

Evaluating Construct Convergence: A Canonical Correlation Analysis of Diagnostic Classification Model-Based SJT and IPIP-NEO-PI-120

Medianta Tarigan^{1,3*}, Rahmat Hidayat¹, and Agung Santoso²

¹Faculty of Psychology, Gadjah Mada University, Yogyakarta, Indonesia

²Faculty of Psychology, Sanata Dharma University, Yogyakarta, Indonesia

³Department of Psychology, Indonesia University of Education, Bandung, West Java, Indonesia

Abstract. The Five-Factor Model (FFM) is the most widely used framework for understanding personality, typically assessed through self-report inventories such as the IPIP-NEO-120. Although these tools show strong psychometric properties, reliance on self-reports introduces vulnerabilities to bias, including social desirability, response distortion, and faking, particularly in high-stakes contexts like personnel selection. To address such limitations, researchers have increasingly explored behavior-oriented measures, notably Situational Judgment Tests (SJTs). When analyzed using Cognitive Diagnostic Models (CDMs), SJTs can yield multidimensional mastery profiles that offer a nuanced representation of personality. However, limited evidence exists regarding the convergence between CDM-based SJT profiles and traditional self-reports. This study applied Robust Canonical Correlation Analysis (RCCA) with a Spearman rank correlation matrix to examine structural alignment between CDM-based SJT scores and IPIP-NEO-120 domain scores in a sample of 289 participants. RCCA was chosen to handle common violations of multivariate assumptions, such as non-normality and outliers. Results showed a moderate canonical correlation in the first root, supporting convergent validity while highlighting distinct trait representations across methods. Subsequent roots yielded weaker and less stable associations. Overall, findings indicate that RCCA is a viable approach for evaluating cross-method validity and emphasize the complementary insights of situational and self-report approaches in personality assessment.

1 Introduction

The Five-Factor Model (FFM), or Big Five, is the most widely accepted framework in personality psychology, organizing traits into Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness [1]. It has been operationalized through instruments such as the NEO and IPIP-NEO-120, both demonstrating strong applicability and cross-cultural generalizability [2]. However, as self-report measures, these instruments are vulnerable to biases including social desirability, response distortion, and faking,

* Corresponding author: mediantatarigan@mail.ugm.ac.id

particularly in high-stakes contexts [3]. To address these limitations, researchers have turned to behavior-oriented assessments such as Situational Judgment Tests (SJTs), which present contextually rich scenarios, mirror real-world decision-making, and are more resistant to faking [4]. While SJTs have traditionally assessed social and non-cognitive skills, growing evidence indicates their potential for measuring personality when responses are mapped to FFM constructs [5]. A major challenge, however, lies in the multidimensional nature of SJT responses, which often reflect multiple traits simultaneously. Cognitive Diagnostic Models (CDMs) offer a solution by linking responses to latent traits through a Q-matrix, yielding nuanced and interpretable profiles aligned with personality theory [6]. Yet, the correspondence between CDM-based SJT profiles and established self-report measures remains to be tested. Canonical Correlation Analysis (CCA) provides a multivariate approach for evaluating shared variance between such methods [7], but its reliance on strict assumptions limits robustness in psychometric applications. To overcome this, the present study employs Robust Canonical Correlation Analysis (RCCA) with a rank-based Spearman approach to compare CDM-based SJT scores with those from the IPIP-NEO-120, aiming to evaluate the structural alignment of these two approaches and to determine the extent to which personality traits are similarly represented across situational and self-report assessments.

2 Methods

2.1 Instrument

2.1.1 *Situational Judgement Test (SJT) for the Five-Factor Model (FFM)*

The SJT employed in this study was a version that had been previously developed during the initial phase of the research. The instrument consists of a series of situational dilemmas designed to elicit behavioral responses that are theoretically aligned with the five core domains of the Five-Factor Model (FFM): Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness to Experience. This instrument was developed using a Diagnostic Classification Model (DCM) approach.

2.1.2 *IPIP-NEO 120*

The IPIP-NEO-120 is a self-report personality inventory consisting of 120 items rated on a five-point Likert scale (1 = “Very Inaccurate” to 5 = “Very Accurate”), measuring the five domains of the FFM. The Indonesian version has been adapted and validated, demonstrating acceptable to high reliability across domains [8].

2.2 Participants

A total of 289 participants were involved, comprising 231 females (79.9%) and 58 males (20.1%). Most were aged 18–25 (80.62%), with smaller proportions aged 26–35 (14.19%) and 36 or older (5.19%). In terms of education, the majority held a Diploma or Bachelor’s degree (77.86%), followed by high school (14.52%) and Master’s (7.61%). Participants were primarily college students (60.21%), alongside employees (20.76%) and fresh graduates (19.03%), reflecting a sample largely in transition from education to the workforce. Data were collected in 2025.

2.3 Data Analysis

To examine the multivariate relationships between personality traits measured by SJTs (via CDM) and the IPIP-NEO-120, Robust Canonical Correlation Analysis (RCCA) with a Spearman rank correlation matrix was used. RCCA identifies maximally correlated canonical variates while addressing violations of linearity and normality, and reducing the impact of multivariate outliers common in psychometric data. By relying on rank-based associations, RCCA preserves the ordinal structure of responses and yields more stable estimates under non-ideal conditions [9]. This analysis was conducted using the `cancor()` function in R (RStudio 4.3.1)

3 Results

3.1 Descriptive Analysis

Descriptive analyses were conducted to provide an overview of the data distribution for each study variable. The reported statistics included the mean, standard deviation, standard error of the mean, variance, minimum score, and maximum score for each trait measured by both the SJT and the IPIP-NEO-120.

Table 1. Descriptive Analysis

	Mean	Std. Error of Mean	Std. Deviation	Variance
N_SJT	9.346	0.180	3.068	9.415
E_SJT	10.678	0.162	2.754	7.587
O_SJT	11.156	0.154	2.626	6.896
A_SJT	10.692	0.153	2.607	6.797
C_SJT	11.913	0.159	2.697	7.274
N_NEO	50.280	0.483	8.206	67.341
E_NEO	75.104	0.339	5.765	33.239
O_NEO	78.246	0.275	4.670	21.811
A_NEO	81.747	0.244	4.149	17.217
C_NEO	80.201	0.555	9.434	88.994

The analysis result on Table 1 showed that SJT-based scores ranged from a mean of 9.35 (Neuroticism) to 11.91 (Conscientiousness), with relatively small variances. For the IPIP-NEO-120, mean scores were higher, ranging from 50.28 (Neuroticism) to 81.75 (Agreeableness) and 80.20 (Conscientiousness), with larger variability, particularly in Conscientiousness (variance = 88.99).

3.2 Canonical Correlation Analysis

Canonical Correlation Analysis (CCA) was conducted to evaluate the multivariate relationships between two sets of personality variables, namely the CDM-based SJT results and the IPIP-NEO-120 scores.

3.2.1 Assumption Testing for CCA

The first assumption tested was normality. The results indicated that the data were not normally distributed, with a p-value < .001 (statistic Henze-Zirkler = 3.237), which is below the conventional threshold of .05.

The next assumption test was conducted by examining the linearity of the data.

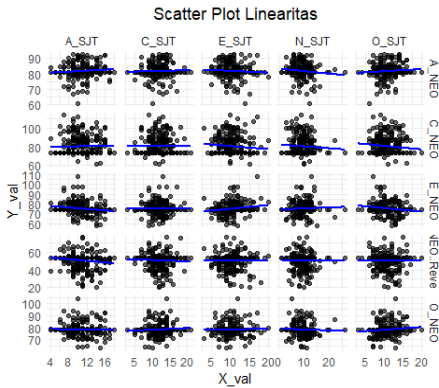


Fig. 1. Scatter Plot Linearity

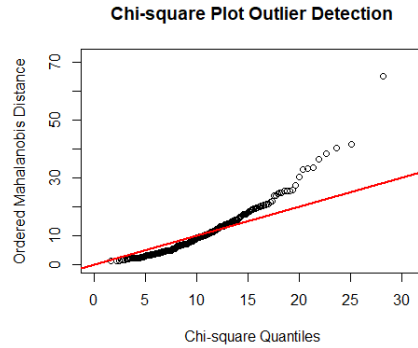


Fig. 2. Scatter Plot Outlier Detection

Figure 1 showed that most variable pairs lacked clear slopes or elliptical patterns, indicating poor support for the linearity assumption in CCA. Multicollinearity was not an issue, with VIF values ranging from 1.020 to 1.474. However, Mahalanobis distance analysis (Figure 2) revealed substantial deviations, suggesting the presence of multivariate outliers.

3.2.2 Robust CCA

The overall results of assumption testing indicated that most classical CCA assumptions were not met. To address these challenges without compromising the uniqueness of the data, the study proceeded with Robust Canonical Correlation Analysis (RCCA).

Table 2. Robust CCA Result

Root	P-value	Coefficient of RCCA	Note
1	0,000	0,373	Medium
2	0,000	0,283	Small to medium
3	0,021	0,157	Small
4	0,015	0,108	Small
5	0,972	0,001	Negligible

Table 2 showed that Robust CCA produced five canonical roots, with Root 1 showing the strongest and statistically significant correlation ($r = 0.373$, $p < .001$). According to [10] guidelines for interpreting correlation strength (small = .10, medium = .30, large = .50), this value represents a moderate association, indicating acceptable convergent validity. Therefore, subsequent analyses focused on Root 1.

Table 3. Root 1 Robust CCA Coefficient

Variable	Root1	
	Coef	Cross loading
X Set (SJT-FFM)		
N SJT	0,422	0,157
E SJT	0,555	0,207
O SJT	-0,525	-0,196
A SJT	-0,722	-0,269
C SJT	-0,183	-0,068
Y set (NEO IPIP-120)		
N NEO Reverse	0,352	0,131

E NEO	0,72	0,269
O NEO	-0,192	-0,072
A NEO	-0,42	-0,157
C NEO	-0,138	-0,052

The root 1 Robust CCA coefficient result on Table 3 showed that in the SJT set (U1), the strongest contribution came from a large negative coefficient of Agreeableness (-0.722), followed by positive Extraversion (0.555) and negative Openness (-0.525). Neuroticism (reverse-scored as emotional stability) showed a moderate positive effect (0.422), while Conscientiousness contributed minimally (-0.183). In the NEO set (V1), Extraversion had the largest positive effect (0.720), with emotional stability also contributing positively (0.352). Agreeableness showed a moderate negative effect (-0.420), while Openness (-0.192) and Conscientiousness (-0.138) were weak. Cross-loadings highlighted Extraversion and emotional stability as positively linked across sets, whereas Agreeableness, Openness, and Conscientiousness showed weaker negative associations.

3.2.3 Evaluation of the Robust CCA Model

To evaluate the effectiveness and stability of cross-set associations, this study applied nonparametric bootstrapping of the canonical correlations for two approaches: classical CCA (Pearson) and RCCA. Analyses were conducted under two data conditions: raw data with outliers included (n = 289) and data with outliers removed (n = 265).

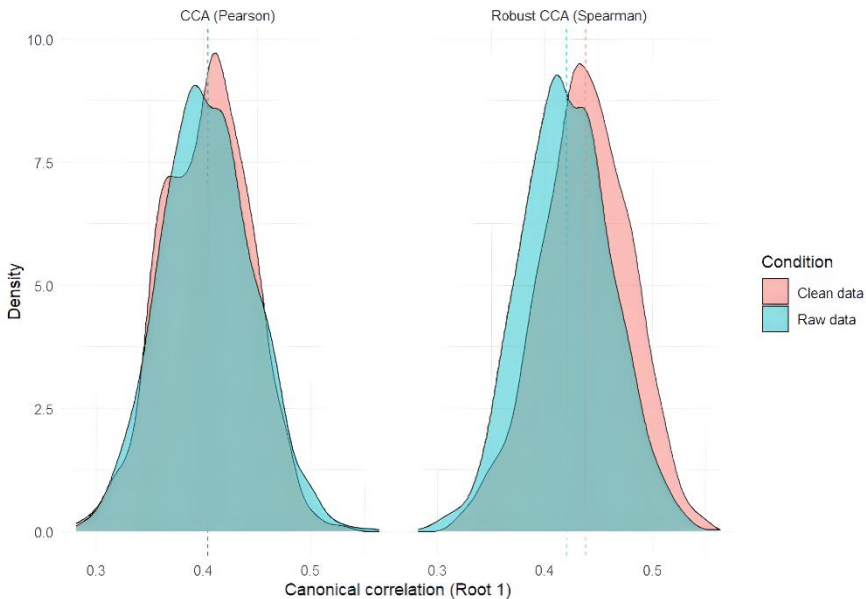


Fig. 3. Nonparametric Bootstrap Robust CCA

Bootstrapping (1,000 resamples) on Figure 3 showed that under CCA–Pearson, the mean canonical correlation was nearly identical between raw (M = 0.404) and cleaned data (M = 0.403), with overlapping distributions. In contrast, RCCA–Spearman showed a slight increase after outlier removal (from M = 0.420 to M = 0.438), indicating that excluding extreme scores made the method more sensitive to sample characteristics. Robustness and

MSE analyses further confirmed the effectiveness of RCCA compared to Pearson CCA across both conditions (with and without outliers).

Table 4. MSE

Method	Condition	Coefficient	MSE rho	MSE alpha	MSE beta
CCA Pearson	Raw	0.3577	0,007	0,644	0,516
CCA Pearson	Clean	0.3571	0,005	0,608	0,560
CCA Spearman	Raw	0.3726	0,004	0,592	0,559
CCA Spearman	Clean	0.3936	0,006	0,583	0,521

From MSE result on Table 4, Pearson CCA showed almost no change after outlier removal ($\Delta r = 0.17\%$), while Spearman CCA showed a moderate increase ($\Delta r = 5.64\%$), highlighting its sensitivity to rank-order changes. MSE analysis revealed that Spearman consistently produced lower errors for canonical coefficients ($MSE\rho$) and weights ($MSE\alpha$, $MSE\beta$) than Pearson, in both raw and cleaned data. This indicates that RCCA–Spearman provides more accurate and robust estimates than Pearson CCA, especially when handling outliers.

4 Discussion

The findings of this study highlight both the methodological and substantive implications of applying Robust Canonical Correlation Analysis (RCCA) to examine the alignment between CDM-based Situational Judgment Tests (SJT–FFM) and self-report inventories. Classical Canonical Correlation Analysis (CCA) assumes multivariate normality, linearity, and absence of outliers, conditions rarely met in psychometric data. In this study, only multicollinearity and sample size assumptions were satisfied, while violations of normality and outlier presence were evident. Such violations can bias Pearson-based estimates, making RCCA with Spearman correlations a more suitable choice, as it minimizes sensitivity to distributional anomalies while preserving the goal of maximizing shared variance between sets [9], [11].

The robust analysis revealed five canonical roots, with the first two demonstrating meaningful associations. Root 1 ($r = 0.373$, $p < .001$) represented the strongest canonical dimension, followed by Root 2 ($r = 0.283$, $p < .001$). Overall, the primary canonical correlation of 0.348 indicated a moderate effect. In personality validation contexts, moderate correlations are expected and substantively valuable, particularly when comparing scenario-based SJTs with Likert-type self-reports, which differ in response format and cognitive demands [4], [12]. Rather than redundancy, such moderate convergence suggests that SJTs reflect the same broad constructs as self-reports while capturing contextualized behavioral expressions [13], [14].

Trait-level contributions illustrated the multidimensionality of SJT responses. Extraversion contributed positively while Agreeableness contributed negatively within the first canonical variate, reflecting suppression effects common when multiple traits are embedded in a single response option [11]. For example, situational cues appeared to activate assertiveness and stability while attenuating openness and agreeableness, highlighting how context modulates trait expression. This pattern aligns with prior work on situational construal, which demonstrates that SJTs elicit differential trait activation depending on scenario demands [15]. By contrast, Conscientiousness showed minimal contribution across sets, reinforcing evidence that SJT performance may rely more on domain-specific behavioral knowledge than on self-reported trait dispositions [13].

Methodologically, RCCA demonstrated value by providing stable canonical estimates under assumption violations. Bootstrap comparisons showed that Spearman-based RCCA

adjusted more meaningfully to outlier removal than Pearson-based CCA, reflecting its sensitivity to monotonic associations while preserving rank-order differences [9]. In psychometric applications, such sensitivity is not a weakness but an asset, as it captures meaningful individual variation often masked by linear-normal methods. Taken together, the results suggest that RCCA is a robust and appropriate analytic framework for linking behavioral and self-report measures, ensuring that convergence estimates reflect substantive relationships rather than artifacts of assumption violations.

5 Conclusion

The application of Spearman-based robust canonical correlation analysis provided more appropriate insights for psychometric data than classical approaches. The analysis showed a moderate canonical relationship, supporting convergent validity between the DCM-based SJT and the NEO-PI. This suggests that, despite originating from different frameworks, the two instruments share meaningful construct-level associations.

While bootstrap results indicated some instability, this reflects not only methodological constraints but also the complexity and heterogeneity of psychological data. Thus, convergence should be understood as both a function of correlation magnitude and the capacity of robust methods to capture non-linear patterns.

Nonetheless, this study has some limitations: the small sample size reduces generalizability, unstable estimates call for replication with larger and more diverse samples, and differences in instrument format may have affected associations. Besides that, most of the participants were university students or young adults, and the majority were women. These sample characteristics may have influenced the expression of personality traits and the response tendencies in the SJT, which should be taken into account when interpreting and generalizing the findings. Nevertheless, the findings demonstrate that robust CCA offers a valid alternative for examining convergent validity when classical assumptions are unmet, highlighting its potential to address data limitations and capture the complexity of psychological constructs.

References

1. L. R. Goldberg, The development of markers for the Big-Five factor structure, *Psychol Assess*, **4**, 1, 26–42 (1992). <http://dx.doi.org/10.1037/1040-3590.4.1.26>
2. J. A. Johnson, Measuring thirty facets of the Five Factor Model with a 120-item public domain inventory: Development of the IPIP-NEO-120, *J Res Pers*, **51**, 78–89 (2014). <http://dx.doi.org/10.1016/j.jrp.2014.05.003>
3. J. McDonald, Measuring personality constructs: The advantages and disadvantages of self-reports, informant reports and behavioural assessments, *Enquire*, **1** (2008)
4. M. McDaniel, N. Hartman, D. Whetzel, and W. Grubb III, Situational judgment tests, response instructions, and validity: A meta-analysis, *Pers Psychol*, **60**, 63–91 (2007). <http://dx.doi.org/10.1111/j.1744-6570.2007.00065.x>
5. F. Lievens, P. Schäpers, and C. N. Herde, Situational judgment tests: From low-fidelity simulations to alternative measures of personality and the person-situation interplay, *Measuring and Modeling Persons and Situations*, Elsevier, 285–311 (2021). <http://dx.doi.org/10.1016/B978-0-12-819200-9.00017-X>

6. M. Sorrel, J. Díaz, F. Abad, J. Torre, D. Aguado, and F. Lievens, Validity and Reliability of Situational Judgement Test Scores: A New Approach Based on Cognitive Diagnosis Models, *Organ Res Methods*, **19** (2016). <http://dx.doi.org/10.1177/1094428116630065>
7. B. Thompson, *Canonical Correlation Analysis*, (SAGE Publications, Inc., 1984)
8. M. Tarigan and Fadillah, The Validity and Reliability of The Indonesian IPIP-NEO-120, *TPM*, **31**, 1, 121–136 (2024)
9. A. Alfons, C. Croux, and P. Filzmoser, Robust Maximum Association Between Data Sets: The R Package *ccaPP*, *Austrian Journal of Statistics*, **45**, 1, 71–79 (2016). <http://dx.doi.org/10.17713/ajs.v45i1.90>
10. J. Cohen, *Statistical power analysis for the behavioural sciences*. (Hillside, 1988).
11. A. Sherry and R. Henson, Conducting and Interpreting Canonical Correlation Analysis in Personality Research: A User-Friendly Primer, *J Pers Assess*, **84**, 37–48 (2005). http://dx.doi.org/10.1207/s15327752jpa8401_09
12. F. Lievens and P. R. Sackett, The validity of interpersonal skills assessment via situational judgment tests for predicting academic success and job performance., *Journal of Applied Psychology*, **97**, 2, 460–468 (2012). <http://dx.doi.org/10.1037/a0025741>
13. G. Olaru, J. Burrus, C. Maccann, F. Zaromb, O. Wilhelm, and R. Roberts, Situational Judgment Tests as a method for measuring personality: Development and validity evidence for a test of Dependability, *PLoS One*, **14** (2019). <http://dx.doi.org/10.1371/journal.pone.0211884>
14. V. S. Schröder, A. L. Heimann, P. V. Ingold, and M. Kleinmann, Enhancing Personality Assessment in the Selection Context: A Study Protocol on Alternative Measures and an Extended Bandwidth of Criteria, *Front Psychol*, **12** (2021). <http://dx.doi.org/10.3389/fpsyg.2021.643690>
15. J. Freudenstein, P. Schäpers, L. Roemer, P. Mussel, and S. Krumm, Is it all in the eye of the beholder? The importance of situation construal for situational judgment test performance, *Pers Psychol*, **73**, 4, 669–700 (2020). <http://dx.doi.org/10.1111/peps.12385>