VALIDITY STUDY

Hiring Certainty with Science.

How the **PI Behavioral Assessment** outperforms alternative selection methods in real-world hiring.

OVERVIEW

Ecological Validity Study

To demonstrate the ecological validity and practical utility of the PI Behavioral Assessment (BA) in real-world selection contexts, a quasi-experimental predictive criterion validity design was employed. This study compares post-hire outcomes between hiring managers who utilized the BA during candidate selection (Experimental Group) and those who relied on alternative selection methods (Control Group).

Study Groups

EXPERIMENTAL GROUP	CONTROL GROUP
Recruited via Rally Research, PI's client network. Active hiring managers who utilized the PI Behavioral Assessment to hire a candidate, with at least one hire completed more than 90 days ago.	Recruited via Cloud Research, a crowdsourcing platform. Active hiring managers who used any hiring instrument besides the PI Behavioral Assessment, with at least one hire completed more than 90 days ago. Represents a "business as usual" baseline.
394 hires evaluated	**183 hires evaluated**

Total: N = 577 responses retained after data cleaning, analyzed using Linear Mixed-Effects Models (LMM) with random intercepts per hiring manager to control for individual manager variance.

Control Group: Selection Tools Used

Among the most popular pre-hire tools used by the Control Group:

Tool	% of Control Group
Pre-employment personality assessments	30%
Pre-employment skill-based or job knowledge tests	29%
Pre-employment cognitive ability assessments	28.4%
In-depth, structured interviews	27.3%

Among personality assessments used pre-hire: The Big Five (50%), Myers Briggs (34%), Hogan Personality Inventory (34%).

METHODOLOGY

Measures

Hiring managers were first prompted to provide a list of names and job titles for up to five recent hires — those hired more than 90 days ago to ensure sufficient time to assess job performance and work outcomes. Participants were asked to list as many hires as they could think of, selecting as randomly as possible, not only listing top performers.

Job Performance

Task proficiency was measured using two items adapted from Williams & Anderson (1991), assessing the extent to which the employee meets formal performance requirements and the relative value of their total contribution compared to the average employee.

Extra Role Performance (OCBs & CWBs)

Voluntary behaviors were assessed using four items adapted from Spector (2006), capturing both Organizational Citizenship Behaviors (OCB) (e.g., volunteering for extra tasks) and Counterproductive Work Behaviors (CWB) (e.g., low effort, disrespect) at the interpersonal and organizational levels.

Retention & Turnover

Operationalized as current employment status. For separated employees, managers indicated whether the exit was voluntary or involuntary and the time to separation; for active employees, tenure duration was recorded.

Perceived Fairness

Hiring process fairness was measured using three items covering Procedural Justice (adapted from Niehoff & Moorman, 1993) and Distributive Justice (perceived qualification of the selected candidate and confidence in the hiring decision).

Utility

Two novel items assessing the degree to which the hiring manager relied on the assessment for the selection decision and subsequent management/communication.

Promotability

Likelihood of handling greater scope; adapted from Kiker & Motowidlo, 1999

Culture Fit

Alignment with core values; adapted from Cable & Judge, 1996

Contextual Data & Data Cleaning

To better understand the role, contextual data were collected including role salary, average tenure for the specific position, and industry. Rigorous screening criteria were applied to ensure data integrity: respondents were removed if they did not satisfy 2 specific attention checks. For the Control Group (crowdsourcing route), 3 additional attention checks were included, plus a ReCaptcha score provided through Qualtrics.

Statistical Analysis

A series of Linear Mixed-Effects Models (LMM) were specified, entering role complexity (salary), employee tenure, and post-hire tool usage intensity as fixed covariates to isolate the treatment effect. To account for the non-independence of observations inherent in the data structure (where multiple hires were nested within a single manager), a random intercept was included for each rater to control for individual manager variance.

FINDINGS

Table 1. Descriptive Statistics for Outcome Variables

Variable	Control Group		Experimental Group	
	Control_Mean	Control_SD	Experimental_Mean	Experimental_SD
Job Performance	4.27	0.87	4.54	0.69
OCBs	3.77	0.95	3.98	0.74
CWBs	1.98	1.08	1.82	0.95
Distributive Justice	4.10	0.94	4.49	0.62
Procedural Justice	4.37	0.93	4.11	1.23
Promotability	3.49	1.23	4.11	0.94
CultureFit	4.07	0.99	4.58	0.63
Value Produced	3.64	1.28	3.89	0.95

** Statistically significant advantage for BA group (p < .01). N = 577; OCBs = Organizational Citizenship Behaviors; CWBs = Counterproductive Work Behaviors; Control Group = Candidates hired through means other than the Behavioral Assessment; Experimental Group = Candidates hired through the Behavioral Assessment.*

Table 2. Linear Mixed-Effects Models Predicting Organizational Outcomes by Selection Method

Estimates represent unstandardized regression coefficients (b), interpreted as the difference in the dependent variable score for the Behavioral Assessment group relative to the Control group, holding covariates constant. Positive values indicate higher scores for the BA group.

Variable	Estimate	p value
Job Performance	0.31	0.00
OCBs	0.19	0.08
CWBs	-0.09	0.65
Distributive Justice	0.42	0.00
Procedural Justice	-0.28	0.08
Promotability	0.48	0.00
CultureFit	0.59	0.00
Value Produced	0.20	0.10

FINDINGS

Robust Support for Predictive Validity

The analysis revealed robust support for the assessment's predictive validity across key organizational criteria as opposed to other tools. The strongest effects were observed in the domain of person-organization fit and future potential.

Culture Fit

b = 0.59, p < .001

Employees selected using the BA demonstrated significantly higher Culture Fit compared to the control group.

Promotability

b = 0.48, p < .001

Employees selected using the BA demonstrated significantly higher Promotability compared to the control group.

Distributive Justice

b = 0.42, p < .001

The assessment demonstrated clear utility in predicting tangible outcomes, with the BA group scoring significantly higher on Distributive Justice.

Job Performance

b = 0.31, p < .01

The BA group scored significantly higher on core Job Performance compared to the control group.

Collectively, these findings suggest that the use of the Behavioral Assessment provides incremental validity over standard selection methods, particularly in identifying candidates who are not only task-proficient but also culturally aligned and capable of long-term growth.

The Behavioral Fit Rating Effect

Within the experimental group, whether managers considered the behavioral fit rating (i.e., how closely aligned their BA results were to the role) as part of the hiring decision was strongly linked to:

Job Performance	**b = 0.42, p < .01**
Culture Fit	**b = 0.53, p < .01**

CONCLUSIONS

Ecological Validity Summary

This study utilizes a quasi-experimental design that directly pits the Behavioral Assessment against alternative selection methods currently active in the market. Rather than relying on theoretical correlations, these results demonstrate ecological validity, suggesting that the BA provides a distinct competitive advantage in real-world hiring scenarios where other tools are present.

Ultimately, this head-to-head comparison confirms that the Behavioral Assessment effectively captures the nuance of candidate fit that other instruments miss, resulting in hires who are not only more capable but significantly more aligned with the organization's future.

Summary of Significant Effects

Outcome	Estimate (b)	p value	Interpretation
Culture fit	.59	<.001	Strongest effect; highest alignment with organizational values
Promotability	.48	<.001	Greater likelihood of handling increased scope and responsibility
Distributive Justice	.42	<.001	Higher confidence in candidate qualification and hiring decision quality
Job performance	.31	<.01	Significantly higher task proficiency and overall contribution

METHODOLOGY NOTE

Quasi-experimental predictive criterion validity design. N = 577 hires nested within 181 managers (Experimental Group: 394 hires via Rally Research; Control Group: 183 hires via Cloud Research). Fixed covariates: role salary, employee tenure, post-hire tool usage intensity. Random intercept per hiring manager. Measures adapted from: Williams & Anderson (1991), Spector (2006), Niehoff & Moorman (1993), Kiker & Motowidlo (1999), Cable & Judge (1996). Data cleaning: attention checks + reCAPTCHA scoring.