Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Stratifying cognitive performance severity in higher education using a two level clustering approach

Nuril Kusumawardani Soeprapto Putri
Institut Teknologi Sepuluh Nopember
Umi Laili Yuhana
Yuhana@its.ac.id

Institut Teknologi Sepuluh Nopember
Daniel Oranova Siaahaan
Institut Teknologi Sepuluh Nopember
Istiani Istiani
Bina Nusantara University

Research Article

Keywords: Spectral clustering, Graph-based learning, Hierarchical clustering, Unsupervised modelling, Cluster validation

Posted Date: April 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9110337/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Stratifying cognitive performance severity in higher education using a two level clustering approach

Nuril Kusumawardani Soeprapto Putri[1,2], Umi Laili Yuhana[1*], Daniel Oranova Siaahaan[1], Istiani[3]

[1]Department of Informatics Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
[2]Information Systems Department, School of Information Systems, Bina Nusantara University, Jakarta, Indonesia 11480
[3]Psychology Department, Faculty of Humanities, Bina Nusantara University, Jakarta, Indonesia 11480

*Corresponding author: Umi Laili Yuhana (Yuhana@its.ac.id).

Abstract

Cognitive ability distributions are multidimensional and may exhibit non-linear structure, which can limit centroid-based clustering. This study proposes a two-level spectral clustering framework for multivariate cognitive severity stratification using psychometric latent trait (theta) estimates of logical reasoning, verbal ability, and abstraction.. The complete sample (N = 36,830) was organised according to three latent ability estimates derived from logical reasoning, verbal ability, and abstraction. In the first stage, graph Laplacian–based spectral clustering identified three severity strata without assuming spherical cluster structure. The lowest-performing cluster (n = 13,252) was then refined in a second stage to obtain more detailed subgroups.

Multivariate cluster separability was evaluated using PERMANOVA, showing significant differentiation among clusters (pseudo-F = 4687.0, R^2 = 0.2613, p = 0.01). Pairwise post hoc tests confirmed significant separation across all cluster pairs, with the strongest divergence between the lowest and highest strata (pseudo-F = 6156.5, R^2 = 0.3978, p = 0.01). NMDS with fitted vectors further indicated distinct contributions of abstraction, verbal ability, and logical reasoning to cluster separation. Rather than providing diagnostic labels, this study presents an interpretable and scalable unsupervised framework for identifying structured severity patterns in complex cognitive data.

Keywords

Spectral clustering, Graph-based learning, Hierarchical clustering, Unsupervised modelling, Cluster validation

1. Introduction

Cognitive abilities such as logical reasoning, verbal ability, and abstraction are inherently multidimensional, yet they are frequently modelled as separate domains or reduced to a single composite score in educational analytics. While such simplifications facilitate large-scale assessment, they risk obscuring meaningful patterns of overall cognitive severity that emerge from interactions across domains. In higher education, where student cohorts are large and cognitively heterogeneous, this limitation constrains the ability of data-driven systems to distinguish between broadly graded severity profiles that may inform scalable academic support and adaptive learning strategies [1–3].

Unsupervised clustering has therefore become a common approach for discovering latent cognitive profiles in both educational and clinical research. Prior studies have applied clustering techniques to multi-domain cognitive indicators in school-aged learners and university students, revealing groups that differ in overall level and domain-specific strengths or weaknesses [3–5]. Similarly, neuropsychological and clinical investigations consistently report severity-graded cognitive subtypes. It often described as near-normal, intermediate, and low-functioning profiles, that are associated with differential functional outcomes and intervention strategies [2,6,7]. Despite these advances, many studies rely on relatively small samples, high-

dimensional test batteries, or descriptive cluster interpretation without formal multivariate validation of cluster separability.

Methodological concerns surrounding unsupervised clustering in cognitive and psychological data have become increasingly prominent. Simulation studies and systematic reviews show that clustering algorithms may detect false groupings when applied to correlated indicators, modest effect sizes, or insufficient sample sizes conditions frequently observed in cognitive datasets [8]. Even distinguishing between one and two true latent clusters can be unreliable without appropriate dimensionality reduction, adequate sample scale, and rigorous validation procedures. Comparative analyses in educational data mining further demonstrate that clustering outcomes are sensitive to algorithm choice and data representation, emphasizing the need for principled methodological design rather than purely exploratory application [4,9]. In parallel, recent contributions emphasize structured and validated approaches to clustering in high-dimensional settings. A comprehensive roadmap on spectral clustering highlights its capacity to capture latent structure in complex spaces and integrate multi-view representations [10], while applications in educational big data demonstrate how multidimensional feature spaces can reveal heterogeneous learner profiles when supported by appropriate validation [11]. Together, these findings support the importance of statistically grounded, interpretable, and scalable clustering frameworks for multivariate cognitive data.

Despite existing studies, there is still a need for approaches that examine cognition as a combined set of abilities rather than separate domains, use psychometric latent traits for clearer measurement, and focus on identifying levels of severity instead of assigning diagnostic labels. To address this gap, the present study proposes a two-level unsupervised clustering framework that embeds logical reasoning, verbal ability, and abstraction into a unified three-dimensional theta and applies Spectral Clustering in a large university sample. The first-level clustering identifies broad multivariate severity groups, while a second-level refinement focuses on re-clustering the most heterogeneous group to reveal finer-grained severity strata. Multivariate separation is formally evaluated using permutational multivariate analysis of variance (PERMANOVA), and non-metric multidimensional scaling is employed to visualize and interpret the directional contributions of each latent trait. Rather than advancing diagnostic claims, this study introduces a statistically grounded exploratory pipeline for multivariate cognitive severity stratification, providing methodological foundations for adaptive educational analytics and AI-driven decision support in higher education.

Accordingly, this study is guided by the following research questions:

RQ1: Do the combined latent cognitive traits form statistically distinct clusters when analysed using spectral clustering?
RQ2: Can a two-level spectral clustering approach reveal more detailed and coherent subgroups within heterogeneous regions of the data?

RQ3: Are the identified clusters stable and predictable enough to be reliably classified using supervised machine learning?

2. Related Works

2.1. Multivariate Cognitive Profiling and Latent Structures

Cognitive diagnosis and psychometric modelling increasingly represent cognition as a multivariate latent construct rather than isolated domains. Higher order latent trait models specify a general aptitude that governs multiple specific attributes, yielding parsimonious yet informative representations for cognitive diagnosis [12]. Recent general response cognitive diagnostic models further extend this idea to rich response types and higher order continuous traits, emphasizing exploratory estimation of item–attribute structures when the Q matrix is unknown [13]. In mental health, cognitive diagnosis models have been proposed as an alternative to scale scores to capture heterogeneous symptom profiles across multiple domains using item level data [14]. Joint latent process and latent class models similarly treat cognition as a latent trajectory linked to multiple longitudinal markers and competing clinical outcomes [15]. These approaches align with the present study's use of psychometric theta scores to define a compact, multivariate cognitive space for exploratory profiling.

2.2. Unsupervised Clustering in Educational and Cognitive Contexts

Unsupervised clustering is widely used to discover learner profiles from behavioral and performance data. Ensemble and two-stage frameworks combining DBSCAN with k-means or hierarchical clustering have been applied to campus behavioural logs and ubiquitous learning data to separate mainstream from anomalous patterns and identify at-risk students [16,17]. Other work clusters temporal traces of online engagement using graph-based methods and dynamic time warping to reveal quasi-hierarchical learner groups with distinct performance risks [18]. Comparative educational data mining studies benchmark k-means, DBSCAN, BIRCH, EM, and GMM for grouping students on multidimensional performance indicators, highlighting sensitivity to algorithm choice and the importance of internal validation indices such as silhouette scores [19,20]. Recent clustering-plus-recommendation frameworks explicitly connect unsupervised profiles to adaptive pedagogical interventions [21]. These studies motivate severity-graded clustering in large cohorts but typically operate on behavioural rather than latent cognitive features and seldom use formal multivariate tests of cluster separability.

2.3. Spectral and Multi View Clustering for Complex Cognitive Data

Spectral clustering has emerged as a powerful approach for complex, high-dimensional data, including cognitive and neurophysiological signals. Surveys based on graph theory review the construction of similarity graphs, Laplacian variants, and eigenvector selection, and emphasize both the strong theoretical foundation and scalability challenges for large datasets [22]. Methodological advances propose robust spectral variants that learn affinity matrices in low-dimensional spaces to reduce outlier influence and automatically infer cluster number [23]. Multi-view spectral clustering further integrates complementary representations into unified graphs or nonnegative embeddings, avoiding sub-optimal two-step pipelines and reducing post-processing uncertainty [24–26]. Within cognitive assessment, spectral clustering has been directly applied to examinee response data to classify attribute-homogeneous groups,

showing competitive or superior performance to k-means and some diagnostic models in simulation and real datasets [9].

2.4. Cognitive Severity Stratification and Latent Profiles

Beyond education, latent profile and transition analyses have been used to derive severity-graded cognitive or functional subtypes, often yielding three-level structures (normal/mild/severe) and modelling transitions between profiles over time [27]. Similar mixture and growth-mixture approaches identify distinct severity trajectories in psychopathology and grief, linking them to cognitive predictors and outcomes [28]. However, these works typically impose parametric mixture structures, focus on clinical populations, or operate in high-dimensional test spaces without emphasizing scalable graph-based clustering or explicit multivariate validation of profile separation.

Collectively, prior research supports multivariate latent representations of cognition, the utility of unsupervised clustering for educational decision-making, and the methodological advantages of spectral and multi-view clustering. The current study extends this literature by combining psychometric theta embeddings with a two-level spectral clustering pipeline and formal multivariate validation (PERMANOVA) to derive scalable, severity-graded cognitive strata in a large higher-education cohort.

3. Methodology

3.1. Overall Analytical Workflow

This study implements an end-to-end analytical pipeline that integrates psychometric latent trait estimation, unsupervised clustering, predictive validation, and multivariate statistical testing to derive interpretable cognitive severity strata. As summarized in Figure 1, the workflow proceeds in five stages. **(1)** Item-level responses are converted into domain-specific latent ability estimates using Item Response Theory (IRT), producing theta scores for logical reasoning, abstraction, and verbal ability. **(2)** These theta values define a compact three-dimensional cognitive feature space and are rescaled using Min–Max normalization to ensure comparable contributions across domains. **(3)** Unsupervised clustering is performed using K-Means, Agglomerative Clustering, HDBSCAN, and Spectral Clustering; Spectral Clustering is selected as the primary method due to its ability to capture non-linear structure in graph-based similarity space. The proposed framework adopts a two-level strategy. In Stage 1, spectral clustering is applied to the full dataset (N = 36,830) to identify three severity strata. In Stage 2, reclustering is applied exclusively to the lowest-performing cluster identified in Stage 1 (n = 13,252) to improve severity resolution within this subgroup. **(4)** Cluster assignments are then treated as pseudo-labels to train supervised classifiers (Random Forest, XGBoost, CatBoost), assessing whether severity level can be predicted reliably for downstream deployment. **(5)** Finally, PERMANOVA is used to statistically validate whether clusters are separable in multivariate cognitive space based on distance matrices and permutation testing. This integrated design supports scalable and statistically grounded severity stratification for large higher-education cohorts.

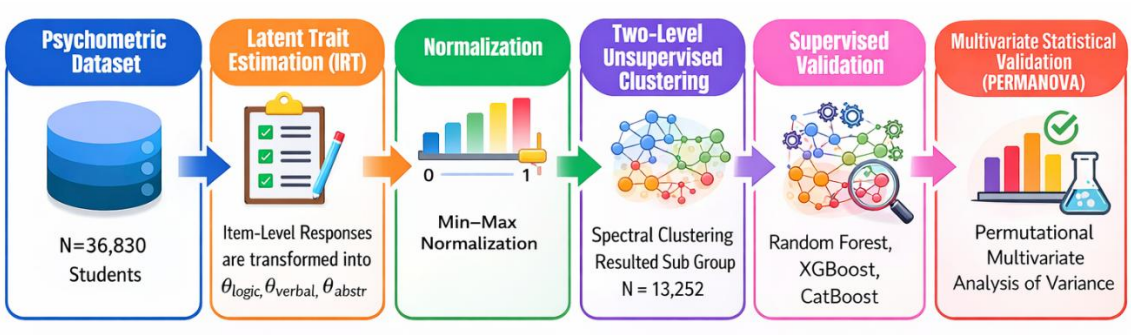


Figure 1: Workflow for Cognitive Severity Modelling Using IRT and Machine Learning

3.2. Psychometric Dataset and IRT Processing

The empirical analysis was based on a large-scale institutional psychometric assessment designed to measure three core cognitive domains: logical reasoning, abstraction, and verbal ability. These domains are widely recognized as central components of higher-order cognitive functioning and academic aptitude in educational and psychological research.

The assessment instrument consisted of 100 multiple-choice items, each with five response options. The distribution of items across domains was as follows: 30 items measured logical reasoning, 35 assessed abstraction, and 35 evaluated verbal ability. The test was administered digitally as part of a university-wide cognitive screening initiative.

The dataset comprised anonymized secondary data from 36,830 undergraduate students. Only item-level response data were available in coded form, with no personally identifiable information. Because the study relied exclusively on de-identified archival data, no additional ethical approval was required. Responses were encoded dichotomously (1 = correct, 0 = incorrect) prior to psychometric modelling.

To obtain robust estimates of students' latent cognitive abilities, the binary response data were analyzed using the three-parameter logistic (3PL) Item Response Theory model. The 3PL model simultaneously estimates:

- Item difficulty (b) – the location of the item on the ability scale
- Item discrimination (a) – the strength of the item in differentiating ability levels
- Guessing parameter (c) – the probability of a correct response due to guessing

Separate IRT calibrations were conducted for each cognitive domain (logical reasoning, abstraction, verbal ability) to preserve domain-specific latent structures.

Items were evaluated based on their parameter estimates. Items exhibiting:

- Very low discrimination,
- Extreme difficulty values,
- Or inflated guessing parameters,

were removed from theta estimation to improve measurement precision and reduce noise in the latent trait space.

Importantly, no student observations were excluded during this process. All 36,830 student records were retained. The refinement process targeted only item-level quality

control, ensuring that latent trait estimation reflected reliable measurement properties rather than test artifact.

3.3. Latent Trait Extraction Processing

For each student, domain-specific theta scores were estimated from the calibrated 3PL models. These theta values represent continuous latent ability estimates for:

- Logical reasoning (θ_logic)
- Abstraction (θ_abstraction)
- Verbal ability (θ_verbal)

The resulting three-dimensional theta representation forms a compact and interpretable cognitive feature space. This latent embedding reduces high-dimensional item-level data into theoretically grounded psychometric traits, providing a stable foundation for subsequent multivariate clustering analysis.

3.4. Preprocessing: Normalization

After obtaining theta scores from the IRT analysis for each of the three cognitive domains (logic, verbal, and abstraction), preprocessing was performed to prepare the data for clustering. MinMaxScaler was employed to scale all values into the [0, 1] range. This step was critical to ensure that each domain contributed equally to the similarity computation and to avoid bias in clustering due to differing value ranges. Normalization also helps maintain consistency when the model is later applied to classify smaller, primary datasets. These preprocessing steps improve the robustness and generalizability of the clustering and classification phases that follow.

3.5. Clustering Methods: KMeans, Agglomerative, HDBSCAN, Spectral Clustering

This study employed four unsupervised clustering techniques to explore latent cognitive profiles among students: K-Means, Agglomerative Clustering, HDBSCAN, and Spectral Clustering. Each method was applied to the preprocessed theta scores (logic, verbal, and abstraction) to identify groups of students with similar cognitive characteristics.

a. K-Means clustering partitions the data into k clusters by minimizing the within-cluster sum of squares. While widely used due to its efficiency, K-Means assumes spherical clusters and equal variance, which may not align with the structure of cognitive data.
b. Agglomerative Clustering, a hierarchical approach, starts by treating each data point as a separate cluster and successively merges the closest pairs. This method does not require specifying the number of clusters a priori and can capture nested group structures. A complete-linkage criterion was used to ensure compact cluster formation.
c. HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise) is an extension of DBSCAN that introduces a hierarchical approach to density-based clustering. It constructs a cluster hierarchy based on varying density levels and then extracts the most stable clusters according to cluster persistence. Unlike DBSCAN, which requires a fixed density threshold (epsilon), HDBSCAN can automatically determine cluster boundaries and handle datasets with varying densities more effectively.

d. Spectral Clustering was employed as the primary method for the final model due to its ability to handle complex, non-convex cluster structures. It computes the graph Laplacian, and performing dimensionality reduction through eigen decomposition. Clustering is then applied in the reduced spectral space. This approach is particularly effective in identifying subtle group differences in multidimensional psychological attributes [29,30].

3.6. Evaluation Metrics

To evaluate the clustering performance, three internal validation metrics were employed: Silhouette Score, Davies-Bouldin Index (DBI), and Calinski-Harabasz Index (CHI). These metrics assess cluster compactness and separation without relying on external labels.

a. The Silhouette Score quantifies how well each data point fits within its assigned cluster compared to other clusters. It ranges from -1 to 1, where higher values indicate better-defined clusters. A score near 1 reflects strong cohesion within clusters and clear separation between clusters.
b. The Davies-Bouldin Index evaluates the average similarity between each cluster and its most similar counterpart. Lower DBI values signify superior clustering quality, indicating that clusters are compact and well-separated from one another.
c. The Calinski-Harabasz Index, also known as the Variance Ratio Criterion, measures the ratio of between-cluster dispersion to within-cluster dispersion. Higher CHI values imply more distinct and well-formed clusters. This metric is particularly effective for assessing the compactness and separation of clusters in high-dimensional data.

All three metrics were computed for each clustering algorithm applied in this study. These metrics informed the selection of the optimal clustering model to represent cognitive profiles.

3.7. Classification Models

To build a predictive model for identifying cognitive difficulty levels, three ensemble-based algorithms were used: Random Forest, XGBoost, and CatBoost. These models were chosen for their effectiveness in handling structured educational data and their ability to capture nonlinear relationships among cognitive variables. The cluster labels derived from Spectral Clustering, representing severe, moderate, and mild cognitive difficulty, were treated as target variables in a supervised classification task. Each model received the normalized theta scores for logical reasoning, abstraction, and verbal ability as input features. The objective was to predict students' cognitive cluster level, enabling early identification of those who might require academic support.

- Random Forest is an ensemble learning method that combines multiple decision trees trained on random subsets of the data. It is known for its stability and resistance to overfitting, making it suitable for datasets with complex relationships.
- XGBoost (Extreme Gradient Boosting) builds trees sequentially, where each new tree corrects the errors made by the previous ones. It uses regularization techniques to improve accuracy and prevent overfitting, resulting in strong predictive performance and computational efficiency.
- CatBoost (Categorical Boosting) is a gradient boosting algorithm that introduces ordered boosting to prevent target leakage and reduce bias. It also uses a symmetric tree structure that ensures consistent generalization. CatBoost

performed best among the three models, achieving the highest accuracy and balanced precision, recall, and F1-scores.

3.8. Multivariate Statistical Validation: PERMANOVA

PERMANOVA was conducted on the second-stage clustering results within the lowest-performing subgroup (n = 13,252) to assess multivariate separability among the refined severity strata. While, MANOVA is commonly used to test group differences across multiple dependent variables; however, valid MANOVA inference depends on assumptions that are difficult to satisfy in large-scale psychometric data and clustering-derived groups, including multivariate normality within each group and homogeneity of covariance matrices across groups. In this dataset, the joint distribution of theta scores across the three cognitive domains showed deviations from multivariate normality and unequal dispersion patterns across clusters, which can inflate Type I error or destabilize MANOVA results when group structure is not derived from an experimental design.

PERMANOVA provides a robust alternative because it operates directly on a dissimilarity matrix and uses permutation testing to evaluate whether between-group variation exceeds within-group variation without requiring multivariate normality [31,32]. Using Euclidean distances computed from normalized theta scores, PERMANOVA tests separation through a pseudo-F statistic:

$$F = \frac{SSA}{SSW} \cdot \frac{N - g}{g - 1} \tag{1}$$

where SSA and SSW denote among-group and within-group sums of squares, N is the total number of observations, and $g$is the number of clusters. The proportion of explained variance is:

$$R^2 = \frac{SSA}{SSA + SSW} \tag{2}$$

Statistical significance is obtained by permuting cluster labels to generate a null distribution of pseudo-F values:

$$p = \frac{\#(F_{perm} \geq F_{observed}) + 1}{\text{number of permutations} + 1} \tag{3}$$

In this study, PERMANOVA was implemented using the adonis() function in the vegan package in R. This approach strengthens inference by confirming that cluster

separability reflects statistically supported multivariate structure rather than an artifact of parametric assumption violations.

4. Results

4.1. Descriptive Statistics of Theta Values

The descriptive statistics for the theta scores are summarized in Table 1. The logical reasoning domain had a mean of approximately 0.000003 with a standard deviation of 0.898, ranging from –1.72 to 1.34. The abstraction domain yielded a slightly lower mean (–0.000004) with a standard deviation of 0.854, spanning from –2.54 to 3.28. Meanwhile, verbal ability had a mean of 0.000072 and a standard deviation of 0.859, with theta values ranging from –3.16 to 2.31.

The interquartile ranges across the three domains indicate moderate variability in student performance. For example, 50% of the students scored between –0.77 and 0.76 in logical reasoning, between –0.58 and 0.56 in abstraction, and between –0.56 and 0.54 in verbal ability.

Table 1. Descriptive Statistics of Theta Values

Cognitive Domain	N	Mean	Std. Deviation	Min	25th Percentile	Median	75th Percentile	Max
Logical Reasoning	36830	0.000003	0.898	-1.7218	-0.7663	0.0620	0.7588	1.3431
Abstraction	36830	-0.000004	0.854	-2.5414	-0.5822	-0.0159	0.5621	3.2843
Verbal Ability	36830	0.000072	0.859	-3.1616	-0.5606	-0.0206	0.5373	2.3107

4.2. Data Transformation: Normalization with MinMaxScaler

The theta scores were normalized using the MinMaxScaler method to ensure that each cognitive feature (logic, abstraction, and verbal) ranged between 0 and 1. This normalization step was essential to eliminate scale-related bias and enable fair distance computation during clustering. This preprocessing step ensured that the clustering algorithms were not influenced by scale disparities, leading to more reliable and comparable clustering outcomes in subsequent stages. Table 2 displays an excerpt of the normalized dataset, demonstrating the uniformity across the three cognitive dimensions.

Table 2. Sample of Normalized Dataset

No	Logical Reasoning	Abstraction	Verbal Ability
1	0.171287	0.377134	0.440102
2	0.662258	0.314519	0.584970
3	0.283503	0.387282	0.603470
4	0.376229	0.362971	0.815820
5	0.137359	0.311592	0.524085

4.3. Clustering Performance Comparisons

This section compares the clustering performance of four unsupervised methods: K-Means, Agglomerative Clustering (Complete Linkage), HDBSCAN, and Spectral Clustering. The evaluation focuses on the separability, cohesion, and visual structure of clusters formed using cognitive ability data (logic, abstraction, verbal) after MinMaxScaler normalization.

To identify the most effective clustering technique for modeling cognitive difficulty levels, four unsupervised algorithms were tested: K-Means, Agglomerative Clustering (Complete Linkage), HDBSCAN, and Spectral Clustering. These methods were evaluated using the Silhouette Score, a widely used metric that considers both cohesion (how close data points in the same cluster are) and separation (how distinct clusters are from one another). Table 3 summarizes the Silhouette Scores for each algorithm.

Table 3. Clustering Comparison Based on Silhouette Score

Clustering Method	Number of Clusters	Silhouette Score
K-Means	3	0.31
Agglomerative (Complete Linkage)	3	0.27
HDBSCAN	2	0.27
Spectral Clustering	3	0.29

Among the four clustering methods, K-Means achieved the highest Silhouette Score (0.3080), indicating that it produced the most distinct and internally consistent clusters. Spectral Clustering followed closely with a score of 0.29, showing comparable performance and suggesting that both methods were effective in capturing structure within the cognitive data. In contrast, HDBSCAN and Agglomerative Clustering yielded lower scores (0.2742 and 0.2701, respectively), reflecting less cohesive cluster formations.

In addition to numerical evaluation, visual inspection was performed on the clustering results to assess the separability of the groups. Table 5 (Visualization) presents a qualitative comparison of how each algorithm partitioned the dataset in 3D space, using scatter plots of the normalized theta variables. The Spectral Clustering visualization exhibited distinct and well-defined cluster formations, particularly along the abstraction and logic axes. In contrast, K-Means produced clusters that were moderately distinguishable, but with some overlap, especially in the verbal dimension. Agglomerative Clustering resulted in more elongated and less coherent cluster shapes, while HDBSCAN, although capable of identifying noise points and arbitrary shapes, was less effective in this context due to the relative uniformity in density across the theta values.

Taken together, the findings from both the quantitative Silhouette Score and visualization analysis indicate that Spectral Clustering offers the best performance for modelling latent cognitive groupings in this dataset. Its ability to expose subtle

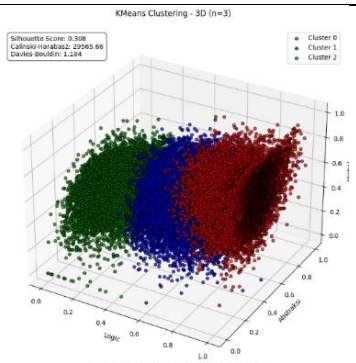
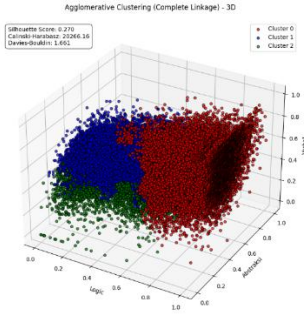
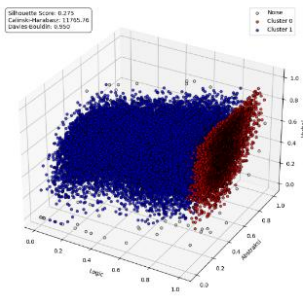
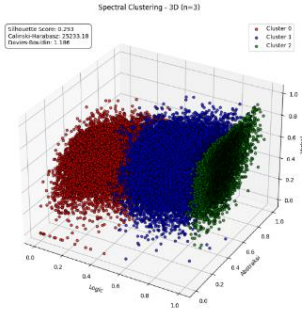
structures in the data makes it particularly useful for further classification tasks aimed at identifying severity levels of cognitive difficulty.

Initial Spectral Clustering Result and Identification of Cluster 0

The first spectral clustering step divided the normalized theta scores (logic, abstraction, and verbal) of 36,830 students into three primary clusters. These clusters represented varying levels of cognitive ability, as shown in the histogram in Table 4.

a. Cluster 0 (red): Students with low cognitive ability
b. Cluster 1 (blue): Students with moderate cognitive ability
c. Cluster 2 (green): Students with high cognitive ability

Table 4. Clustering Comparison Based on Visualization

Cluster Method	Evaluation Metrics	Visualization
K-Means	• *Silhouette Score*: 0.31 • Calinski-Harabaz: 29565.66 • Davies-Bouldin : 1.18	
Agglomerative *Clustering* dengan Complete Linkage	• *Silhouette Score*: 0.27 • Calinski-Harabaz: 20266.16 • Davies-Bouldin: 1.66	
HDBSCAN	• *Silhouette Score*: 0.27 • Calinski-Harabaz: 11765.76 • Davies-Bouldin: 0.95	
Spectral	• *Silhouette Score*: 0,29 • Calinski-Harabaz: 25233,18 • Davies-Bouldin: 1,19	

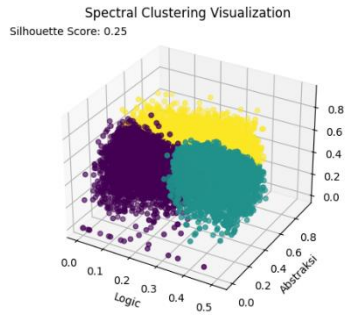
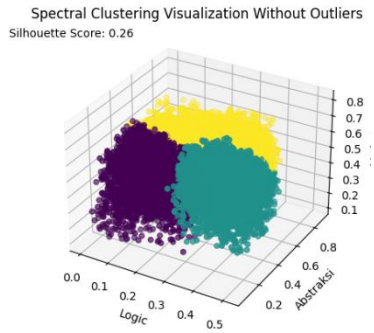
Spectral Clustering on Cluster 0

The second clustering was applied specifically to the data points within Cluster 0 using the same Spectral Clustering settings as in the initial stage (e.g., number of clusters and graph construction parameters). The objective was to reveal internal subgroupings that represent varying severity levels of cognitive challenges among low-performing students.

Following the clustering, some data points appeared far from the cluster centres, suggesting potential outliers. These outliers were subsequently removed to enhance the clarity of the cluster structure. This post-clustering cleaning led to a small increase in the Silhouette Score, from 0.25 to 0.26, suggesting improved separation and compactness.

Table 6 presents a visual comparison of the clustering results before and after outlier removal. In particular, clusters became more distinct, especially between green and yellow groups, while the purple cluster retained denser internal variability.

Table 5. Comparison Cluster 0 With and Without Outlier

Cluster 0 with Outlier	**Cluster 0 with no outlier**
	

To describe the resulting sub-clusters, Table 6 provides descriptive statistics for each sub-cluster. The descriptive statistics of theta scores across the three final clusters—Cluster 0 (Severe), Cluster 1 (Moderate), and Cluster 2 (Mild)—revealed distinct patterns of cognitive performance. Cluster 0, associated with the most severe cognitive challenges, exhibited the lowest average scores across all three domains: logical reasoning (M = 0.138), abstraction (M = 0.346), and verbal ability (M = 0.449).

Interestingly, Cluster 2—interpreted as the mild group—had a higher mean score in abstraction (M = 0.548) and verbal ability (M = 0.579) compared to the other clusters. However, its mean score for logical reasoning (M = 0.246) was lower than Cluster 1, and even notably lower than expected for a group labeled as mild. In contrast, Cluster 1 showed the highest average in logical reasoning (M = 0.382), while having intermediate scores in the other two domains.

This pattern suggests that logical reasoning may not follow the same severity trajectory as abstraction and verbal skills, highlighting a domain-specific variation in cognitive challenges among students. These results emphasize that the severity of cognitive issues is not uniformly distributed across domains and reinforce the importance of analyzing each domain independently before inferring overall cognitive performance.

Table 6. Statistic Descriptive of Each Cluster

Cluster	*Statistic*	*Logic*	Abstraction	Verbal
0	*count*	4929	4929	4929
	mean	0.138463	0.345728	0.449179
	std	0.068379	0.100881	0.115161
	min	0	0.056871	0.108913
	25%	0.084498	0.279524	0.36875
	50%	0.136854	0.351463	0.452292
	75%	0.190651	0.416907	0.533369
	max	0.306943	0.632158	0.733813
1	***Statistic***	***Logic***	**Abstraction**	**Verbal**
	count	3932	3932	3932
	mean	0.381729	0.376546	0.490474
	std	0.061028	0.100125	0.107326
	min	0.252155	0.080213	0.160668
	25%	0.331449	0.309656	0.416239
	50%	0.384653	0.380531	0.495496
	75%	0.433232	0.449231	0.567074
2	**Statistic**	**Logic**	**Abstraction**	**Verbal**
	count	4391	4391	4391
	mean	0.245964	0.547878	0.579366
	std	0.099698	0.083881	0.083685
	min	0	0.327209	0.286688
	25%	0.169204	0.489678	0.522468
	50%	0.246815	0.540505	0.580806
	75%	0.322044	0.60315	0.637324

4.4. Cluster Evaluation Metrics

Following the clustering process, the Spectral Clustering model was evaluated using two internal validation metrics: the Davies-Bouldin Index (DBI) and the Calinski-Harabasz Index (CHI). The results yielded a DBI value of 1.2354 and a CHI value of 5333.627, indicating an overall satisfactory clustering performance.

The Davies-Bouldin Index, which quantifies the average similarity between each cluster and its most similar one (lower values are better), suggests that while the clustering structure is generally acceptable, some degree of overlap or internal variance remains within the clusters. This implies that the boundaries between some clusters may not be fully distinct. Future improvements could be achieved by optimizing the clustering parameters or further refining the dataset, such as by removing ambiguous data points.

Conversely, the Calinski-Harabasz Index, which evaluates the ratio of between-cluster dispersion to within-cluster dispersion (higher values are better), scored 5333.627,

signifying a strong cluster structure. This high value reflects well-separated and internally cohesive clusters, indicating that the Spectral Clustering model effectively captures meaningful patterns and distinctions within the data.

4.5. Classification Evaluation

This section presents a comparative summary of the classification performance of the three machine learning models used in this study: Random Forest, CatBoost, and XGBoost. All models were evaluated using a consistent dataset and assessed through four key metrics: accuracy, macro-averaged precision, recall, and F1-score. Table 7 provides a comparison among those three selected classification models.

Among the three models, CatBoost achieved the highest accuracy (98.79%) and maintained balanced precision, recall, and F1-score across all clusters. Its confusion matrix also indicates the lowest number of misclassified instances, particularly in Cluster 2, which exhibited slightly more overlap with other clusters. While Random Forest and XGBoost also demonstrated strong performance, their slightly higher misclassification rates and marginally lower accuracy suggest that CatBoost provides a more robust and reliable classification outcome for this dataset. Based on these findings, CatBoost is identified as the most suitable model for classifying the clustered data, offering both high accuracy and minimal classification error.

Table 7. Classification Evaluation Results

Model	Accuracy (%)	Precision (Macro)	Recall (Macro)	F1-Score (Macro)	Notes on Confusion Matrix
Random Forest	98.68	0.99	0.99	0.99	Cluster 0: 966 correct, 11 misclassified
					Cluster 1: 807 correct, 8 misclassified
					Cluster 2: 843 correct, 16 misclassified
CatBoost	98.79	0.99	0.99	0.99	Cluster 0: 970 correct, 7 misclassified
					Cluster 1: 807 correct, 8 misclassified
					Cluster 2: 846 correct, 13 misclassified
XGBoost	98.62	0.99	0.99	0.99	Cluster 0: 970 correct, 7 misclassified
					Cluster 1: 804 correct, 11 misclassified
					Cluster 2: 845 correct, 14 misclassified

4.6. Cognitive Severity Model

Figure 2 illustrates the final model developed to classify cognitive severity levels among university students based on their psychometric test responses. The model

workflow consists of three main stages: data pre-processing, classification, and severity level output.

The input to the model comprises psychometric test responses measuring three key cognitive domains: logical reasoning, abstraction, and verbal ability. These raw responses are first pre-processed through a structured pipeline. Each response is binary encoded (1 for correct answers, 0 for incorrect), then passed through two stages of Item Response Theory (IRT) analysis using the 3-Parameter Logistic (3PL) model. The first stage filters out low-quality items based on discrimination (a), difficulty (b), and guessing (g) parameters. The second stage computes the latent ability estimates (θ values) for each cognitive domain. These theta scores are then normalized using the MinMaxScaler to ensure consistent scaling between 0 and 1.

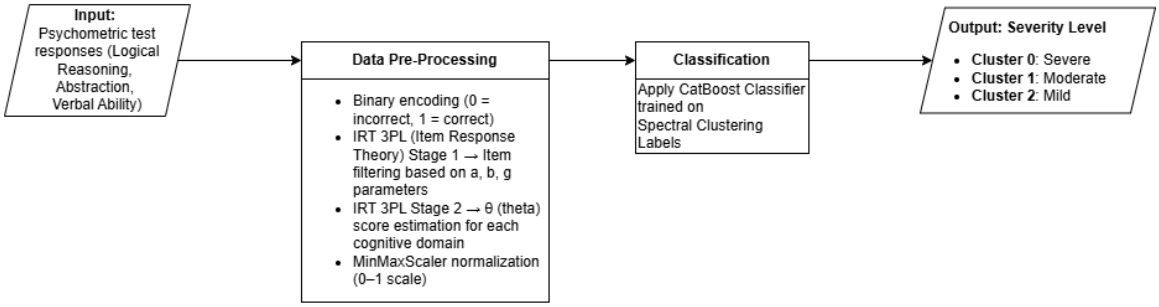


Figure 2 Cognitive Severity Classification Model Using CatBoost and Spectral Clustering

The normalized theta values are fed into a CatBoost classifier that has been previously trained on cluster labels derived from Spectral Clustering. These cluster labels were generated from an unsupervised clustering process on a large psychometric dataset and reflect three severity levels of cognitive difficulty.

The final output of the model assigns each student to one of the three identified clusters:

a. Cluster 0: Severe – indicating students with low cognitive functioning across domains,
b. Cluster 1: Moderate – indicating intermediate cognitive challenges,
c. Cluster 2: Mild – indicating relatively less severe or isolated cognitive difficulties.

This classification model offers a practical tool for early identification of students requiring differentiated academic support based on latent cognitive profiles.

4.7. Multivariate Validation of Cluster Separation (PERMANOVA Results)

To formally assess whether the identified cognitive severity clusters represent statistically distinct multivariate structures, a PERMANOVA analysis was conducted on the combined theta variables (logical reasoning, abstraction, and verbal ability). The analysis was performed on the full clustered sample (n = 13,252). Pairwise distances between observations were computed using Euclidean distance based on the three normalized theta variables.

The test began with the formulation of the following hypotheses:

- **Null Hypothesis (H_0):** There is no significant difference in the combined theta values (logical reasoning, abstraction, and verbal ability) across the identified clusters.
- **Alternative Hypothesis (H_1):** There is a significant difference in the combined theta values (logical reasoning, abstraction, and verbal ability) across the identified clusters.

PERMANOVA results indicated a highly significant multivariate separation among the three clusters:

- Pseudo-F = 4687.0
- R^2 = 0.2613
- p = 0.01
- N = 13,252

The large pseudo-F statistic indicates that between-cluster variability substantially exceeds within-cluster variability in multivariate space. The R^2 value suggests that approximately 26.13% of the total variance in combined cognitive profiles is explained by cluster level. In high-dimensional behavioral and psychometric data, R^2 values of this magnitude are considered substantial, given the inherent heterogeneity of human cognitive performance.
The permutation-based p-value (p = 0.01) confirms that the observed separation is unlikely to arise from random label assignment. Therefore, the null hypothesis of no multivariate difference between clusters is rejected.

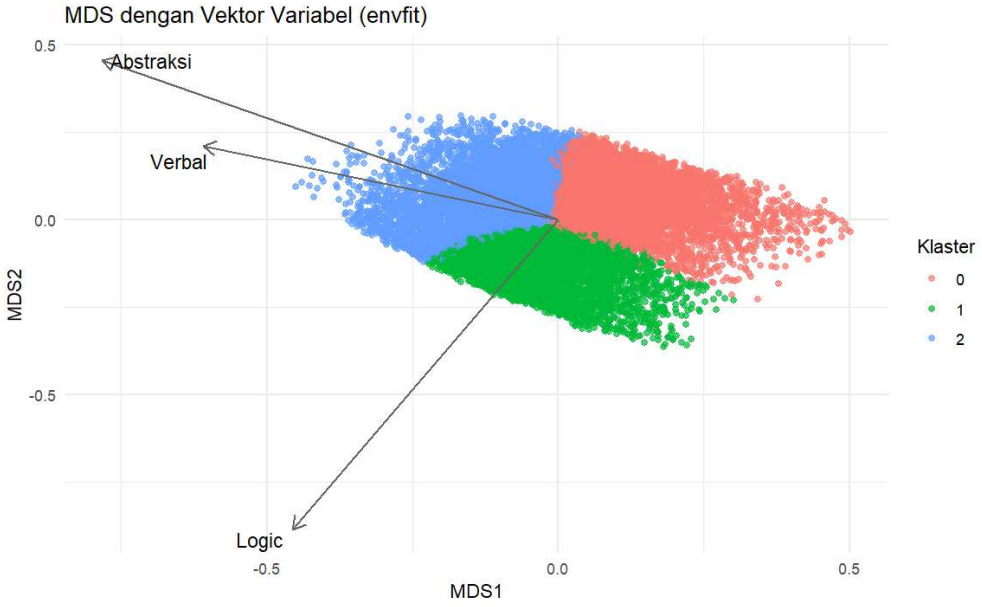


Figure 3 NMDS ordination of cognitive theta space with fitted vectors (envfit).

The NMDS projection in figure 3, reveals clear spatial separation among the three severity groups, consistent with the PERMANOVA findings. Cluster 0 (Severe) occupies a region associated with lower abstraction and verbal loadings, while Cluster 2 (Mild) aligns in the opposite direction. The fitted vectors indicate that abstraction and verbal ability contribute primarily along the positive MDS2 axis, whereas logical reasoning exhibits a distinct directional loading. This confirms that cluster separation

reflects structured multivariate differentiation rather than random geometric partitioning.

4.7.1. Pairwise Post Hoc PERMANOVA

To further examine the structural differentiation between severity groups, pairwise PERMANOVA comparisons were conducted across all cluster combinations. The results are presented in Table 8.

Table 8. Pairwise Comparisons

Cluster Comparison	Pseudo-F	R^2	p-value
Cluster 0 vs Cluster 1	5036.6	0.3625	0.01
Cluster 0 vs Cluster 2	6156.5	0.3978	0.01
Cluster 1 vs Cluster 2	4698.3	0.3608	0.01

All pairwise comparisons yielded statistically significant differences ($p = 0.01$), indicating that each severity group occupies a distinct position within the three-dimensional cognitive theta space. The comparison between Cluster 0 (Severe) and Cluster 2 (Mild) produced the highest pseudo-F value ($F = 6156.5$) and the largest effect size ($R^2 = 0.3978$), suggesting the strongest multivariate separation between the lowest and highest cognitive severity groups. This result confirms that students classified as severe and mild exhibit substantially different overall cognitive profiles across logical reasoning, abstraction, and verbal ability. The separation between Cluster 0 (Severe) and Cluster 1 (Moderate) was also statistically significant ($F = 5036.6$, $R^2 = 0.3625$), indicating that the moderate group represents a distinct intermediate severity structure rather than a gradual extension of the severe group. Similarly, the comparison between Cluster 1 (Moderate) and Cluster 2 (Mild) demonstrated significant differentiation ($F = 4698.3$, $R^2 = 0.3608$). Although the effect size was slightly smaller than the Severe–Mild contrast, the magnitude of explained variance remains substantial, supporting the validity of a three-level severity stratification.

Overall, the pairwise PERMANOVA findings confirm that the clusters identified through Spectral Clustering are not arbitrary partitions produced by the algorithm. Instead, they reflect statistically meaningful and structurally distinct multivariate cognitive severity profiles. The moderate-to-large R^2 values across all comparisons further indicate that cluster level explains a considerable proportion of variance in the latent theta space.

5. Discussions

This study examined whether psychometrically derived latent traits can be meaningfully stratified using spectral clustering, whether a two-level clustering strategy can reveal finer subgroup structures, and whether the resulting clusters are stable enough to support supervised classification.

5.1. RQ1: Do the combined latent cognitive traits form statistically distinct clusters when analyzed using spectral clustering?

The findings provide strong support for RQ1. Spectral clustering applied to the three latent cognitive traits (logical reasoning, abstraction, and verbal ability) produced three clearly differentiated clusters. To formally evaluate whether these clusters represent statistically distinct multivariate structures, a PERMANOVA analysis was conducted based on Euclidean distances computed from the three normalized theta variables.

The PERMANOVA results showed significant multivariate separation (pseudo-F = 4687.0, $R^2 = 0.2613$, $p = 0.01$). These findings indicate that approximately 26% of the total variance in the combined cognitive profiles is explained by cluster membership. In large and heterogeneous behavioral datasets, this magnitude of explained variance reflects meaningful structural differentiation rather than random grouping.

Because the permutation-based p-value was statistically significant ($p = 0.01$), the null hypothesis (H_0) is rejected. This confirms that the clusters differ significantly in their joint cognitive trait composition.

Pairwise post hoc PERMANOVA further confirmed that all cluster pairs were significantly different, with the strongest separation observed between the lowest and highest severity groups ($R^2 = 0.3978$). This suggests that the extreme clusters occupy clearly distinct positions within the multidimensional cognitive representation.

The NMDS visualization with fitted vectors also showed consistent directional patterns. Abstraction and verbal ability contributed strongly along one axis, whereas logical reasoning showed a distinct directional influence. This indicates that the cluster structure reflects systematic multivariate differentiation across domains rather than arbitrary partitioning. Overall, these results demonstrate that spectral clustering identifies statistically supported multivariate severity partitions based on combined latent cognitive traits.

5.2. RQ2: Can a two-level spectral clustering approach reveal more detailed and coherent subgroups within heterogeneous regions of the data?

The second-stage clustering focused on the most heterogeneous low-performing cluster identified in the first stage. This refinement revealed additional subgroup structure within this region.

Although the numerical improvement in Silhouette Score was modest after outlier removal, the refined clusters showed clearer internal separation and improved interpretability in visualization. The second-level clustering revealed domain-specific differences, particularly in logical reasoning compared to abstraction and verbal ability. This suggests that severity patterns are not uniform across domains and that additional structure can exist within broad global clusters.

The two-level approach therefore allowed the model to capture finer variation within a heterogeneous region of the latent space without destabilizing the overall partitioning.

Rather than forcing a single global segmentation to explain all variability, hierarchical refinement enabled more detailed and structurally coherent subgroup identification.

These findings support the value of multi-level spectral clustering when analyzing structured latent cognitive data.

5.3. RQ3: Are the identified clusters stable and predictable enough to be reliably classified using supervised machine learning?

The supervised classification results indicate high predictive stability of the derived clusters. CatBoost achieved 98.79% accuracy with balanced precision, recall, and F1-scores across clusters. Random Forest and XGBoost also demonstrated strong performance.

This high predictive accuracy suggests that the cluster labels represent internally consistent and reproducible structures in the data. If cluster assignments were unstable or weakly defined, classification performance would likely deteriorate. Instead, the results indicate that severity strata identified through spectral clustering can be reliably approximated using supervised models.

Importantly, this predictive validation does not imply diagnostic ground truth. Rather, it confirms that the unsupervised partitions are structurally coherent and sufficiently stable to support scalable inference in new datasets.

5.4. Methodological Implications

The integration of IRT-based latent trait modeling, spectral clustering, hierarchical refinement, and permutation-based multivariate validation provides a structured and scalable pipeline for modeling severity patterns in multidimensional cognitive data.

Three methodological contributions emerge:

1. Low-dimensional psychometric embeddings provide a stable basis for unsupervised clustering.
2. Spectral clustering effectively captures structured separation in correlated latent spaces.
3. Hierarchical refinement enables finer subgroup discovery within heterogeneous regions without compromising global cluster integrity.

By combining geometric validation (PERMANOVA and NMDS) with predictive evaluation, the study moves beyond purely descriptive clustering and demonstrates a statistically grounded framework for multivariate severity stratification.

6. Conclusion

This study proposed a two-level spectral clustering framework for multivariate cognitive severity stratification using psychometrically derived latent traits. By

embedding logical reasoning, abstraction, and verbal ability into a unified three-dimensional theta space, the approach demonstrated that structured severity partitions can be identified and statistically validated within a compact latent representation.

The results confirmed that spectral clustering produced statistically distinct multivariate clusters, supported by significant PERMANOVA findings and consistent geometric separation in NMDS visualization. Hierarchical refinement further revealed finer subgroup structure within heterogeneous regions of the latent space, improving granularity without compromising global partition stability. The high predictive performance of supervised classifiers indicates that the derived severity strata are internally coherent and reproducible, supporting scalable deployment in applied settings.

Methodologically, this work contributes to artificial intelligence research by integrating psychometric latent modeling with graph-based clustering, hierarchical refinement, and permutation-based multivariate validation in a single analytical pipeline. The study demonstrates how unsupervised learning can be applied in structured latent spaces to produce interpretable and statistically grounded partitions without relying on strong parametric assumptions.

Several limitations should be acknowledged. First, the framework was evaluated within a three-dimensional latent representation derived from domain-specific IRT models; extensions to higher-dimensional or cross-institutional datasets require further investigation. Second, although spectral clustering demonstrated effective separability, parameter sensitivity and alternative graph construction strategies were not exhaustively explored. Future research may examine stability under different similarity metrics, cross-validation across independent cohorts, and integration with adaptive learning systems.

Overall, the proposed framework provides a scalable and statistically supported approach for modeling structured severity patterns in multidimensional cognitive data. By combining latent trait modeling, hierarchical spectral clustering, and formal multivariate validation, this study advances interpretable unsupervised learning for complex behavioral datasets.

References

1. Benassi, M. *et al.* Using Two-Step Cluster Analysis and Latent Class Cluster Analysis to Classify the Cognitive Heterogeneity of Cross-Diagnostic Psychiatric Inpatients. *Front. Psychol.* **11**, (2020).

2. Green, M. J., Girshkin, L., Kremerskothen, K., Watkeys, O. & Quidé, Y. A Systematic Review of Studies Reporting Data-Driven Cognitive Subtypes across the Psychosis Spectrum. *Neuropsychol. Rev.* **30**, 446–460 (2020).

3. Orsoni, M. *et al.* Preliminary evidence on machine learning approaches for clusterizing students' cognitive profile. *Heliyon* **9**, e14506 (2023).

4. Hooshyar, D., Yang, Y., Pedaste, M. & Huang, Y.-M. Clustering Algorithms in an Educational Context: An Automatic Comparative Approach. *IEEE Access* **8**, 146994–147014 (2020).

5. Masyfa, F. H., Siahaan, D., Yuhana, U. L. & Hartono, P. Optimizing Cognitive Profiling Assessment Through Comparative Study of Hybrid Clustering-Classification Approaches. in *2024 IEEE Technology & Engineering Management Conference - Asia Pacific (TEMSCON-ASPAC)* 1–8 (IEEE, 2024). doi:10.1109/TEMSCON-ASPAC62480.2024.11025043.

6. Garcia-Rudolph, A., Garcia-Molina, A., Opisso, E. & Tormos Muñoz, J. Personalized Web-Based Cognitive Rehabilitation Treatments for Patients with Traumatic Brain Injury: Cluster Analysis. *JMIR Med. Inform.* **8**, e16077 (2020).

7. Nguyen, T. T. T. & Chang, Y.-L. Insights into the Heterogeneity of Cognitive Aging: A Comparative Analysis of Two Data-Driven Clustering Algorithms. *J. Gerontol. B Psychol. Sci. Soc. Sci.* **80**, (2025).

8. Toffalini, E., Girardi, P., Giofrè, D. & Altoè, G. Entia Non Sunt Multiplicanda … Shall I look for clusters in my cognitive data? *PLoS One* **17**, e0269584 (2022).

9. Guo, L., Yang, J. & Song, N. Spectral Clustering Algorithm for Cognitive Diagnostic Assessment. *Front. Psychol.* **11**, (2020).

10. Mondal, R. *et al.* Clustering graph data: the roadmap to spectral techniques. *Discover Artificial Intelligence* vol. 4 Preprint at https://doi.org/10.1007/s44163-024-00102-x (2024).

11. Jin, J. Student behavior patterns in vocational education big data based on clustering algorithm. *Discover Artificial Intelligence* **5**, (2025).

12. de la Torre, J. & Douglas, J. A. Higher-Order Latent Trait Models for Cognitive Diagnosis. *Psychometrika* **69**, 333–353 (2004).

13. Liu, J., Lee, S. & Gu, Y. Exploratory General-Response Cognitive Diagnostic Models with Higher-Order Structures. *Psychometrika* **90**, 1651–1678 (2025).

14. Tan, Z. *et al.* A Tutorial on Cognitive Diagnosis Modeling for Characterizing Mental Health Symptom Profiles Using Existing Item Responses. *Prevention Science* **24**, 480–492 (2023).

15. Proust-Lima, C., Dartigues, J. & Jacqmin-Gadda, H. Joint modeling of repeated multivariate cognitive measures and competing risks of dementia and death: a latent process and latent class approach. *Stat. Med.* **35**, 382–398 (2016).

16. Li, X., Zhang, Y., Cheng, H., Zhou, F. & Yin, B. An Unsupervised Ensemble Clustering Approach for the Analysis of Student Behavioral Patterns. *IEEE Access* **9**, 7076–7091 (2021).

17. Utami, I. Q., Hwang, W.-Y. & Ningrum, R. A. Student's Behavior Clustering based on Ubiquitous Learning Log Data using Unsupervised Machine Learning. *Journal of Advanced Technology and Multidiscipline* **3**, 13–20 (2024).

18. Peach, R. L., Yaliraki, S. N., Lefevre, D. & Barahona, M. Data-driven unsupervised clustering of online learner behaviour. *NPJ Sci. Learn.* **4**, 14 (2019).

19. Pecuchova, J. & Drlik, M. Enhancing the Early Student Dropout Prediction Model Through Clustering Analysis of Students' Digital Traces. *IEEE Access* **12**, 159336–159367 (2024).

20. Durachman, Y. Clustering Student Behavioral Patterns: A Data Mining Approach Using K-Means for Analyzing Study Hours, Attendance, and Tutoring Sessions in Educational Achievement. *Artificial Intelligence in Learning* **1**, 35–53 (2025).

21. Ouassif, K., Ziani, B., Herrera-Tapia, J. & Kerrache, C. A. Empowering Education: Leveraging Clustering and Recommendations for Enhanced Student Insights. *Educ. Sci. (Basel).* **15**, 819 (2025).

22. Ding, L., Li, C., Jin, D. & Ding, S. Survey of spectral clustering based on graph theory. *Pattern Recognit.* **151**, 110366 (2024).

23. Zhu, X., Gan, J., Lu, G., Li, J. & Zhang, S. Spectral clustering via half-quadratic optimization. *World Wide Web* **23**, 1969–1988 (2020).

24. Tang, C. *et al.* Unified One-Step Multi-View Spectral Clustering. *IEEE Trans. Knowl. Data Eng.* **35**, 6449–6460 (2023).

25. Hu, Z., Nie, F., Wang, R. & Li, X. Multi-view spectral clustering via integrating nonnegative embedding and spectral embedding. *Information Fusion* **55**, 251–259 (2020).

26. El Hajjar, S., Dornaika, F. & Abdallah, F. Multi-view spectral clustering via constrained nonnegative embedding. *Information Fusion* **78**, 209–217 (2022).

27. Wang, Y. *et al.* Latent Transition and Predictors of Cognitive Function Among Community-Dwelling Chinese Older Adults With Cardiovascular Metabolic Diseases. *Int. J. Geriatr. Psychiatry* **40**, (2025).

28. Smith, K. V. & Ehlers, A. Cognitive predictors of grief trajectories in the first months of loss: A latent growth mixture model. *J. Consult. Clin. Psychol.* **88**, 93–105 (2020).

29. von Luxburg, U. A tutorial on spectral clustering. *Stat. Comput.* **17**, 395–416 (2007).

30. Jia, H., Ding, S., Xu, X. & Nie, R. The latest research progress on spectral clustering. *Neural Comput. Appl.* **24**, 1477–1486 (2014).

31. Anderson. Permutational Multivariate Analysis of Variance (PERMANOVA). *Wiley StatsRef: Statistics Reference Online* 1–15 (2017) doi:10.1002/9781118445112.stat07841.

32. Anderson, M. J. A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* **26**, 32–46 (2001).

Ethical Consideration and Consent to Participate

This study received ethical clearance with exempt status from the Research Ethics Committee of Bina Nusantara University (Ethical Clearance Result No: 016/VRRTT/I/2026) for a retrospective analysis of fully anonymized secondary data. Informed consent was obtained at the time of the original data collection as part of the institutional admission assessment process. No additional consent was required for the present retrospective analysis. All procedures were conducted in accordance with relevant institutional guidelines and the Declaration of Helsinki.

Consent for Publication

Not applicable. The manuscript does not contain any identifiable personal data.

Funding

The authors received no external funding for this research.

Clinical Trial Number

Not applicable.

Author Contributions

All authors contributed to the study conception and design. Data collection and analysis were carried out by Nuril Kusumawardani Soeprapto Putri. Umi Laili Yuhana and Daniel Oranova Siahaan served as supervisors, ensuring the robustness of the research design and providing critical academic guidance throughout the research process. The first draft of the manuscript was written by Nuril Kusumawardani Soeprapto Putri, and all authors commented on previous versions of the manuscript. Istiani served as the psychologist and psychometrics expert, responsible for conceptualizing, developing, and evaluating the psychometric test instruments used in this study. All authors read and approved the final version of the paper.

Data Availability Statement

The anonymized dataset generated and analyzed during the current study is publicly available on Zenodo at https://doi.org/10.5281/zenodo.17504949 under the Creative Commons Attribution 4.0 International License (CC-BY 4.0). The dataset includes only codified theta values and contains no personally identifiable information. Researchers may freely reuse the data for non-commercial academic purposes with proper citation.