring through unstructured interviews and wanted to move to an evidence-based process. Our job was to figure out which items actually predict job performance, and whether using them to hire would unfairly disadvantage any group of applicants. We worked with data from 225 current sales associates, including responses to 15 assessment items, a cognitive ability test, performance ratings, and records of rule infractions. Three useful predictor scales emerged from the assessment items: Counterproductive Work Behavior (CWB), Conscientiousness (CON), and Emotional Stability (ES). The CWB scale measures self-reported attitudes that signal a tendency toward rule-breaking; it is used here as a predictor administered before hire, not as an outcome. Cognitive ability rounded out the set of four predictors. Our main recommendation is that the company use three of these predictors (cognitive ability, conscientiousness, and CWB) to score applicants, weighted equally. This combination predicts job performance and infractions, and produces less adverse impact than the alternatives we tested. Emotional stability should not be used because it adds little predictive value and creates a gender gap in scores. The rest of this report walks through how we got to that recommendation.

Data Overview

The dataset for this analysis included responses from 225 current sales associates. It contained demographic information (e.g., race, gender), cognitive ability scores, responses to 15 job-relevant assessment items, two performance ratings for each employee, and records of infractions.

Data Preparation

We observed that some of the 15 items appeared to measure the same thing. To identify and group these items, we conducted an exploratory factor analysis (EFA).

Three key factors emerged:

1. Counterproductive Work Behaviors (CWB): Items 1, 2, 9, 10, and 11, reflecting undesirable workplace behaviors.
2. Emotional Stability (ES): Items 4, 6, 7, 13, and 15, assessing emotional control.
3. Conscientiousness (CON): Items 3, 5, 8, 11, and 14, reflecting effort, quality, and thoroughness in work.

These factors, along with the score of cognitive ability, form our four key predictor variables.

Methodology

We first evaluated the reliability and validity of the measures. To assess the reliability of the extracted factors from EFA, internal consistency was evaluated using Cronbach's alpha for

each of the identified factors, and a minimum threshold of 0.70 was set for acceptable reliability. Content validity was evaluated through subjective judgment by psychologists familiar with the domain, and the divergent validity was examined by checking that the factors were not highly correlated with each other.

The criterion variables (i.e., outcomes related to success) were the performance ratings and the number of infractions. The performance ratings, provided by two judges, were assessed for interrater reliability and then averaged. To create a composite score combining both outcomes, infractions were reverse-coded, and both the ratings and infractions were converted to z-scores before averaging. Criterion-related validity was assessed by analyzing correlations between predictors and the three outcome variables, with adjustments made for interrater unreliability in performance ratings. Subgroup differences by race and gender were analyzed using t-tests.

Next, we conducted multiple regression analyses to explore the relationship between the predictors (the extracted factors and cognitive ability) and the criterion (performance ratings, infractions, and a composite score). Incremental validity was also assessed to evaluate how much each predictor contributed to the model's overall predictive power.

To provide a clear instruction on how to use the predictors, we explored two methods: unit weighting (equal weight for each predictor) and optimal weighting (weights based on regression analysis). The z-scores of the meaningful predictors were combined into an overall score for each employee, with CWB reverse-coded before combining. Adverse impact was assessed for both methods to ensure fairness across demographic groups.

Results and Recommendations

The three extracted factors each had a Cronbach's alpha of .80, demonstrating good reliability. A panel of professional I-O psychologists confirmed that the measures reflect the intended constructs. Additionally, the predictors do not correlate with each other too highly, and they correlate with each of the three outcome variables; please see Table 2. The interrater reliability of the ratings of performance was estimated to be .75. We corrected for attenuation due to unreliability in criterion, please see Table 3. Overall, the predictors are shown to have good reliability and validity.

T-tests (Tables 4 and 5) revealed subgroup differences. Cognitive ability scores were higher for white employees compared to non-white employees, with a large difference (Cohen's $d = 1.00$). Emotional stability scores were higher for male employees than female employees, with a small-to-moderate difference (Cohen's $d = 0.27$). Given these subgroup differences, caution is advised when using cognitive ability and emotional stability as predictors to avoid potential adverse impact.

The results and recommendations from the multiple regression and incremental validity analyses are summarized below. For detailed results, refer to Tables 6(a), 6(b), 7(a), 7(b), 8(a), and 8(b).

1. Outcome: **Infractions** – CWB is the only meaningful predictor of infractions. Those who score lower on CWB are less likely to commit infractions than those who score higher on CWB. If reducing infractions is a priority, we recommend adding tools to assess CWB in the hiring process.
2. Outcome: **Performance Ratings** – Cognitive ability is a strong predictor of performance, explaining 24% of the variance. Despite showing subgroup differences across races, given its predictive ability of performance, we recommend keeping cognitive ability as part of the selection process while being mindful of its negative effect on the company's diversity goals. Conscientiousness also emerged as a significant predictor, explaining 14% of additional variance, so including this trait as a predictor is recommended. Emotional stability adds only 2% to the explained variance and shows gender-based differences, increasing the risk of adverse impact. We suggest excluding emotional stability. Finally, CWB does not predict performance ratings, so it should not be used for this outcome.
3. Outcome: **Composite Score (Infractions and Performance Ratings)** – For a combined measure of infractions and performance, cognitive ability, conscientiousness, and CWB are all effective predictors. Emotional stability, however, is not significant and should be excluded.

In this report, we examined how to use predictors to maximize the composite score, with cognitive ability, conscientiousness, and CWB as predictors. To calculate the adverse impact ratio, we created a hypothetical scenario in which the top 100 applicants were hired.

For unit weighting, results showed that 83 white applicants and 17 non-white applicants would be hired. The adverse impact ratio for race was calculated as $(17/61)/(83/164)=0.55$. For gender, 49 females and 51 males would be hired, with an adverse impact ratio of $(49/120)/(51/105)=0.84$. In optimal weighting, the predictors were assigned the following weights: CWB (0.31), conscientiousness (0.31), and cognitive ability (0.38). Results showed that 86 white applicants and 14 non-white applicants would be hired. The adverse impact ratio for race was $(14/61)/(86/164)=0.437$. For gender, 49 females and 51 males would be hired, with an adverse impact ratio of $(49/120)/(51/105)=0.84$.

Although both scenarios had a race-related adverse impact ratio lower than ideal (0.8), the use of multiple predictors besides cognitive ability helped offset this effect, making the impact acceptable. We recommend using unit weighting, because the adverse impact ratio for race was lower than in optimal weighting. Additionally, because the sample size is small, optimal weighting might be susceptible to shrinkage, making it less applicable to other samples. Overall, the company should use CWB, CON, and COG and weigh them equally to select future employees.

Appendix

Tables Representing the Statistical Analyses Conducted for this Report

Table 1

Frequency Table for Race and Sex

Race		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Non-White	61	27.11	27.11	27.11
	White	164	72.89	72.89	100
	Total	225	100	100	—
Sex		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Male	105	46.67	46.67	46.67
	Female	120	53.33	53.33	100
	Total	225	100	100	—

Table 2

Means, standard deviations, and correlations for our predictor and outcome variables

Variable	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8
1. COG	107.07	13.54								
2. Rating 1	6.05	1.71	.41***							
3. Rating 2	6.18	1.71	.46***	.60***						
4. Avg Rating	6.12	1.53	.49***	.90***	.90**					
5. Infractions	0.63	0.88	.02	.08	.17*	.14*				
6. CWB	3.98	0.86	-.06	.03	.02	.03	.36***			
7. CON	3.98	0.83	-.00	.36***	.32**	.38***	.07	.16*		
8. ES	3.99	0.85	.07	.18**	.19**	.21**	.09	.23***	.10	
9. Composite	0.00	0.66	.36***	.62***	.55***	.66***	-.66***	-.26***	.24***	.09

Note. $n = 225$ for all variables. COG represents cognitive ability, CWB represent counterproductive workplace behaviour, CON represents conscientiousness, ES represents emotional stability, and Composite refers to the composite score combining the z-scores of average ratings and reversed infractions. * indicates $p < .05$. ** indicates $p < .01$. *** indicates $p < .001$

Table 3*Corrected Correlations Between Predictors and Performance*

Variable	Performance
1. CWB	.03
2. ES	.24**
3. CON	.44***
4. COG	.56***

Note. $n = 225$ for all variables. COG represents cognitive ability, CWB represent counterproductive workplace behaviour, CON represents conscientiousness, and ES represents emotional stability.

Table 4*T-Tests for Gender Differences in Predictors*

Variable	Levene's Test for Equality of Variances		T-test for Equality of Means				95% Confidence Interval			
	<i>F</i>	<i>Sig.</i>	<i>t</i>	<i>df</i>	<i>Sig(1-tailed)</i>	<i>Sig(2-tailed)</i>	Mean Difference	Std. Error Difference	Lower	Upper
ES	3.87	.050	2.00	223	.023	.047	0.27	0.13	0.004	0.53
CON	0.36	.552	1.61	223	.054	.108	0.21	0.13	-0.05	0.48
CWB	1.43	.232	-0.32	223	.374	.747	-0.04	0.13	-0.31	0.22
COG	0.08	.773	1.58	223	.058	.116	0.21	0.13	-0.05	0.47

Note. $n = 225$ for all variables. COG represents cognitive ability, CWB represent counterproductive workplace behaviour, CON represents conscientiousness, and ES represents emotional stability.

Table 5*T-Tests for Race Differences in Predictors*

Variable	Levene's Test for Equality of Variances				T-test for Equality of Means				95% Confidence Interval	
					Significance		Mean Difference	Std. Error Difference	Lower	Upper
	<i>F</i>	<i>Sig.</i>	<i>t</i>	<i>df</i>	<i>Sig(1-tailed)</i>	<i>Sig(2-tailed)</i>				
ES	0.16	.694	0.21	223	.170	.341	-0.14	0.15	-0.44	0.15
CON	0.07	.792	-0.96	223	.339	.677	0.06	0.15	-0.23	0.36
CWB	0.91	.341	-0.21	223	.419	.838	-0.03	0.15	-0.32	0.27
COG	0.19	.666	-6.67	223	<.001	<.001	-0.92	0.14	-1.19	0.65

Note. $n = 225$ for all variables. COG represents cognitive ability, CWB represent counterproductive workplace behaviour, CON represents conscientiousness, and ES represents emotional stability.

Table 6(a)*Multiple Regression Analysis Predicting Infractions*

Predictors	B	SE	β	t	p	R ²	adj. R ²	F
Intercept	-1.20	0.59		-2.05	.042	.14	.12	F(4, 220) = 8.58, p < .001
COG	0.00	0.00	0.04	0.64	.521			
CWB	0.37	0.07	0.36	5.59	<.001			
CON	0.01	0.07	0.01	0.13	.893			
ES	0.01	0.07	0.01	0.12	.908			

Note. $n = 225$ for all variables. COG represents cognitive ability, CWB represent counterproductive workplace behaviour, CON represents conscientiousness, and ES represents emotional stability. * indicates $p < .05$. ** indicates $p < .01$. *** indicates $p < .001$.

Table 6(b)*Summary of Hierarchical Regression Analysis Using Cognitive Ability, Counterproductive Workplace Behaviour, Conscientiousness, and Emotional Stability as Predictors for Infractions*

Predictor Variables	Step 1			Step 2			Step 3			Step 4		
	B	(SE)	B	B	(SE)	β	B	(SE)	β	B	(SE)	β
COG	0.00	0.00	0.02	0.00	0.00	0.04	0.00	0.00	0.04	0.00	0.00	0.04
CWB				0.38	0.06	0.37***	0.37	0.07	0.37***	0.37	0.07	0.36***
CON							0.01	0.07	0.01	0.01	0.07	0.01
ES										0.01	0.07	0.01
R ² _{Cumulative}	0.00			0.14			0.14			0.14		
F _{Change}	0.08			34.49			0.02			0.01		

Note. $n = 225$ for all variables. COG represents cognitive ability, CWB represent counterproductive workplace behaviour, CON represents conscientiousness, and ES represents emotional stability. * indicates $p < .05$. ** indicates $p < .01$. *** indicates $p < .001$.

Table 7(a)

Multiple Regression Analysis Predicting Job Ratings

Predictors	B	SE	β	t	p	R ²	adj. R ²	F
Intercept	-3.15	0.85		-3.69	<.001	.40	.39	F(4, 220) = 36.63, p <.001
COG	0.05	0.01	0.47	9.03	<.001			
CWB	-0.06	0.10	-0.03	-0.62	.537			
CON	0.68	0.10	0.37	6.94	<.001			
ES	0.27	0.10	0.15	2.75	.006			

Note. n = 225 for all variables. COG represents cognitive ability, CWB represent counterproductive workplace behaviour, CON represents conscientiousness, and ES represents emotional stability. * indicates p < .05. ** indicates p < .01. *** indicates p < .001.

Table 7(b)

Summary of Hierarchical Regression Analysis Using Cognitive Ability, Counterproductive Workplace Behaviour, Conscientiousness, and Emotional Stability as Predictors for Job Performance Ratings

Predictor Variables	Step 1			Step 2			Step 3			Step 4		
	B	(SE)	β	B	(SE)	β	B	(SE)	β	B	(SE)	β
COG	0.05	0.01	0.49***	0.06	0.01	0.49***	0.06	0.01	0.49***	0.05	0.01	0.47***
CWB				0.10	0.10	0.06	0.00	0.10	0.00	-0.06	0.10	-0.03
CON							0.70	0.10	0.38***	0.68	0.10	0.37***
ES										0.27	0.10	0.15**
R ² _{Cumulative}	0.24			0.24			0.38			0.40		
F _{Change}	68.85			1.00			49.74			7.57		

Note. n = 225 for all variables. COG represents cognitive ability, CWB represent counterproductive workplace behaviour, CON represents conscientiousness, and ES represents emotional stability. * indicates p < .05. ** indicates p < .01. *** indicates p < .001.

Table 8(a)

Multiple Regression Analysis Predicting Composite Criterion

Predictors	B	SE	β	t	p	R ²	adj. R ²	F
Intercept	-1.98	.40		-4.90	<.001	.27	.26	F(4, 220) = 20.22, p <.001
COG	.02	.00	.33	5.69	<.001			
CWB	-.23	.05	-.30	-5.05	<.001			
CON	.22	.05	.27	4.67	<.001			
ES	.08	.05	.11	1.80	.073			

Note. $n = 225$ for all variables. COG represents cognitive ability, CWB represent counterproductive workplace behaviour, CON represents conscientiousness, and ES represents emotional stability. * indicates $p < .05$. ** indicates $p < .01$. *** indicates $p < .001$.

Table 8(b)

Summary of Hierarchical Regression Analysis Using Cognitive Ability, Counterproductive Workplace Behaviour, Conscientiousness, and Emotional Stability as Predictors for Composite Criterion

Predictor Variables	Step 1			Step 2			Step 3			Step 4		
	B	(SE)	β	B	(SE)	β	B	(SE)	β	B	(SE)	β
COG	0.02	0.00	0.36***	0.02	0.00	0.34***	0.02	0.00	0.34***	0.02	0.00	0.33***
CWB				-0.18	0.05	-0.23***	-0.21	0.04	-0.28***	-0.23	0.05	-0.30***
CON							0.22	0.05	0.28***	0.22	0.05	0.27***
ES										0.08	0.05	0.11
R ² _{Cumulative}	.13			.18			.26			.27		
F _{Change}	32.19			14.88			22.93			3.24		

Note. $n = 225$ for all variables. COG represents cognitive ability, CWB represent counterproductive workplace behaviour, CON represents conscientiousness, and ES represents emotional stability. * indicates $p < .05$. ** indicates $p < .01$. *** indicates $p < .001$.