Article

Not peer-reviewed version

Assessment of Long-Term Learning Through Item Response Theory in Moodle and H5P Virtual Environments: A Case Study of Leveling Students at the Escuela Politécnica Nacional

Paola Cabascango-Flores and Erick P. Herrera-Granda *

Posted Date: 4 March 2026

doi: 10.20944/preprints202603.0330.v1

Keywords: item response theory; confirmatory factorial analysis; statistical analysis; edumetrics

Article

Assessment of Long-Term Learning Through Item Response Theory in Moodle and H5P Virtual Environments: A Case Study of Leveling Students at the Escuela Politécnica Nacional

Paola Cabascango-Flores [1,2] and Erick P. Herrera-Granda [3,*]

[1] Departamento de Formación Básica, Escuela Politécnica Nacional, Quito 170525, Ecuador
[2] Posgrado, Universidad Politécnica Estatal del Carchi, Av. Universitaria y Antisana, Tulcán 040101, Carchi, Ecuador
[3] Departmento de Informática y Ciencias de la Computación, Escuela Politécnica Nacional, Quito 170525, Ecuador
* Correspondence: erick.herrera@epn.edu.ec ; Tel.: +593989460084

Abstract

This study integrated Item Response Theory (IRT) models with ordinal survey instruments to assess academic performance trajectories and identify multidimensional factors associated with academic achievement among first-semester leveling students (N=1,558 pre-test; N=1,676 post-test) at the Escuela Politécnica Nacional, Ecuador. A dual-component methodology was employed: (1) an 80-item ordinal survey measuring eight latent constructs (socioeconomic, academic, motivational, vocational, social integration, psychological/emotional, institutional, and biological/health factors), validated through Confirmatory Factor Analysis ($CFI > 0.95$, $RMSEA < 0.06$); and (2) structured diagnostic assessments in mathematics, physics, chemistry, geometry, and language, calibrated using three-parameter logistic (3PL) IRT models via Expected A Posteriori (EAP) estimation. Results demonstrated high internal consistency ($r = 0.93$ between IRT and raw scores), with mean IRT-scaled ability $\bar{\theta} = 10.45$ ($SD = 3.51$) on a 1–20 scale. Item parameters indicated adequate discrimination $\bar{a} = 1.92$) and centered difficulty ($\bar{b} = 0.05$), though 13.75% of items exhibited poor model fit ($S - X^2\, p < 0.01$), concentrated in physics and chemistry domains. Factorial scores and performance outcomes were statistically contrasted against 24 categorical demographic variables, revealing differential performance patterns across student subgroups. This research provides validated psychometric instruments, reproducible IRT-LMS integration protocols, and empirical evidence supporting targeted interventions to strengthen university transition in resource-constrained contexts.

Keywords: item response theory; confirmatory factorial analysis; statistical analysis; edumetrics

---

1. Introduction

Over the past decade, educational assessments have evolved from static approaches to more complex, adaptive frameworks that model student performance and learning progression with greater precision [1]. Item Response Theory (IRT) is one such framework. IRT estimates the probability of correct responses based on students' latent abilities and item-specific parameters, such as difficulty, discrimination, and guessing. This generates comparable interpretations across cohorts and applications [2,3]. The utility of this approach was solidified through international assessments, such as PISA, which revealed significant differences between educational systems and informed public policy decisions aimed at equity and improvement [4]. However, adopting it in contexts with

 doi:10.20944/preprints202603.0330.v1

technological constraints and limited teacher training remains challenging, especially when integrating it into virtual learning ecosystems [5,6].

In Ecuador, the Learning Standards for the Unified General Baccalaureate (BGU) defined the expected competencies in key subject areas and were used as a reference for secondary-level exit exams [7,8]. Despite these standards, multiple higher education institutions reported entry-level deficiencies that manifested in leveling courses, especially in foundational competencies in mathematics, physics, and chemistry [9,10]. The Escuela Politécnica Nacional encountered the same challenge, requiring initial diagnostic assessments that estimated point-in-time mastery, characterized learning trajectories, and provided evidence of long-term learning gains.

Learning management platforms, such as Moodle, when combined with interactive resources like H5P, create an environment conducive to diagnostic assessments with immediate feedback and continuous monitoring [5,6]. However, traditional practices in these environments typically focused on point-in-time measurements with limited capacity to infer long-term learning or distinguish between observable performance and actual knowledge gains. Integrating IRT models within Moodle/H5P enables: (1) more precise estimation of latent abilities, (2) monitoring of change between initial and final measurements, and (3) alignment of score interpretation with BGU performance standards. This provides evidence for personalized pedagogical decisions [2,3,7,8]. The primary barriers were teacher preparation for item design, calibration, and institutional technical orchestration [5].

Within this framework, the present study evaluated the feasibility and added value of a comprehensive methodological approach combining ordinal survey instruments and IRT-based diagnostic assessments for leveling students at the Escuela Politécnica Nacional, aligned with BGU standards. The research employed a dual-component instrument: first, an ordinal survey comprising 80 items measuring factors associated with academic success or attrition among first-year students—including socioeconomic, academic, motivational, vocational, social integration, psychological/emotional, institutional, and biological/health dimensions—which was validated and calibrated through Confirmatory Factor Analysis (CFA) to extract factorial scores for each latent construct. Second, a structured performance assessment measured student competency at three critical timepoints: before the first semester, upon completion of the first academic period, and at the conclusion of each subject (mathematics, physics, chemistry, geometry, language, literature, among others), enabling the characterization of learning trajectories in response to instructional interventions. These performance data were calibrated and weighted using IRT models to derive more precise estimates of students' actual abilities. Subsequently, factorial scores from each dimension and subject-specific performance outcomes were statistically contrasted against 24 categorical variables, complemented by numerical variable comparisons, which facilitated the precise identification of factors and circumstances influencing first-semester student success or attrition. The expected contribution was threefold: psychometric—validated instruments capturing multidimensional risk profiles; methodological—a reproducible workflow for integrated assessment applicable in resource-constrained contexts; and practical—evidence-based inputs for adjusting leveling strategies oriented toward gap reduction and strengthening the transition to university-level education [11].

1.1. Related Works

The quantification of factors associated with academic performance in newly admitted university students has necessitated integrating rigorous psychometric methodologies capable of processing structured diagnostic assessments, ordinal survey responses, and institutional statistical records. Foundational advances centered on Item Response Theory frameworks to calibrate diagnostic instruments and ensure measurement precision across heterogeneous populations. Ntumi (2025) [12] employed advanced Multidimensional Item Response Theory combined with Confirmatory Factor Analysis and Rasch analysis to evaluate STEM, language, and cognitive competencies in Sub-Saharan African students, utilizing Mardia's skewness and Henze-Zirkler tests

 doi:10.20944/preprints202603.0330.v1

for multivariate non-normality while implementing Differential Item Functioning via Mantel-Haenszel and logistic regression to detect socioeconomic bias, achieving 35% test length reduction through adaptive algorithms while maintaining precision. Ortiz-Rojas et al. (2019) [13] applied two-parameter IRT models to calibrate programming assessments for first-year engineering students, integrating Confirmatory Factor Analysis validation of motivation surveys with Structural Equation Modeling to test mediation hypotheses, revealing that gamification produced direct performance effects rather than motivationally mediated outcomes. Ryan (2018) [14] extended IRT through Rasch Analysis, utilizing Dichotomous and Rating Scale Models to calibrate assessment literacy instruments, employing Confirmatory Factor Analysis with Unweighted Least Squares estimation on polychoric correlation matrices, subsequently applying Moderated Multiple Regression to quantify knowledge-performance relationships.

Psychological screening instrument calibration via IRT received substantial attention. Sugawara et al. (2023) [15] implemented two-parameter IRT models to estimate discrimination and difficulty for the University Personality Inventory, administered to 1,185 first-year medical students. They confirmed unidimensionality through an Exploratory Factor Analysis using tetrachoric coefficients and generated Test Information Functions to visualize optimal measurement ranges. Avcu (2021) [16] employed the Graded Response Model to evaluate the Satisfaction with Life Scale among 471 undergraduates, verifying dimensionality through Exploratory and Confirmatory Factor Analysis, assessing local independence via Yen’s Q3 statistics, and conducting Differential Item Functioning via logistic regression, ensuring gender invariance. The results revealed that the category response functions indicated seven-point scale redundancy. Appolloni et al. (2025) [17] integrated Classical Test Theory with three-parameter logistic IRT, validating a 30-item dichotomous questionnaire, utilizing Kuder-Richardson 20 for internal consistency while modeling discrimination, difficulty, and pseudo-guessing parameters. Wang et al. (2022) [18] adapted the Dark Factor of Personality Scale for Chinese students exclusively through Graded Response Model calibration, conducting unidimensionality assessment via Exploratory Factor Analysis and Differential Item Functioning analysis, ensuring gender invariance, applying Item Information Functions to develop maximally efficient short-forms. Aza-Espinosa et al. (2023) [19] calibrated 73 multiple-choice Natural Sciences items for 583 first-year students using Rasch logistic models, weighting items by difficulty and discrimination, subsequently employing Analysis of Variance and pairwise t-tests correlating weighted performance with internet accessibility and socioeconomic variables.

2. Materials and Methods

2.1. Study Area and Population Description

This research was conducted at the Leveling Department of the Escuela Politécnica Nacional (EPN), located in Quito, Ecuador (Figure 1). Founded in 1869, EPN constitutes one of Ecuador’s leading public technical universities, with an institutional mission to “train highly qualified, ethical professionals committed to national development through the generation and application of knowledge in science, technology, and culture,” and an institutional vision to consolidate itself as “a leading university in science, technology, and innovation in Latin America, recognized for its academic excellence and social impact.” The Leveling Department serves as a mandatory transitional academic program designed to bridge competency gaps between secondary education outcomes and university-level requirements across fundamental disciplines, including mathematics, physics, chemistry, geometry, language, and literature.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 March 2026 doi:10.20944/preprints202603.0330.v1

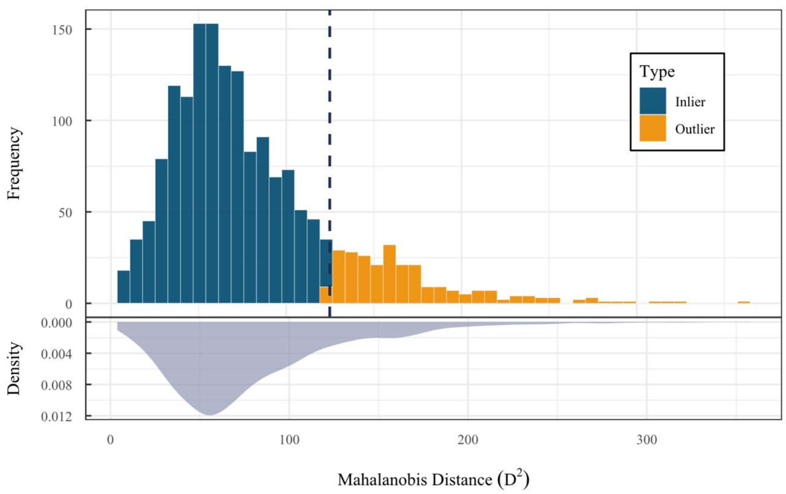


Figure 1. Histogram and density curve of Mahalanobis distances for each survey administered to students.

The target population comprised 1,762 students enrolled in the 2025-A leveling cohort, representing diverse academic backgrounds from Ecuador's Unified General Baccalaureate (BGU) system. Persistent institutional evidence indicated substantive entry-level deficiencies in foundational competencies, particularly in STEM disciplines, necessitating systematic diagnostic assessment and evidence-based intervention strategies. The institutional context presented a unique opportunity to examine the interplay between psychometric assessment, learning management system integration, and student success trajectories in a resource-constrained, high-stakes transition environment.

A representative sample was calculated using finite population sampling formulae with 95% confidence level and 5% margin of error, yielding a minimum required sample of n = 316. Inclusion criteria required formal enrollment status and active access to the institutional Moodle platform. The final analytical sample comprised 1,558 valid participants for the initial diagnostic assessment (Pre-Test) and 1,676 participants for the final assessment (Post-Test), following data cleaning procedures to remove incomplete responses or irregular participation patterns. This sample exceeded statistical requirements and represented 88.4% and 95.1% of the enrolled cohort, respectively, thereby minimizing selection bias and enhancing generalizability.

Diagnostic assessments were administered at three critical timepoints: (1) baseline assessment prior to instructional intervention (Week 1), (2) formative assessment upon completion of the first academic period (Week 8), and (3) summative assessment at course conclusion (Week 16). This longitudinal design enabled characterization of learning trajectories and quantification of instructional effects. Key stakeholders included enrolled students across all academic tracks and faculty members responsible for subject-specific instruction, who provided curricular alignment guidance and participated in item development workshops to ensure content validity and alignment with BGU learning standards and institutional competency frameworks.

2.2. Procedures

The research was executed through a sequential three-phase experimental design integrating psychometric calibration, longitudinal assessment, and multivariate statistical analysis. All procedures adhered to institutional ethical protocols with informed consent and data anonymization implemented throughout.

Phase 1: Baseline Diagnostic Assessment and Data Collection

The initial diagnostic assessment comprised two integrated components. The performance assessment consisted of 80 multiple-choice items distributed across five content domains: Fundamentals of Mathematics (20 items), Language and Communication (20 items), Geometry (20

 doi:10.20944/preprints202603.0330.v1

items), Physics (10 items), and Chemistry (10 items). Items were mapped to Ecuador's Unified General Baccalaureate curriculum standards (73.8%, n = 59) and institutional competency requirements (26.2%, n = 21), ensuring content validity and curricular alignment.

The assessment was deployed through Moodle 3.11 LMS, utilizing H5P interactive content framework with randomized item presentation, automatic timestamping, and real-time data capture. Participants completed the 90-minute supervised assessment during Week 1 of the 2025-A leveling period (April 26-27, 2025). Concurrently, an 80-item ordinal survey measuring eight latent constructs (socioeconomic, academic, motivational, vocational, social integration, psychological/emotional, institutional, and biological/health factors) was administered using five-point Likert scales, requiring approximately 30 minutes. Additionally, 24 categorical demographic variables were collected, including age, gender, geographic origin, high school type, household composition, parental education, and technology access.

Raw data were extracted and imported into R 4.3.1 for preprocessing. Quality control procedures included duplicate removal, multivariate outlier detection using Mahalanobis distance, missing data pattern assessment, and response time validation. Following these procedures, 1,558 complete protocols were retained (88.4% retention rate).

Phase 2: Psychometric Calibration and Item Analysis

The ordinal survey underwent Confirmatory Factor Analysis (CFA) using the lavaan package with diagonally weighted least squares estimation. Model fit was evaluated using CFI, TLI, RMSEA, and SRMR indices. Upon achieving acceptable fit (CFI ≥ 0.95, RMSEA ≤ 0.06), factorial scores were extracted and standardized.

For performance assessment, three-parameter logistic (3PL) Item Response Theory models were implemented using the mirt package. Item parameters (difficulty, discrimination, pseudo-guessing) were estimated via marginal maximum likelihood with the expectation-maximization algorithm. Student abilities were estimated using the Expected A Posteriori method and transformed to a 1-20 institutional scale. Item fit was evaluated using S-X^2 statistics ($p < 0.01$ threshold), and Test Information Functions identified regions of maximum precision.

Phase 3: Longitudinal Assessment and Comparative Analysis

Parallel-form assessments were administered at Week 8 (formative) and Week 16 (summative) under identical conditions. Item parameter anchoring enabled the placement of all ability estimates on a common metric for growth modeling. The final integrated database (n = 1,676) combined IRT ability estimates, factorial scores, performance trajectories, and demographic variables. Statistical analyses included descriptive characterization, differential item functioning analysis, reliability assessment, paired-samples comparisons, and multivariate modeling using both parametric and non-parametric methods (Kruskal-Wallis, Dunn-Šidák post-hoc) conducted in R 4.3.1 with $\alpha = 0.05$.

2.2. Statistical Analysis

Mahalanobis Distances: The database compiled for research often contains missing data and outliers, suggesting the need to begin any statistical analysis by implementing a comprehensive data analysis protocol. Among the analytical tools for data processing in multivariate samples, the use of Mahalanobis distances stands out. This technique quantifies the number of standard deviations that a specific observation is from the mean of a distribution. Since outliers do not follow a behavior pattern analogous to regular observations, applying this measure allows for identifying such anomalies. In contrast, from a geometric perspective, the Euclidean distance represents the shortest length between two points; however, it does not consider the correlation between highly interrelated variables. The Mahalanobis distance differs from the Euclidean distance in its ability to incorporate the correlation between variables in its calculation.

In this study, the Mahalanobis distance was employed as a scale-invariant metric that allows for calculating the distance between a point $x \in \mathbb{R}^p$ from a p-variate probability distribution $f_X(.)$, a p-value, and the mean $\mu = E(X)$ of that distribution. It was assumed that the distribution $f_X(.)$ has a finite number of second-order moments, facilitating the definition of the covariance matrix $\Sigma =$

 doi:10.20944/preprints202603.0330.v1

$E(X-\mu)$. Under these conditions, Mahalanobis distances are established through the following mathematical relationship:

$$D_M(x) = \sqrt{(x-\mu)^T \Sigma^{-1}(x-\mu)} \quad (1)$$

Spearman's correlation test.

Once the latent variables were obtained and outlier screening had been completed, the monotonic association among the selected constructs was assessed using Spearman's rank-order correlation, a nonparametric test widely employed when study variables are measured on an ordinal scale or when the assumption of normality is not met in the observed data [49,50]. This test evaluates the null hypothesis of no monotonic relationship between two variables:

$$H_0: p_s = 0, \quad (2)$$

where ρ_s denotes the population Spearman correlation coefficient. Let (X_i, Y_i) be a set of paired observations with $i = 1,2,\dots,n$ and let (X_i), $R(Y_i)$ represent the ranks assigned to each observation. The Spearman coefficient is defined as:

$$p_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}, \quad (3)$$

where $d_i = R(X_i) - R(Y_i)$ is the rank difference for the i-th pair. In the presence of tied ranks, the correlation is computed using the general formula based on rank covariances:

$$p_s = \frac{cov(Rx, Ry)}{\sigma R_x \sigma R_y}, \quad (4)$$

which ensures an unbiased estimate even when ties exist in the data [51]. For the estimation of the test statistic and the associated p-value, under the null hypothesis and for sufficiently large sample sizes, the transformed statistic:

$$t = \rho_s \sqrt{\frac{n-2}{1-\rho_s^2}}, \quad (5)$$

follows approximately a Student's t-distribution with $n-2$ degrees of freedom [52]. The p-value is then obtained as:

$$p = 2[1 - F_{t_{n-2}}(|\, t \,|)], \quad (6)$$

where $F_{t_{n-2}}(\cdot)$ denotes the cumulative distribution function of the Student's t-distribution. For small sample sizes, the test is conducted through exact methods based on rank permutations, thereby ensuring a precise estimation of the probability associated with the observed statistic [53].

In this manner, Spearman's rank correlation enabled the determination of both the strength and direction of monotonic relationships among variables, regardless of their underlying distributions, providing a robust analysis free of strict parametric assumptions and well suited to the objectives of the present study.

Confirmatory Factor Analysis. A confirmatory factor analysis (CFA) was conducted to assess the instrument's validity and reliability. CFA empirically tests the correspondence between observed indicators and latent factors, making it particularly suitable for ordinal response data. The proposed model considers a $p \times 1$ response vector constructed from observable random variables, capable of explaining one or more unobserved variables, referred to as factors η. Through this procedure, each observable variable contributes to explaining the behavior of the corresponding latent variable. Accordingly, the model considers a vector of observed responses $\boldsymbol{Y}_i$ used to explain the unobserved latent variable ξ by means of:

$$\boldsymbol{Y} = \boldsymbol{\Lambda}\,\xi + \epsilon, \quad (7)$$

where $\boldsymbol{Y}$ represents the observed random variables, ξ corresponds to the unobserved variables, and $\boldsymbol{\Lambda}$ denotes a $p \times k$ matrix, with k being the number of unobserved latent variables. Since $\boldsymbol{Y}$ is composed of a set of unobserved latent variables ξ, the model incorporates the error term ϵ.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 March 2026 doi:10.20944/preprints202603.0330.v1

Parameter estimation was carried out using the maximum likelihood (ML) method, derived from the iterative minimization of the following function:

$$F_{ML} = ln|\boldsymbol{\Lambda\Omega\Lambda}' + \boldsymbol{I} - diag(\boldsymbol{\Lambda\Omega\Lambda}')| + tr\big(R(\boldsymbol{\Lambda\Omega\Lambda}' + \boldsymbol{I} - -diag(\boldsymbol{\Lambda\Omega\Lambda}')^{-1})\big) - \ln(R) - p, \tag{8}$$

where $\boldsymbol{\Lambda\Omega\Lambda}'$ represents the variance-covariance matrix derived from the factor analytic model and $\boldsymbol{R}$ [43,44,54–57]**.**

Item Response Theory. Item response theory (IRT) was employed to model the probability that a student with a given level of latent ability (θ) would respond correctly to an item, taking into account the psychometric properties of each question. To this end, the Rasch model (1PL), proposed by [58], and the three-parameter logistic model (3PL), developed by [59] and formalized in the work of [60], were utilized. The Rasch model assumes equal discrimination capacity across items and depends solely on the difficulty parameter b_i, expressed as:

$$P_i\,(\theta) = \frac{e^{(\theta-b_i)}}{1+e^{(\theta-b_i)}}, \tag{10}$$

where $P_i\,(\theta)$ denotes the probability of a correct response by an individual with ability level θ. This model provides invariant and comparable measurement estimates. The 3PL model, in turn, introduces the discrimination parameter (a_i) and the pseudo-guessing parameter (c_i), and is expressed as:

$$P_i\,(\theta) = c_i + (1-c_i)\frac{1}{1+e^{-a_i(\theta-b_i)}}. \tag{11}$$

Parameter estimation was performed using maximum likelihood, optimizing the following function:

$$L(\theta,a,b,c) = \prod_{i=1}^{n}[\,P_i\,(\theta)]^{ui}[1-P_i]^{(1-u_i)}. \tag{12}$$

Kruskal–Wallis and Dunn–Šidák Tests. Once the estimated latent factors were obtained, a dataset free from distortions caused by atypical observations was derived, comprising a set of variables capable of adequately explaining the factors of interest in the investigation. A comparative analysis among groups was then conducted using the nonparametric Kruskal–Wallis test, which is appropriate for ordinal variables. This test evaluates the null hypothesis:

$$H_0\colon \eta_1 = \eta_2 = \cdots = \eta_k, \tag{13}$$

where η_i denotes the median of the i-th group and n represents the total number of observations. Defining $n = \sum_{i=1}^{k} n_i$, where n_i is the sample size of each group $i = 1,2,\cdots,k$, and k represents the number of groups to be compared, $R(X_{ij})$ denotes the rank assigned to the j-th observation of the i-th group X_{ij} and R_i represents the sum of ranks assigned to the i-th group, $R_i = \sum_{i=1}^{n_i} R(X_{ij})$. The test statistic is thus expressed as:

$$T = \frac{1}{S^2}\left(\sum_{i=1}^{k}\frac{R_i^2}{n_i} - \frac{n(n+2)^2}{4}\right), \tag{14}$$

where:

$$S^2 = \frac{1}{n-1}\left(\sum_{allrank} R\big(X_{ij}\big)^2 - \frac{n(n-1)^2}{2}\right), \tag{15}$$

in the absence of tied ranks, S^2 reduces to $n(n+1)/12$, and the test statistic simplifies to Equation 6:

$$T = \frac{12}{n(n+1)}\sum_{i=1}^{k}\frac{R_i^2}{n_i} - 3(n+1), \tag{16}$$

under the null hypothesis H_0 and the previously stated assumptions, T is asymptotically distributed as a chi-squared distribution with $k-1$ degrees of freedom, $T \sim \chi^2_{k-1}$ [53,61].

When the overall test yields a statistically significant result, the Dunn–Šidák post hoc test is applied for pairwise multiple comparisons, which is based on the Šidák inequality and is more precise than the Bonferroni correction [62]. For a given family-wise error rate (FWER) α, the Dunn–Šidák contrast defined as $\mu_i - \mu_j$ is expressed as:

$$\mu_i - \mu_j = \bar{y}_i - \bar{y}_j \pm t_{\alpha',\nu} \sqrt{s^2 \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}, \tag{17}$$

where:

$$\alpha' = \frac{1}{2}\left(1 - (1-\alpha)^{\frac{1}{c}}\right), \tag{18}$$

$\bar{y}_i$ and $\bar{y}_j$ are the means of the samples under consideration, c indicates the total number of contrasts in the family, and the quantile $t_{\alpha',\nu}$ is obtained from the Student's t-distribution for a given degrees-of-freedom parameter ν. Finally, the confidence intervals for each possible Dunn–Šidák contrast are given by the expression:

$$\sum_{i=1}^{k} c_i \bar{y}_i \pm t_{\alpha',\nu} \sqrt{s^2 \sum_{i=1}^{k} \frac{c_i^2}{n_i}}. \tag{19}$$

3. Results

The instrument proposed in the present study comprises two components: an 80-item survey designed to measure eight unobserved variables associated with the academic performance of newly admitted university students, and an 80-question, structured test that evaluates five subjects taught in the EPN leveling course. The structured test was administered to students before they began their leveling course studies at the university. Databases collected with each instrument were processed to mitigate noise, bias, and outliers. Both instruments were then validated, enabling the generation of models for each that facilitated the extraction of factor scores for the survey and weighted scores for the structured test. These scores served as valid indicators of factors associated with performance and achievement among students entering higher education institutions. The processed scores for each student were used to perform all relevant hypothesis tests guided by categorical variables and research hypotheses as post hoc analyses. Section 3.1 details the validation and score extraction for the survey; Section 3.2 presents the validation and score extraction for the structured tests; and Section 3.3 reports the hypothesis tests conducted.

3.1. Survey Analysis

This component of the proposed methodology consists of a factorial instrument consisting of 80 questions on a 7-point Likert scale, where students had to rate their level of agreement with each question on a scale ranging from strongly disagree (level 1) to strongly agree (level 7). Each factor was designed as an unobserved latent variable, taking into account 10 questions related to the factor, which were designed by the team of professors who teach undergraduate courses to newly admitted students at the institution, supervised and validated by a team of educational psychologists. In this way, the factorial instrument took into account the following factors: socioeconomic, academic, motivational, vocational, social integration, psychological/economic, institutional and biological/health. The questions designed and the full factorial structure are presented in Table 1.

Table 1. Questions that make up the designed factorial instrument.

Variable	Question
	Socioeconomic factor
p_1	My home has the basic resources I need to study (books, internet, supplies).
p_2	I have a comfortable and quiet space to do my homework.
p_3	My parents or guardians can support me with my studies.
p_4	My family's financial situation is stable.
p_5	I have regular access to the internet at home.

p_6	My family's financial situation does not interfere with my studies.
p_7	In my home, education is valued as a means of self-improvement.
p_8	I have sufficient school supplies for the academic year.
p_9	Economic concerns affect my concentration when studying..
p_{10}	There is sufficient financial support to cover my educational needs.
	Academic factor
p_{11}	Do you think it is important to set aside a week solely for exams, i.e., without regular classes?
p_{12}	I regularly attend all my classes.
p_{13}	I organize my time to complete my academic tasks.
p_{14}	I have good study habits.
p_{15}	I review the content covered in class before exams.
p_{16}	Aplico estrategias para mejorar mi comprensión lectora.
p_{17}	I am able to relate new knowledge to previous learning.
p_{18}	I prepare my work in advance.
p_{19}	I actively participate during classes.
p_{20}	I strive to learn even when the content is complex.
	Motivational Factor
p_{21}	Motivational Factor Studying generates a sense of personal satisfaction.
p_{22}	I put effort into my studies because I want to achieve my goals.
p_{23}	I have confidence in my ability to learn.
p_{24}	I enjoy learning new things, even when they are not assessed.
p_{25}	I exert greater effort when I know I will receive some form of reward.
p_{26}	I feel motivated when I observe progress in my academic outcomes.
p_{27}	I believe that what I learn will be useful in my future life.
p_{28}	I am committed to my studies.
p_{29}	I strive in my studies so that my family feels proud.
p_{30}	I persist in my tasks even when I encounter difficulties.
	Vocational Factor
p_{31}	I am certain that I chose the right program of study for me.
p_{32}	My decision to pursue this program was personal and deliberate.
p_{33}	I am passionate about the subjects I study.
p_{34}	I am well acquainted with the professional field related to my area of study.
p_{35}	My personal interests align with what I study.
p_{36}	I have clear goals regarding my professional future.
p_{37}	What I study is related to what I want to do in my life.
p_{38}	I identify with the profession for which I am being trained.
p_{39}	My family's influence was a determining factor in my career choice.
p_{40}	I envision myself practicing this profession in the coming years.
	Social Integration Factor
p_{41}	I have friendships within the academic environment.
p_{42}	I feel part of the group in my classes.
p_{43}	I find it easy to interact with my classroom peers.
p_{44}	I feel supported by my instructors.
p_{45}	I participate in extracurricular activities at my institution.
p_{46}	I feel respected in my educational environment.
p_{47}	I collaborate with other students on assignments and group projects.
p_{48}	I feel a sense of belonging to this educational institution.
p_{49}	I relate well with individuals from cultures or backgrounds different from my own.
p_{50}	I feel included within the student body.
	Psychological/Emotional Factor
p_{51}	I feel confident in my intellectual abilities.

 doi:10.20944/preprints202603.0330.v1

p_{52}	I am able to manage stress related to examinations.
p_{53}	I have positive self-esteem regarding my academic performance.
p_{54}	I become easily frustrated when I receive poor grades.
p_{55}	I tend to experience anxiety before taking exams.
p_{56}	When I have an emotional problem, I find it difficult to concentrate.
p_{57}	I feel motivated most of the time to study.
p_{58}	I am able to express my emotions in an appropriate manner.
p_{59}	I maintain emotional balance during academically demanding periods.
p_{60}	I acknowledge my achievements and feel proud of them.
	Institutional Factor
p_{61}	The instruction provided by my instructors is clear and comprehensible.
p_{62}	The assessments administered are fair and consistent with the content taught.
p_{63}	My institution has sufficient didactic resources available.
p_{64}	The physical facilities are adequate for learning.
p_{65}	I have access to tutoring or additional academic support.
p_{66}	Instructors demonstrate willingness to address our questions.
p_{67}	The institution fosters a climate of respect and inclusion.
p_{68}	Educational technologies that support learning are utilized.
p_{69}	The form of assessment motivates my learning.
p_{70}	The experience of studying at the EPN has met my academic and personal expectations, and this has positively influenced my performance.
	Biological and Health Factor
p_{71}	I sleep at least seven hours each night.
p_{72}	I maintain a balanced diet throughout the week.
p_{73}	I engage in physical activity at least twice a week.
p_{74}	I feel energized during the day to study.
p_{75}	I do not have health problems that interfere with my learning.
p_{76}	I visit a physician regularly when I feel unwell.
p_{77}	My physical condition (health) allows me to remain focused.
p_{78}	I do not consume substances that affect my academic performance.
p_{79}	When I am ill, I find it difficult to perform academically.
p_{80}	In general, I have good health.

The survey presented in Table 1 was administered to the entire student population enrolled in the 2025A career-leveling semester, yielding a total of 1,670 observations. The main descriptive statistics—mean, median, quartiles, standard deviation, and distribution—were examined, revealing that the items exhibited behavior within similar ranges and that all variables presented nonzero standard deviations; consequently, no evidence of invalidity was identified during the data visualization phase.

Subsequently, a data treatment protocol was implemented, consisting solely of outlier detection, given that no variable contained missing values and, therefore, no imputation procedures were required. For data treatment, Mahalanobis distances were employed, which allow estimation of the distance of each multivariate observation in the dataset relative to the data centroid. As reported in the literature [41,54,55,63], these distances follow a multivariate distribution that approximates the χ^2 distribution; accordingly, 79 degrees of freedom were used for a 99.9% probability region of distances considered as inliers, while the remaining 0.1% falling outside the distribution were classified as outliers. In this manner, a cutoff score of 124.8392 was established, through which 259 atypical observations were identified and subsequently removed to mitigate the effects of bias and noise that these could generate in the analysis. As a result, the final dataset comprised 1,411 observations. The histogram and the distribution of Mahalanobis distances for the dataset are displayed in Figure 1.

Once the dataset had been processed and atypical observations removed, the next step was to verify whether the database met the parametric assumptions required to assess the instrument's

doi:10.20944/preprints202603.0330.v1

validity and reliability through Confirmatory Factor Analysis (CFA). To this end, the assumptions of additivity, normality, homogeneity, and homoscedasticity were examined. For the verification of the additivity assumption, the multivariate correlation matrix of all items composing the survey was obtained, where each cell represents the bivariate correlation of every possible pair of items. The multivariate correlation matrix for the designed survey is presented in Figure 2.

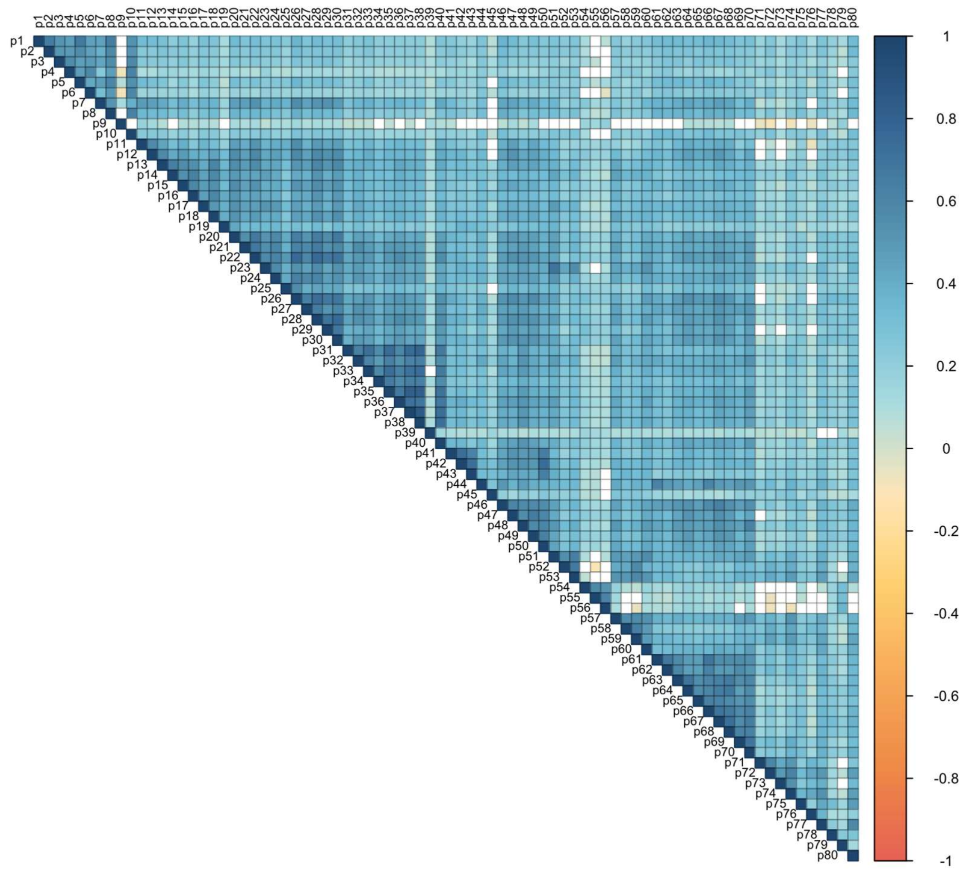
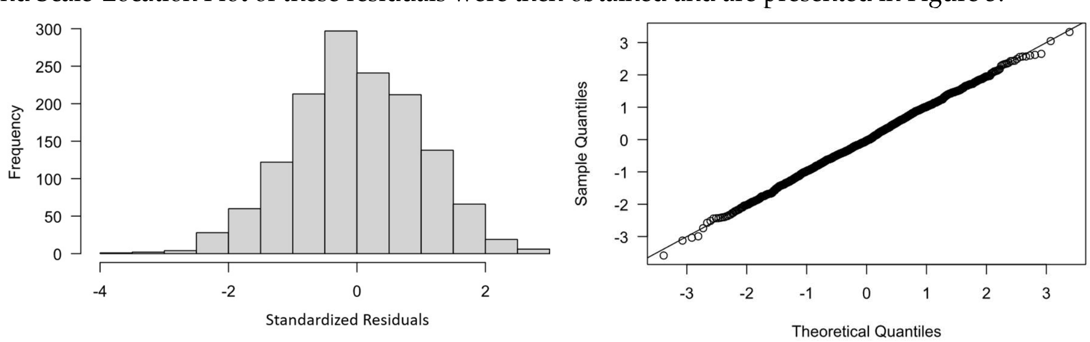


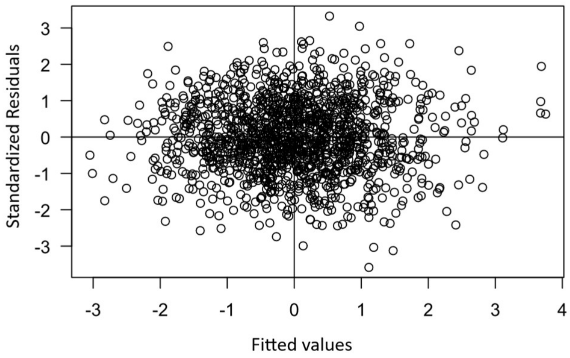
Figure 2. Bivariate correlation matrix for each pair of questions in the survey administered to students.

As can be observed in Figure 2, none of the possible variable pairs reached a perfect correlation within the range of 0.99 to 1.00; in fact, all bivariate correlations reported values below 0.872, thus the additivity assumption was accepted. Subsequently, to verify the assumptions of normality, homoscedasticity, and linearity of the predictors, a spurious regression analysis was employed, in which a set of random quantiles based on the normal distribution was generated and used as the response variable in a linear regression model that included all 80 survey variables as predictors, from which the standardized residuals of this regression were extracted. The Histogram, Q–Q Plot, and Scale-Location Plot of these residuals were then obtained and are presented in Figure 3.



 doi:10.20944/preprints202603.0330.v1

(a) (b)



(c)

Figure 3. Results of the false regression analysis: a) histogram, b) Q-Q Plot, c) Scale-Location plot.

To verify the suitability of the multivariate data structure, a random regression analysis was conducted, employing a Gaussian white noise vector as the dependent variable and the survey items as predictors. As shown in Figure 3a, the histogram of the studentized residuals exhibited a symmetric and mesokurtic pattern, consistent with a normal distribution. This finding was corroborated by the Q–Q (Quantile–Quantile) plot displayed in Figure 3b, where the residuals aligned along the theoretical 45-degree diagonal, providing robust evidence to support the assumption of approximate multivariate normality. Finally, the assumptions of linearity and homoscedasticity were examined through visual inspection of the scatter plot of fitted values versus studentized residuals (Scale-Location plot). In accordance with the criterion of residual randomness, the point cloud exhibited no systematic patterns, curvatures, or funnel-shaped formations (heteroscedasticity), distributing uniformly around the zero baseline. Consequently, the conditions of error variance homogeneity and predictor linearity were considered to be satisfied.

Once it was verified that the sample met all of the required assumptions, a Confirmatory Factor Analysis was performed using the R statistical programming language through the RStudio interface, employing the *lavaan*, *semplot* and *semtools* functions. The results of this analysis are presented in Figure 4 and Table 2.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 March 2026 doi:10.20944/preprints202603.0330.v1

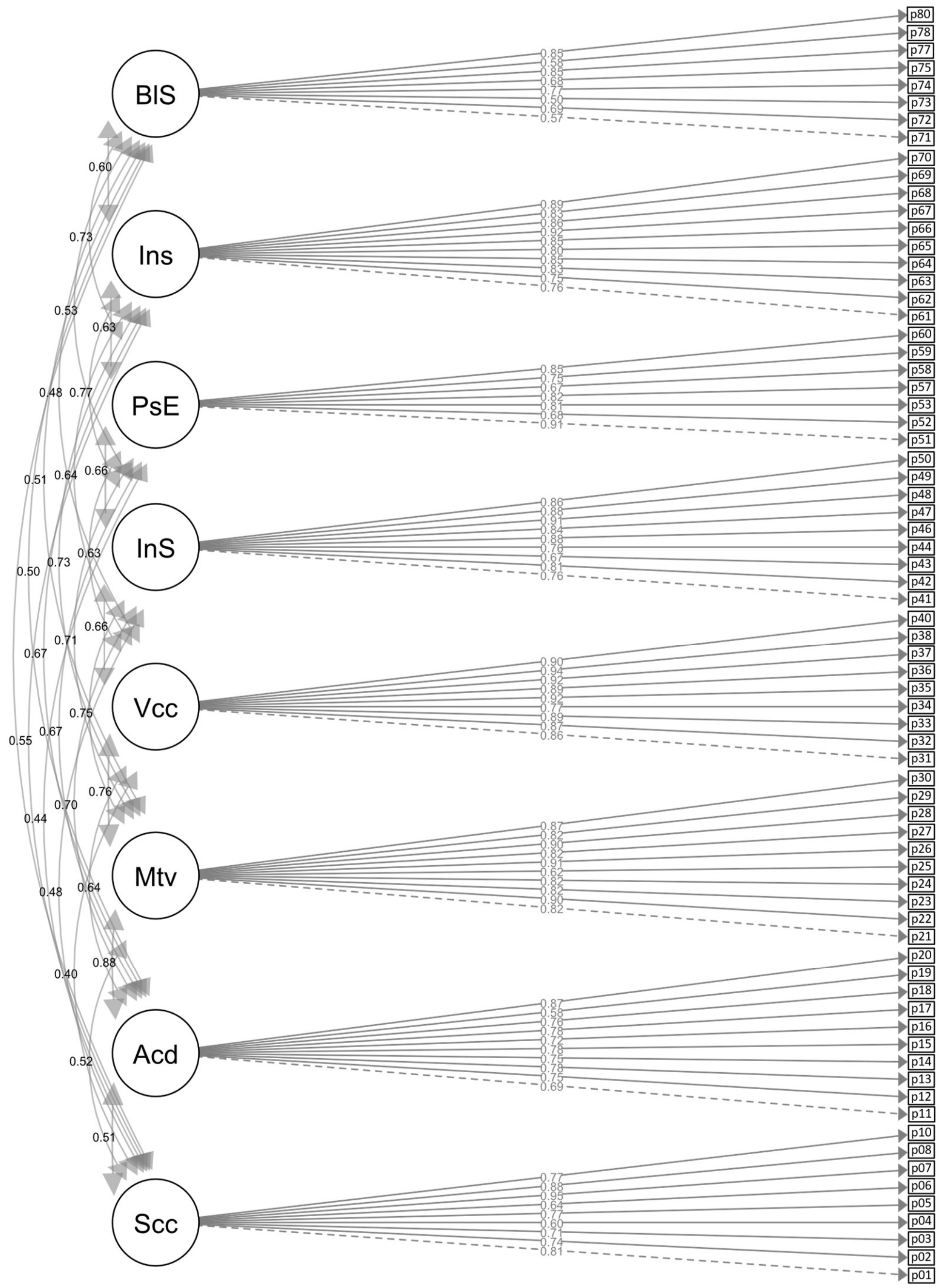


Figure 4. Path diagram, factor structure, and saturations per question for Confirmatory Factor Analysis.

Table 2. Goodness-of-fit indices for the factor structure evaluated using ACF.

npar	*fmin*	*chisq*	*df*	*pvalue*
532.000	8.147	22989.770	2456.000	2.2e-16
chisq.scaled	***df.scaled***	***pvalue.scaled***	***chisq.scaling.fact***	***baseline.chisq***
21387.989	2456.000	0.000	1.156	1506058.200
cfi	***tli***	***nnfi***	***rmsea***	***srmr***
0.986	0.986	0.986	0.057	0.061

The analysis of the measurement model, represented in the path diagram of Figure 4, enabled the examination of the standardized factor loadings (λ) of the proposed items. Following the initial inspection, the instrument was refined by excluding items p_9, p_{39}, p_{45}, p_{54}, p_{55}, p_{56}, p_{76}, and p_{79},

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 March 2026 doi:10.20944/preprints202603.0330.v1

as they exhibited loadings below the .50 criterion. The final version of the instrument comprised 72 items, in which all retained indicators displayed satisfactory factor loadings (above .50) toward their respective factor. Furthermore, the correlations among the latent variables did not approach unity, thereby ruling out discriminant validity concerns and confirming the proposed factorial structure. Regarding the overall goodness of fit of the Confirmatory Factor Analysis (CFA), the obtained indices corroborated an excellent fit of the model to the empirical data. Specifically, both the Comparative Fit Index (CFI) and the Tucker–Lewis Index (TLI) reached values above .95, exceeding the threshold required to be categorized as indicators of optimal fit. Convergently, a value of .057 was obtained for the approximation error (SRMR/RMSEA)*, comfortably satisfying the reference criterion of being below .08. Finally, the psychometric properties detailed in Table 3 confirmed the robustness of the scale. In terms of reliability, the ordinal alpha for each factor surpassed .90, while McDonald's omega coefficient (ω) was consistently above .84, exceeding .90 in nearly all cases, which provides evidence of outstanding internal consistency. With respect to convergent validity, the Average Variance Extracted (AVE) analysis indicated that all constructs attained indices above .50, thus verifying that the variance accounted for by the factors was greater than the error variance. On the basis of these findings, it was concluded that the designed survey possesses satisfactory metric properties of validity and reliability.

Table 3. Internal consistency indicators (Alpha and Omega) and Average Variance Extracted (AVE) of unobserved latent variables.

Factor	Socioeconomic	Academic	Motivational	Vocational	Social Integration	Psychological Emotional	Institutional	Biological and Health
Alpha ordinal	0.9202	0.9201	0.9548	0.9659	0.9386	0.9120	0.9554	0.8531
McDonald's Omega ω	0.9123	0.9051	0.9397	0.9601	0.9322	0.9030	0.9448	0.8401
Average Variance Extracted - AVE	0.5942	0.5605	0.6937	0.7837	0.6783	0.6239	0.6965	0.5722

For the estimation of factor scores $\hat{\eta}_B$ Bartlett's (1937) method was implemented, adapted to the continuous latent response variable framework proposed by Muthén (1984) [64,65]. Given the ordinal nature of the indicators, it is assumed that each observed item discretizes an underlying continuous variable. Formally, this measurement model is defined as:

$$y^* = \nu + \Lambda\eta + \epsilon \qquad (6)$$

where y^* represents the $(p \times 1)$ vector of continuous latent responses for the $p = 80$ items, ν is the intercept vector, Λ is the estimated factor loading matrix of dimension $(p \times m)$ for the $m = 8$ factors, and ϵ constitutes the residual vector (uniquenesses). Under this framework, Bartlett factor scores are derived by minimizing the variance of the unique factor errors subject to the constraint of asymptotic unbiasedness ($\mathbb{E}[\hat{\eta}_B] = \eta$). The matrix-based analytical solution governing this computation is:

$$\hat{\eta}_B = \left(\hat{\Lambda}^\top \hat{\Theta}^{-1} \hat{\Lambda}\right)^{-1} \hat{\Lambda}^\top \hat{\Theta}^{-1} (\hat{y}^* - \hat{\nu}) \qquad (7)$$

In this expression, $\widehat{\mathbf{\Theta}}$ denotes the $(p \times p)$ diagonal matrix composed of the unique error variances ($Var(\boldsymbol{\epsilon})$), which weights the precision of each item. The term $\left(\widehat{\mathbf{\Lambda}}^\top \widehat{\mathbf{\Theta}}^{-1} \widehat{\mathbf{\Lambda}}\right)^{-1}$ acts as the normalization factor that guarantees the unbiasedness property of the estimator. Finally, a critical component is the vector $\hat{\boldsymbol{y}}^*$, whose values are inferred from the polychoric correlation matrix and the set of estimated thresholds $\boldsymbol{\tau}$ through the probit link function, thereby ensuring that the score

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 March 2026 doi:10.20944/preprints202603.0330.v1

estimation maximizes construct validity while respecting the nonlinear nature of the original ordinal data.

Operationally, the estimated factor score for each dimension is obtained through a weighted linear combination of the latent response variables (y^*), where the weights (ω) reflect the net contribution of each item adjusted for its measurement error. The resulting equations for the eight factors of the model are presented below:

$$\hat{\eta}_{Soc} = 0.090(y^*_{p1}) + 0.062(y^*_{p2}) + 0.053(y^*_{p3}) + 0.035(y^*_{p4}) + 0.070(y^*_{p5}) + 0.041(y^*_{p6}) + 0.394(y^*_{p7}) + 0.144(y^*_{p8}) + 0.069(y^*_{p10}) \tag{8}$$

$$\hat{\eta}_{Acad} = 0.066(y^*_{p11}) + 0.085(y^*_{p12}) + 0.097(y^*_{p13}) + 0.084(y^*_{p14}) + 0.100(y^*_{p15}) + 0.072(y^*_{p16}) + 0.097(y^*_{p17}) + 0.088(y^*_{p18}) + 0.043(y^*_{p19}) + 0.172(y^*_{p20}) \tag{9}$$

$$\hat{\eta}_{Mot} = 0.075(y^*_{p21}) + 0.143(y^*_{p22}) + 0.073(y^*_{p23}) + 0.073(y^*_{p24}) + 0.030(y^*_{p25}) + 0.161(y^*_{p26}) + 0.076(y^*_{p27}) + 0.141(y^*_{p28}) + 0.075(y^*_{p29}) + 0.108(y^*_{p30}) \tag{10}$$

$$\hat{\eta}_{Voc} = 0.077(y^*_{p31}) + 0.084(y^*_{p32}) + 0.096(y^*_{p33}) + 0.044(y^*_{p34}) + 0.131(y^*_{p35}) + 0.098(y^*_{p36}) + 0.145(y^*_{p37}) + 0.174(y^*_{p38}) + 0.110(y^*_{p40}) \tag{11}$$

$$\hat{\eta}_{IntSoc} = 0.060(y^*_{p41}) + 0.081(y^*_{p42}) + 0.042(y^*_{p43}) + 0.060(y^*_{p44}) + 0.129(y^*_{p46}) + 0.098(y^*_{p47}) + 0.180(y^*_{p48}) + 0.133(y^*_{p49}) + 0.113(y^*_{p50}) \tag{12}$$

$$\hat{\eta}_{Psico} = 0.133(y^*_{p51}) + 0.036(y^*_{p52}) + 0.066(y^*_{p53}) + 0.073(y^*_{p57}) + 0.035(y^*_{p58}) + 0.051(y^*_{p59}) + 0.089(y^*_{p60}) \tag{13}$$

$$\hat{\eta}_{Inst} = 0.066(y^*_{p61}) + 0.058(y^*_{p62}) + 0.092(y^*_{p63}) + 0.107(y^*_{p64}) + 0.076(y^*_{p65}) + 0.103(y^*_{p66}) + 0.204(y^*_{p67}) + 0.120(y^*_{p68}) + 0.091(y^*_{p69}) + 0.151(y^*_{p70}) \tag{14}$$

$$\hat{\eta}_{Bio} = 0.027(y^*_{p71}) + 0.073(y^*_{p72}) + 0.037(y^*_{p73}) + 0.108(y^*_{p74}) + 0.071(y^*_{p75}) + 0.168(y^*_{p77}) + 0.048(y^*_{p78}) + 0.167(y^*_{p80}) \tag{15}$$

3.2. Structured Test Analysis

The diagnostic assessment was operationalized through the Moodle Learning Management System (LMS) platform integrated with the H5P interactive framework, ensuring a standardized and controlled administration environment. The instrument consisted of 80 multiple-choice items with a single correct response, designed using a table of specifications that ensured content validity and curricular alignment. The test structure equally weighted the domains of Fundamentals of Mathematics, Language and Communication, and Geometry ($k = 20$ items per domain), while the areas of Physics and Chemistry comprised the remaining block ($k = 10$ items each). Regarding item provenance, 73.8% ($n = 59$) were grounded in the Learning Standards of the Unified General Baccalaureate (*Bachillerato General Unificado*), supplemented by 26.2% ($n = 21$) of criteria specific to the Escuela Politécnica Nacional, thereby ensuring both state-level and institutional relevance.

Accordingly, this diagnostic instrument was administered to 1,676 newly admitted students at the institution, originating from various provinces and regions of Ecuador, who constituted the analytical population of the 2025-A career-leveling semester. Data collection was automated in real time through the virtual learning environment interface, enabling the direct consolidation of academic records without manual intermediation. Structurally, the database was organized as a response matrix of dimension $N \times k$ (1.676×80), where each row vector represented the performance pattern of an individual student. For purposes of psychometric modeling, the original polytomous responses were recoded under a dichotomous scheme ($x_{ij} \in \{0,1\}$), assigning a value of 1 to correct responses and 0 to errors or distractor selections. Prior to inferential analysis, this dataset underwent a rigorous data treatment protocol to ensure the integrity of the observations and the

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 March 2026 doi:10.20944/preprints202603.0330.v1

validity of the estimated parameters. Data collection via Moodle allowed responses to each item to be configured as mandatory, resulting in no blank cells in the database; consequently, no imputation techniques were required. For the detection of atypical observations, as in the preceding section, Mahalanobis distances were applied using a χ^2 distribution statistic that considered 99.9% of the observations as inliers for 79 degrees of freedom, thereby classifying the 0.1% of the most distant observations as outliers. In this manner, a cutoff score of 124.8392 was established, through which 7 atypical observations were identified, so that the final database comprised 1,669 observations. The descriptive statistics of the original student scores prior to instrument validation and calibration are presented in Table 4.

Table 4. Descriptive statistics of the scores achieved by students prior to calibration of the instrument.

Min	1st Qu.	Median	Mean	3rd Qu.	Max	St. Dev.
0.0	23.0	31.0	32.3	41.0	76.0	12.9263

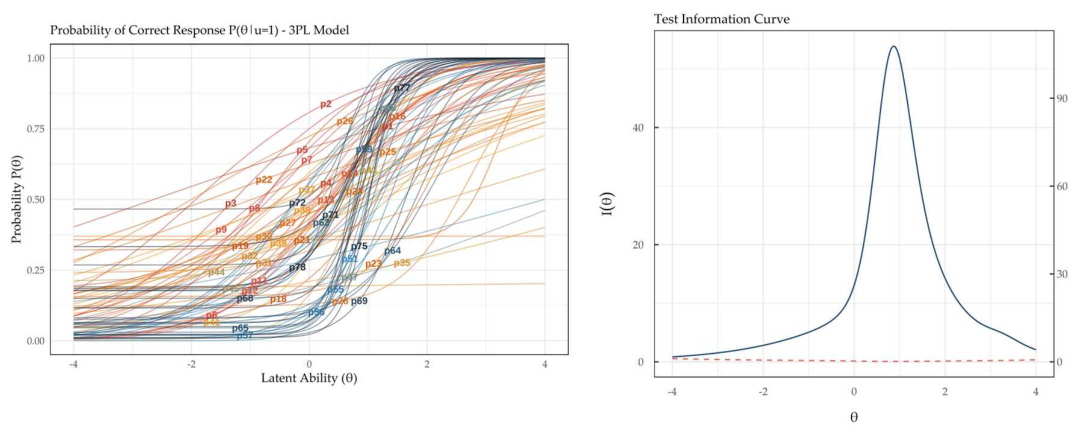
Subsequently, the psychometric calibration process of the instrument was carried out within the framework of Item Response Theory (IRT). In contrast to restrictive single-parameter models, the Three-Parameter Logistic Model (3PL) was selected due to the nature of the multiple-choice items, which introduce the possibility of correct responses by chance. Estimation was performed under a unidimensionality assumption using the mirt statistical package in the R statistical programming language, which allowed modeling the probability that a student with a given level of latent ability (θ) would respond correctly to item i. This mathematical model incorporates three fundamental parameters: discrimination (a_i), difficulty (b_i), pseudo-guessing (c_i), formalized in the seminal works [59,60].

The conditional probability function $P_i(\theta)$ that defines the Item Characteristic Curves (ICCs) is expressed by the following equation:

$$P_i(\theta) = c_i + (1 - c_i)\frac{1}{1 + e^{-a_i(\theta - b_i)}} \tag{16}$$

where $P_i(\theta)$represents the probability that an individual with ability θresponds correctly to item i. The parameter b_i denotes item difficulty, indicating the point on the ability scale at which the probability of a correct response equals 50% (adjusted for guessing). The parameter $a_i$corresponds to item discrimination, proportional to the slope of the curve at its inflection point, reflecting the item's capacity to differentiate among examinees with similar ability levels. The parameter c_i represents pseudo-guessing, constituting the lower asymptote of the curve and reflecting the probability that students with low ability levels answer the item correctly purely by chance.

Through this iterative marginal maximum likelihood estimation procedure, Item Characteristic Curves were generated for the entire item bank, allowing visualization of the functional behavior of each item along the latent ability continuum, as presented in Figure 5.



(a) (b)

Figure 5. Results of the TRI-3PL model fit: a) characteristic curves for each item, b) test information curves.

Beyond individual item calibration, the psychometric analysis assessed instrument reliability at the systemic level using the Test Information Function (TIF). Unlike deterministic approaches from Classical Test Theory (e.g., Cronbach's alpha), which assume a constant measurement error, the IRT framework conceives of precision as a conditional function dependent on the ability level θ. The analytical processing aggregated the local information functions of the $n = 80$ items, modeling the cumulative precision through the additive relationship:

$$I(\theta) = \sum_{i=1}^{n} I_i(\theta) = \sum_{i=1}^{n} \frac{[P_i'(\theta)]^2}{P_i(\theta)Q_i(\theta)} \tag{17}$$

where $I_i(\theta)$quantifies the informational contribution of item i, $P_i'(\theta)$is the first derivative of the logistic response function, and $Q_i(\theta) = 1 - P_i(\theta)$. Given that the test architecture was based on the 3PL model, the contribution of each item to the final score was not uniform; it was positively weighted by its discriminative capacity (a_i) and penalized in the lower ability ranges by the pseudo-guessing asymptote (c_i). This inverse relationship with metric uncertainty was formalized through the Asymptotic Standard Error of Estimation:

$$SE(\theta) = \frac{1}{\sqrt{I(\theta)}} \tag{18}$$

The graphical projections in Figure 5b revealed that the topology of the information curve reached its global maximum within the interval $-0.5 < \theta < +1.5$. This finding confirms that the instrument maximizes its discriminative power and minimizes the standard error for students with average and above-average ability levels. Conversely, the decreasing information in the lower tail ($\theta < -2.0$) evidences the intrinsic difficulty of precisely estimating actual knowledge in the presence of high probabilities of correct responses by chance among students with severe academic lag. Simultaneously with the evaluation of overall precision, the individual goodness of fit of the items was examined using the $S\text{-}X^2$ statistic developed by Orlando and Thissen (2000). This metric quantifies the discrepancy between observed response frequencies and the theoretical probabilities estimated by the 3PL model, conditioned on the observed total score. A conservative statistical significance criterion ($p < .01$) was established to identify those items whose empirical behavior deviated substantially from the model's predictions. The detection of these misfits proved critical for metric quality assurance, flagging those items with functional anomalies or structural ambiguities that required removal or recalibration to ensure construct validity in future applications of the instrument. Applying this methodology to the 80 items in the database yielded the results presented in Table 5.

Table 5. Discrimination, difficulty, pseudo-fortune telling, and Orlando Thisen's test.

Item	Discrimination (a_i)	Difficulty (b_i)	Guessing (c_i)	Statistic $S\text{-}X^2$	P-value $P_i(> S\text{-}X^2)$
		Difficulty level – Very easy			
p_7	0.6770	-2.4527	0.0007	70.9686	0.1179
p_{41}	0.8795	-2.1547	0.0006	52.5564	0.4914
p_1	0.6184	-1.8487	0.2220	139.3602	0.0000
p_{17}	0.9572	-1.5196	0.0003	69.5142	0.0901
		Difficulty level – Easy			
p_{47}	1.1518	-1.4854	0.0002	57.3316	0.2841
p_{12}	1.1537	-1.3527	0.0005	60.0155	0.1814
p_{21}	0.2548	-1.3379	0.0020	83.8726	0.0405
p_4	0.6196	-1.2860	0.0007	76.4155	0.0750
p_{35}	0.4811	-1.2814	0.0012	81.9464	0.0458

p_{15}	0.9383	-1.1539	0.0004	49.2546	0.6928
p_{3}	0.6976	-1.1345	0.0040	110.2665	0.0001
p_{13}	1.0103	-1.1114	0.0002	69.9504	0.0844
p_{60}	1.4031	-1.1019	0.0002	60.4925	0.1257
p_{49}	1.4153	-1.0952	0.0007	56.6408	0.2114
p_{27}	0.4680	-1.0243	0.0020	102.0723	0.0010
p_{6}	0.9429	-1.0079	0.0002	61.0016	0.3009
p_{46}	0.5515	-0.8600	0.0004	73.6509	0.1477
p_{51}	1.1064	-0.7687	0.0001	97.1959	0.0002
p_{52}	1.7628	-0.7600	0.0002	37.2595	0.7539
p_{56}	1.3841	-0.6857	0.0153	55.9331	0.2949
p_{36}	0.4675	-0.6214	0.0068	69.0824	0.2506
p_{18}	0.9633	-0.6029	0.0015	68.9171	0.1152
p_{22}	0.5870	-0.6015	0.0022	51.4544	0.8031
p_{24}	0.4182	-0.5277	0.0041	76.6768	0.0994
Difficulty level – Moderate					
p_{53}	1.7072	-0.4421	0.0030	91.3920	0.0002
p_{50}	0.9411	-0.3871	0.0001	59.5068	0.3492
p_{8}	1.0931	-0.3708	0.0002	67.1619	0.1259
p_{20}	1.4874	-0.3124	0.0372	42.0925	0.8351
p_{11}	1.1771	-0.2695	0.0242	69.7623	0.0868
p_{48}	1.2820	-0.1031	0.0594	71.6275	0.0654
p_{10}	0.9876	-0.0473	0.0002	69.7171	0.0874
p_{14}	1.4875	-0.0470	0.1979	63.7956	0.2214
p_{42}	1.0673	0.0047	0.0001	45.8724	0.8049
p_{57}	1.2356	0.0133	0.0001	65.3338	0.1388
p_{9}	0.9904	0.0250	0.0001	57.6290	0.3782
p_{58}	1.7317	0.0314	0.0551	57.8482	0.3010
p_{19}	1.5592	0.1356	0.1268	39.0912	0.9483
p_{29}	0.6153	0.1457	0.0477	85.8009	0.0161
p_{62}	3.2127	0.1598	0.1431	81.3521	0.0019
p_{44}	1.0353	0.2902	0.0402	69.6113	0.1220
p_{71}	7.5761	0.3551	0.6188	81.7812	0.0009
p_{63}	3.9246	0.3699	0.0618	100.8064	0.0000
p_{30}	0.6165	0.3732	0.0005	65.8620	0.2516
p_{72}	4.0549	0.4008	0.5701	71.4545	0.0380
p_{43}	2.3558	0.4034	0.0001	49.5503	0.2283
p_{16}	0.9300	0.4065	0.0016	77.9640	0.0340
p_{65}	3.9605	0.4083	0.0911	75.6368	0.0029
p_{77}	3.6370	0.4514	0.4023	63.6131	0.1297
Difficulty level – Difficult					
p_{54}	2.7102	0.5089	0.0212	50.9309	0.2196
p_{73}	4.1959	0.5474	0.3296	67.5053	0.0728
p_{79}	3.7975	0.5586	0.3534	78.3584	0.0168
p_{66}	3.2381	0.5729	0.0350	128.3997	0.0000
p_{74}	6.0325	0.5825	0.2173	140.8328	0.0000
p_{59}	2.3863	0.5973	0.0000	64.8601	0.0102
p_{75}	6.3863	0.6035	0.2473	143.1152	0.0000
p_{61}	2.7557	0.6349	0.0431	80.5748	0.0017
p_{26}	0.7739	0.6547	0.1633	93.5642	0.0046
p_{78}	3.9679	0.6666	0.2444	88.6043	0.0021
p_{80}	3.7122	0.6867	0.3598	96.1831	0.0005
p_{55}	2.2923	0.6938	0.0369	63.1391	0.0844

p_{70}	4.6685	0.7327	0.0173	186.7704	0.0000
p_{68}	3.2927	0.7670	0.1583	71.7355	0.0535
p_{76}	3.4202	0.8443	0.1772	64.6292	0.1525
p_{45}	0.3638	0.9178	0.0016	58.9932	0.5125
p_{67}	2.0804	1.0667	0.1095	68.3669	0.1242
p_{38}	0.8340	1.0917	0.2057	68.0968	0.2486
p_{23}	1.8228	1.2057	0.2239	52.2064	0.7528
p_{2}	0.8339	1.2170	0.0236	60.5326	0.3157
p_{39}	6.5611	1.2897	0.3308	86.9314	0.0105
p_{37}	20.6236	1.2928	0.4118	80.7999	0.0256
p_{33}	4.9687	1.3105	0.4147	64.3039	0.2962
p_{69}	2.0448	1.3347	0.0127	61.0462	0.0556
p_{34}	12.7702	1.3409	0.3512	93.0390	0.0024
p_{28}	0.6799	1.3581	0.2108	81.4841	0.0410
p_{32}	14.4308	1.3802	0.3347	92.2266	0.0037
p_{5}	0.5401	1.3816	0.0003	47.5422	0.8348
p_{31}	1.0022	1.4164	0.1728	76.8729	0.0701
		Difficulty level – Very difficult			
p_{64}	1.0546	1.7959	0.0665	45.2997	0.8461
p_{40}	0.5293	2.5587	0.2655	55.0339	0.7225
p_{25}	3.2882	2.8101	0.3314	52.4748	0.8252

Although the psychometric inspection considered multiple indicators of technical quality—such as heuristic discrimination thresholds ($a_i < 0.65$) or upper bounds for pseudo-guessing ($c_i > 0.35$)—the critical decision regarding the internal validity of the items was strictly grounded in statistical goodness of fit. Unlike individual parameters, which describe the functional properties of an item, the $S\text{-}X^2$ statistic proposed by Orlando and Thissen (2000) verifies the structural congruence between the empirical data and the theoretical predictions of the model. Consequently, this probabilistic criterion was prioritized over conventional rules of thumb, establishing the significance level $p < .01$as the determining filter. This approach ensures that items flagged for review are not merely those with low marginal information, but rather those that violate the mathematical assumptions of the 3PL model, thereby compromising the stability of ability estimates (θ). In this manner, and as can be observed in Table 5, the items identified as exhibiting fit problems that were removed for the computation of weighted scores for each student on the administered diagnostic test were items 70, 75, 74, 66, 1, 63, 3, 53, 51, 80, 71, 27, 61, 62, 78, 34, 65, 32, and 26.

Finally, for the generation of individual weighted scores, the analysis moved beyond the simple summation of correct responses (raw score). The observed dichotomous response vector for each student $u = (u_1, u_2, \ldots, u_n)$ was transformed into a continuous estimator $\hat{\theta}$ using the Bayesian Expected A Posteriori (EAP) method. This algorithm computes the mean of the posterior distribution of ability, weighting the likelihood of the response pattern $L(u \mid \theta)$ with a normative prior distribution $\phi(\theta) \sim N(0,1)$. The estimate was obtained through numerical integration of the following expression:

$$\hat{\theta}_{EAP} = \frac{\int_{-\infty}^{+\infty} \theta L(u|\theta)\phi(\theta)d\theta}{\int_{-\infty}^{+\infty} L(u|\theta)\phi(\theta)d\theta} \tag{19}$$

Through this mechanism, the score assigned to each student (e.g., $\theta = 1.52$or $\theta = -0.31$) depends not only on the number of correct responses but also on the psychometric quality of the items answered correctly. A correct response on an item with high discrimination and difficulty contributes a greater marginal "gain" to the estimation of θ than a correct response on an easy item or one susceptible to guessing. This approach minimized the Mean Squared Error (MSE) of the estimates, providing a robust and calibrated metric for subsequent hypothesis testing.

 doi:10.20944/preprints202603.0330.v1

The visual inspection of the distributions (Figure 6), contrasted with the descriptive statistics, revealed significant metric differences between the classical score and the calibrated estimate, thereby rejecting the hypothesis of direct equivalence between the two methods. Unlike the linear summation, the IRT calibration evidenced a substantial positive shift in the measures of central tendency, reporting a higher overall mean ($\bar{x}_{IRT} = 11.50;\ SD = 3.49$) compared to the raw score ($\bar{x}_{Raw} = 10.56;\ SD = 3.78$). This divergence was statistically corroborated through the Wilcoxon rank-sum test ($W = 1{,}641{,}640,\ p < 2.2 \times 10^{-16}$), which confirmed a significant shift in the location of the distributions. Specifically, the increase in the median of the Bayesian distribution ($Mdn = 11.9$ vs. $Mdn = 10.5$), coupled with a pronounced negative skewness (-0.64), suggests that the 3PL model effectively corrected the underestimation of ability among students who answered highly discriminating items correctly, overcoming the limitations of the linear scale. Consequently, the application of IRT proved indispensable for ensuring psychometric equity: it transformed a measure influenced by the mere accumulation of correct responses into a weighted estimate of actual competence, providing superior granularity for student classification.

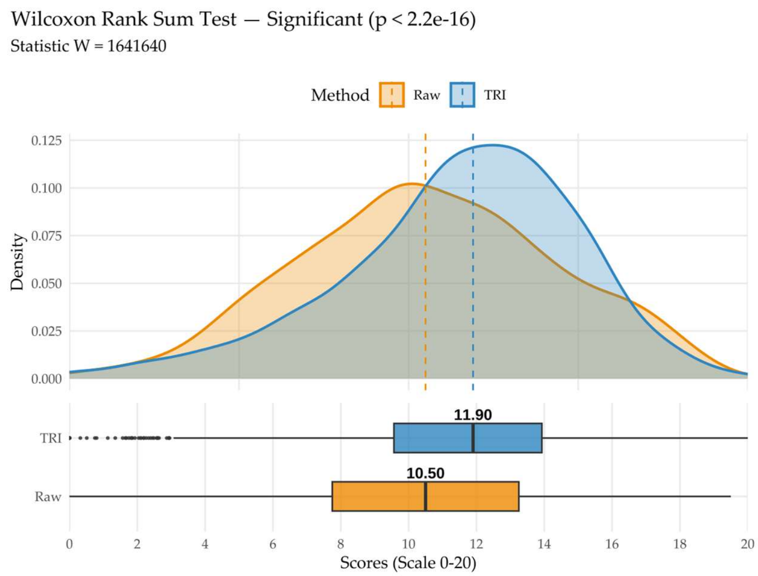


Figure 6. Distributional density comparison and box plots between observed raw scores and ability estimates calibrated using the TRI-3PL model.

4. Discussion

The empirical findings substantiated the effectiveness of integrating Confirmatory Factor Analysis (CFA) with three-parameter logistic Item Response Theory (3PL-IRT) models for comprehensive assessment of newly admitted university students. This dual-component framework yielded psychometrically robust instruments capable of capturing both latent non-cognitive constructs and calibrated academic competency estimates, enabling multidimensional characterization of student entry profiles that transcended conventional assessment limitations.

The factorial survey demonstrated exemplary psychometric properties, with CFA fit indices substantially exceeding conventional thresholds ($CFI = 0.986$, $TLI = 0.986$, $RMSEA = 0.057$, $SRMR = 0.061$), confirming that the eight-factor structure adequately reproduced observed covariance patterns among the 72 retained items. Internal consistency coefficients surpassed stringent reliability standards, with ordinal alpha values exceeding 0.85 and McDonald's omega ranging from 0.84 to 0.96 across all constructs. The Average Variance Extracted indices uniformly exceeded 0.50, establishing convergent validity for the proposed dimensional structure. The implementation of Bartlett factorial score estimation, adapted for ordinal variables through Muthén's latent continuous

 doi:10.20944/preprints202603.0330.v1

response framework, constituted a methodological advancement by weighting each item's contribution according to its error variance, yielding unbiased estimators that maximized construct validity while respecting non-linear Likert-type data characteristics.

Concurrently, 3PL-IRT calibration revealed substantial heterogeneity in item psychometric properties across academic domains. Discrimination parameters ($\bar{a}$ = 1.92) indicated adequate differentiation capacity, while difficulty parameters exhibited appropriate centering ($\bar{b}$ = 0.05). However, S-X^2 statistics identified 19 items (23.75%) with significant misfit ($p < 0.01$), concentrated in physics and chemistry domains, suggesting potential multidimensionality warranting revision. The Test Information Function peaked within $-0.5 < \theta < +1.5$, indicating optimal calibration for students near institutional proficiency thresholds while identifying measurement gaps at lower ability tails. Expected A Posteriori estimation yielded ability scores diverging significantly from raw distributions (W = 1,641,640, $p < 2.2e-16$), with calibrated scores exhibiting higher central tendency ($\bar{M}$_TRI = 11.50 vs. $\bar{M}$_Raw = 10.56) and pronounced negative skewness (−0.64), demonstrating effective correction of underestimation bias inherent to sum-score methods.

The principal methodological contribution resided in integrating CFA-derived factorial scores with IRT-calibrated ability estimates within a unified analytical framework. This synergistic approach enabled systematic cross-tabulation of nine continuous outcome variables against 24 categorical demographic indicators, facilitating precise identification of differential performance patterns.

Table 6. Comparative analysis of methodological approaches for university student assessment.

Name	Technique	Advantages	Limitations
Present Study	3PL-IRT calibration with EAP estimation + CFA with Bartlett factorial scores	Comprehensive 8-factor non-cognitive profiling; pseudo-guessing correction for multiple-choice items; weighted factorial score extraction; distribution-free inference; LMS-integrated deployment; open-source reproducibility	Requires large samples for stable 3PL estimation; limited to structured assessment formats
Ntumi (2025)	MIRT + CFA + Rasch + Differential Item Functioning (Mantel-Haenszel)	Multidimensional ability modeling; adaptive testing efficiency (35% reduction); socioeconomic bias detection	Complex computational requirements; requires specialized software; reduced content coverage through adaptive algorithms
Ortiz-Rojas et al. (2019)	2PL-IRT + CFA + Structural Equation Modeling	Mediation hypothesis testing; integration with motivational constructs; established Ecuadorian context	Single academic domain (programming); omits pseudo-guessing parameter; limited non-cognitive scope (motivation only)
Aza-Espinosa et al. (2023)	Rasch models + ANOVA/t-tests	Parsimonious item weighting; direct socioeconomic variable integration; straightforward implementation	Equal discrimination assumption; parametric inference despite potential violations; lacks multidimensional non-cognitive assessment

The distinctive innovations comprised three elements: Bartlett factorial score extraction through latent continuous response modeling for ordinal indicators, providing error-weighted estimates

 doi:10.20944/preprints202603.0330.v1

maximizing construct validity; 3PL-IRT calibration incorporating pseudo-guessing parameters correcting random response patterns; and systematic integration within non-parametric frameworks enabling robust inference across categorical variables without distributional assumptions.

Building upon this validated infrastructure, subsequent research will develop predictive classification algorithms identifying students at elevated risk for course failure or institutional withdrawal. Machine learning architectures—including ensemble methods and neural network classifiers—will be trained on multidimensional feature vectors comprising IRT-calibrated estimates, CFA-derived factorial scores, and demographic indicators. The objective consists of constructing early warning systems enabling proactive intervention during initial academic weeks, translating diagnostic precision into actionable policies enhancing retention rates among vulnerable populations entering higher education.

5. Conclusions

This investigation accomplished its primary objective of developing and validating an integrated psychometric framework combining Confirmatory Factor Analysis with three-parameter logistic Item Response Theory models for the comprehensive assessment of newly admitted university students in a resource-constrained Latin American context. The dual-component methodology successfully operationalized the measurement of eight latent non-cognitive constructs alongside calibrated academic competency estimates across five foundational disciplines, thereby fulfilling the research aim of characterizing multidimensional factors associated with academic achievement during the university transition period.

The principal contributions of this research were threefold. First, the validated 72-item factorial instrument provided a reliable and construct-valid tool for assessing socioeconomic, academic, motivational, vocational, social integration, psychological/emotional, institutional, and biological/health dimensions among leveling program students. The psychometric evidence—comprising excellent model fit indices, robust internal consistency coefficients, and adequate convergent validity—established this survey as a dependable mechanism for institutional diagnostic purposes. Second, the 3PL-IRT calibration protocol demonstrated that weighted ability estimation via Expected A Posteriori methods yielded substantially more accurate competency measurements than conventional raw scoring approaches, correcting systematic biases introduced by item-level heterogeneity in discrimination and pseudo-guessing parameters.

In conclusion, this research contributed validated psychometric instruments, reproducible IRT-LMS integration protocols, and robust empirical evidence supporting the feasibility and added value of comprehensive diagnostic assessment for strengthening university transition outcomes. The findings substantiated that the synergistic combination of factor-analytic and item response theoretic approaches constituted a powerful methodology for understanding how students arrive at higher education institutions—their academic strengths, knowledge gaps, and associated risk factors—thereby enabling the design of precision interventions capable of enhancing retention and academic success among incoming student populations.

Funding: This work was supported by the National Polytechnic School - EPN of Ecuador (www.epn.edu.ec accessed on 28 May 2025).

Acknowledgments: We acknowledge the National Polytechnic School - EPN of Ecuador for their support in carrying out this research.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. J. Gudamud, I. Chiriboga, J. Zumba, R. Briceño, J. Jiménez, and A. Palma, "Innovations and trends in educational evaluation systems Resumen," *Redilat*, vol. 5, no. 3, pp. 1724–1733, 2024, doi: https://doi.org/10.56712/latam.v5i3.2157.

2. R. De Ayala, *The Theory and Practice of Item Response Theory.*, 2º. Guilford Press, 2022. doi: 10.1007/s11336-010-9179-z.
3. S. Embretson and S. Reise, *Item Response Theory for Psychologists*, 1st ed. Psychology Press, 2000. doi: 10.4324/9781410605269.
4. B. Villarreal, V. Maila, H. Figueroa, and E. Pérez, "Retos y logros de la aplicación de grupos interactivos en una comunidad de aprendizaje," *Cátedra*, vol. 4, no. 1, pp. 56–80, 2021, doi: 10.29166/CATEDRA.V4I1.2676.
5. J. Mero, "Herramientas digitales educativas y el aprendizaje significativo en los estudiantes," *Rev. Cient.*, vol. 7, no. 1, pp. 712–724, 2021.
6. B. Quilca, J. López, M. Guamán, E. Casagallo, and W. Briones, "EVALUACIÓN EDUCATIVA EN ENTORNOS VIRTUALES DE APRENDIZAJE EDUCATIONAL," *Rev. científica multidiciplinar*, vol. 8, no. 1, 2024, doi: https://doi.org/10.37811/cl_rcm.v8i1.9832.
7. Ministerio de educación, *Currículo de EGB y BGU*, 1st ed. Quito, 2016.
8. Ministerio de educación del Ecuador, "Estándares de aprendizaje," 2020.
9. M. Altamirano and G. Alarcón, "Importancia del semestre de nivelación en el ingreso a las universidades ecuatorianas," *Rev. Conrado*, pp. 362–368, 2020.
10. M. Zumárraga and G. Cevallos, "Autoeficacia, procrastinación y rendimiento académico en estudiantes universitarios de Ecuador," *Alteridad*, vol. 17, no. 2, pp. 277–290, 2022, doi: 10.17163/alt.v17n2.2022.08.
11. Secretaría Nacional de Planificación 2021, "Plan de Creación de Oportunidades 2021-2025," 2021.
12. S. Ntumi, "Advanced Multidimensional Item Response Theory Modeling for High-Stakes, Cross-Disciplinary Competency Assessments in Sub-Saharan Africa: A Psychometric Approach to Equity, Adaptivity, and Policy Integration," Apr. 11, 2025. doi: 10.21203/rs.3.rs-6418690/v1.
13. M. Ortiz-Rojas, K. Chiluiza, and M. Valcke, "Gamification through leaderboards: An empirical study in engineering education," *Comput. Appl. Eng. Educ.*, vol. 27, no. 4, pp. 777–788, Jul. 2019, doi: 10.1002/cae.12116.
14. K. A. Ryan, "AN INVESTIGATION OF PRE-SERVICE TEACHER ASSESSMENT LITERACY AND ASSESSMENT CONFIDENCE: MEASURE DEVELOPMENT AND EDTPA PERFORMANCE," Kent State University, 2018. [Online]. Available: https://etd.ohiolink.edu/acprod/odb_etd/ws/send_file/send?accession=kent1522746692960315&disposition=inline
15. N. Sugawara, N. Yasui-Furukori, M. Sayama, and K. Shimoda, "Item response theory analysis of the University Personality Inventory in medical students," *Neuropsychopharmacol. Reports*, vol. 43, no. 3, pp. 446–452, Sep. 2023, doi: https://doi.org/10.1002/npr2.12362.
16. A. Avcu, "Item Response Theory-Based Psychometric Investigation of SWLS for University Students," *Int. J. Psychol. Educ. Stud.*, vol. 8, no. 2, pp. 27–37, 2021, [Online]. Available: https://dergipark.org.tr/en/pub/pes/issue//935929
17. L. Appolloni, D. Valeri, and D. D'Alessandro, "Knowledge on Indoor Air Quality (K-IAQ): Development and Evaluation of a Questionnaire Through the Application of Item Response Theory," *Atmosphere (Basel).*, vol. 16, no. 10, p. 1163, Oct. 2025, doi: 10.3390/atmos16101163.
18. X. Wang, S. Zhang, and T. Xin, "Item Response Theory Analysis of the Dark Factor of Personality Scale for College Students in China," *Int. J. Environ. Res. Public Health*, vol. 19, no. 19, p. 12787, Oct. 2022, doi: 10.3390/ijerph191912787.
19. M. Aza-Espinosa, L. Guerra Torrealba, E. Herrera-Granda, M. Aza-Espinosa, M. Burbano-Pulles, and J. Pozo-Burgos, "Learning Performance Indicators a Statistical Analysis on the Subject of Natural Sciences During the COVID-19 Pandemic at the Tulcán District BT - Trends in Artificial Intelligence and Computer Engineering," M. Botto-Tobar, O. S. Gómez, R. Rosero Miranda, A. Díaz Cadena, and W. Luna-Encalada, Eds., Cham: Springer Nature Switzerland, 2023, pp. 139–154.
20. J. ADEGBUYI, "CONSTRUCTION AND USE OF MULTIDIMENSIONAL STUDENT MATHEMATICS ENGAGEMENT SCALE IN PREDICTING MATHEMATICS ACHIEVEMENT AMONG SENIOR SECONDARY SCHOOL STUDENTS IN EKITI STATE, NIGERIA," UNIVERSITY OF IBADAN, 2019. [Online]. Available: http://140.105.46.132:8080/xmlui/bitstream/handle/123456789/802/adegbuyi.pdf?sequence=1&isAllowed=y

 doi:10.20944/preprints202603.0330.v1

21. T. L. Welles, "An Analysis of the Academic Success Inventory for College Students: Construct Validity and Factor Scale Invariance," FLORIDA STATE UNIVERSITY, 2010. [Online]. Available: https://repository.lib.fsu.edu/islandora/object/fsu:175747/datastream/PDF/view
22. Z. Zhou, H. Yin, R. Y. Lam, and F. Lai, "Predicting students' academic performance: A comparative study using machine learning techniques," *Knowledge-Based Syst.*, vol. 163, pp. 16–27, 2019.
23. K. Buckley, S. Subedi, and S. Krachman, "Measurement Properties of Student Social-Emotional Competency and School Culture Climate Surveys in the NewSchools Invent Cohort," *Transform. Educ.*, pp. 1–44, 2018, [Online]. Available: https://eric.ed.gov/?id=ED605383
24. K. T. NGUYEN, T. M. DUONG, N. Y. TRAN, A. T. HA, and Y. N. T. PHUNG, "The Impact of Emotional Intelligence on Performance: A Closer Look at Individual and Environmental Factors," *J. Asian Financ. Econ. Bus.*, vol. 7, no. 1, pp. 183–193, Jan. 2020, doi: 10.13106/jafeb.2020.vol7.no1.183.
25. J. Zang, Y. Kim, and J. Dong, "New evidence on technological acceptance model in preschool education: Linking project-based learning (PBL), mental health, and semi-immersive virtual reality with learning performance," *Front. Public Heal.*, vol. 10, Sep. 2022, doi: 10.3389/fpubh.2022.964320.
26. R. C. Ho and H. K. Chua, "Bring-your-own-device learning environment: a platform for enhancing student's learning ability and innovativeness," *Int. J. Technol. Enhanc. Learn.*, vol. 7, no. 2, p. 178, 2015, doi: 10.1504/IJTEL.2015.072031.
27. F. Sáez-Delgado, J. Mella-Norambuena, and Y. López-Angulo, "Psychometric properties of the SocioEmotional Skills Instrument for Teachers using network approach: English and Spanish version," *Front. Psychol.*, vol. 15, Sep. 2024, doi: 10.3389/fpsyg.2024.1421164.
28. M. G. Wolf, "Validity and Validation: A Pragmatic Path Forward Wolf, Melissa Gordon," UNIVERSITY OF CALIFORNIA, 2022. [Online]. Available: https://escholarship.org/uc/item/3d80s5r2
29. D. Hewagallage, "Examining the Relations among Academic and Non-Cognitive Factors and Student Achievement," West Virginia University, 2023. [Online]. Available: https://par.nsf.gov/servlets/purl/10525601
30. K. Yeung, "PERCEPTION OF TEACHER EMOTIONAL SUPPORT AND PARENTAL EDUCATION LEVEL: THE IMPACTS ON STUDENTS' MATH PERFORMANCE," University of Leicester, 2007. [Online]. Available: https://figshare.le.ac.uk/articles/thesis/Perception_of_Teacher_Emotional_Support_and_Parental_Education_Level_The_Impacts_on_Students_Math_Performance/10098614?file=18204611
31. R. Fabbricatore, "Latent Class Analysis for proficiency assessment in Higher Education: Integrating multidimensional latent traits and learning topics," University of Naples Federico II, 2023. [Online]. Available: http://www.fedoa.unina.it/15056/1/fabbricatore_rosa_35.pdf
32. A. D'Agostino, F. Schirripa Spagnolo, and N. Salvati, "Studying the relationship between anxiety and school achievement: evidence from PISA data," *Stat. Methods Appt.*, vol. 31, no. 1, pp. 1–20, Mar. 2022, doi: 10.1007/s10260-021-00563-9.
33. M. D. Manzar et al., "Depression, Anxiety, and Stress Scale-21 (DASS-21): Further psychometric exploration using robust item response theory and classical theory measures among university students," *PLoS One*, vol. 20, no. 7, p. e0325238, Jul. 2025, [Online]. Available: https://doi.org/10.1371/journal.pone.0325238
34. I. García-Martínez, J. M. A. Landa, and S. P. León, "The Mediating Role of Engagement on the Achievement and Quality of Life of University Students," *Int. J. Environ. Res. Public Health*, vol. 18, no. 12, p. 6586, Jun. 2021, doi: 10.3390/ijerph18126586.
35. R. Juliá-Sanchis, M. J. Cabañero-Martínez, C. Leal-Costa, M. Fernández-Alcántara, and S. Escribano, "Psychometric Properties of the Health Professionals Communication Skills Scale in University Students of Health Sciences," *Int. J. Environ. Res. Public Health*, vol. 17, no. 20, p. 7565, Oct. 2020, doi: 10.3390/ijerph17207565.
36. I. Tsaousis and A. Al-Owidha, "Development of a Forced-Choice Personality Inventory via Thurstonian Item Response Theory (TIRT)," *Behav. Sci. (Basel).*, vol. 14, no. 12, p. 1118, Nov. 2024, doi: 10.3390/bs14121118.
37. C. Laranjeira, A. Querido, P. Sousa, and M. A. Dixe, "Assessment and Psychometric Properties of the 21-Item Depression Anxiety Stress Scale (DASS-21) among Portuguese Higher Education Students during the

 doi:10.20944/preprints202603.0330.v1

COVID-19 Pandemic," *Eur. J. Investig. Heal. Psychol. Educ.*, vol. 13, no. 11, pp. 2546–2560, Nov. 2023, doi: 10.3390/ejihpe13110177.

38. B. Paladines-Costa, V. López-Guerra, P. Ruisoto, S. Vaca-Gallegos, and R. Cacho, "Psychometric Properties and Factor Structure of the Spanish Version of the Acceptance and Action Questionnaire-II (AAQ-II) in Ecuador," *Int. J. Environ. Res. Public Health*, vol. 18, no. 6, p. 2944, Mar. 2021, doi: 10.3390/ijerph18062944.
39. G. T. Zewude, D. G. Bereded, E. Abera, G. Tegegne, S. Goraw, and T. Segon, "The Impact of Internet Addiction on Mental Health: Exploring the Mediating Effects of Positive Psychological Capital in University Students," *Adolescents*, vol. 4, no. 2, pp. 200–221, Apr. 2024, doi: 10.3390/adolescents4020014.
40. A. E. Jácome-Ortega, E. P. Herrera-Granda, I. D. Herrera-Granda, J. A. Caraguay-Procel, and A. V. Basantes-Andrade, "Análisis temporal y pronóstico del uso de las TIC, a partir del instrumento de evaluación docente de una Institución de Educación Superior," *Rev. Ibérica Sist. e Tecnol. Informação*, no. E22, pp. 399–412, 2019, [Online]. Available: https://www.proquest.com/openview/96910c7cb0c260ae2409940921c7f71b/1?pq-origsite=gscholar&cbl=1006393
41. E. P. Herrera-Granda, J. G. Loor-Bautista, and J. I. Mina-Ortega, "Incidence of Metaphorical Virtual Classrooms and Interactive Learning Objects in the Interaction of Online Students: An Ecuadorian Case Study," *Appl. Sci.*, vol. 14, no. 15, p. 6447, Jul. 2024, doi: 10.3390/app14156447.
42. A. E. Jácome Ortega, J. A. Caraguay Procel, E. P. Herrera-Granda, and I. D. Herrera Granda, "Confirmatory Factorial Analysis Applied on Teacher Evaluation Processes in Higher Education Institutions of Ecuador," 2020, pp. 157–170. doi: 10.1007/978-3-030-37221-7_14.
43. D. E. Imbaquingo, E. P. Herrera-Granda, I. D. Herrera-Granda, S. R. Arciniega, V. L. Guamán, and M. C. Ortega-Bustamante, "Evaluation of university informatic security systems: Teacher evaluation system a case study," *RISTI - Rev. Iber. Sist. e Tecnol. Inf.*, vol. 2019, no. E22, pp. 349–362, 2019.
44. C. P. Guevara-Vega, W. P. Chamorro-Ortega, E. P. Herrera-Granda, I. D. García-Santillán, and J. A. Quiña-Mera, "Incidence of a web application implementation for high school students learning evaluation: A case study," *Rev. Ibérica Sist. e Tecnol. Informação*, vol. 2020, no. E32, pp. 509–523, 2020, [Online]. Available: https://www.proquest.com/openview/bfe21dc96eab6a1dd96d132373a9eefc/1?pq-origsite=gscholar&cbl=1006393
45. J. García-Martín and J.-N. García-Sánchez, "The Digital Divide of Know-How and Use of Digital Technologies in Higher Education: The Case of a College in Latin America in the COVID-19 Era," *Int. J. Environ. Res. Public Health*, vol. 19, no. 6, p. 3358, Mar. 2022, doi: 10.3390/ijerph19063358.
46. C. Estrada-Muñoz, A. Vega-Muñoz, D. Castillo, S. Müller-Pérez, and J. Boada-Grau, "Technostress of Chilean Teachers in the Context of the COVID-19 Pandemic and Teleworking," *Int. J. Environ. Res. Public Health*, vol. 18, no. 10, p. 5458, May 2021, doi: 10.3390/ijerph18105458.
47. L. A. Almusfar, "Improving Learning Management System Performance: A Comprehensive Approach to Engagement, Trust, and Adaptive Learning," *IEEE Access*, vol. 13, pp. 46408–46425, 2025, doi: 10.1109/ACCESS.2025.3550288.
48. C. Primi, G. Fioravanti, S. Casale, and M. A. Donati, "Measuring Problematic Facebook Use among Adolescents and Young Adults with the Bergen Facebook Addiction Scale: A Psychometric Analysis by Applying Item Response Theory," *Int. J. Environ. Res. Public Health*, vol. 18, no. 6, p. 2979, Mar. 2021, doi: 10.3390/ijerph18062979.
49. C. Spearman, "The proof and measurement of association between two things," *Am. J. Psychol.*, vol. 15, no. 1, pp. 72–101, 1904, doi: 10.2307/1412159.
50. W. J. Conover, *Practical Nonparametric Statistics*, 3rd ed. Wiley, 1999.
51. J. Hauke and T. Kossowski, "Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data," *Quaest. Geogr.*, vol. 30, no. 2, pp. 87–93, 2011, doi: 10.2478/v10117-011-0021-1.
52. J. H. Zar, *Biostatistical Analysis*, 5th ed. Pearson, 2010.
53. E. Lehmann, *Nonparametrics Statistical Methods Based on Ranks*. New York: Springer New York, NY, 2006. [Online]. Available: https://link.springer.com/book/9780387352121
54. A. E. Jácome Ortega, J. A. Caraguay Procel, E. P. Herrera-Granda, and I. D. Herrera Granda, "Confirmatory Factorial Analysis Applied on Teacher Evaluation Processes in Higher Education Institutions of Ecuador,"

in *Advances in Intelligent Systems and Computing*, Springer, 2020, pp. 157–170. doi: 10.1007/978-3-030-37221-7_14.

55. J. Manuel Batista-Foguet, G. Coenders, and J. Alonso, "Análisis factorial confirmatorio. Su utilidad en la validación de cuestionarios relacionados con la salud," *Med. Clin. (Barc).*, vol. 122, no. Supl.1, pp. 21–27, Feb. 2004, doi: 10.1157/13057542.
56. Y. Rosseel, "lavaan: An R Package for Structural Equation Modeling," *J. Stat. Softw.*, vol. 48, no. 2, pp. 1–36, 2012, doi: 10.18637/jss.v048.i02.
57. F. Yang-Wallentin, K. G. Jöreskog, and H. Luo, "Confirmatory Factor Analysis of Ordinal Variables With Misspecified Models," *Struct. Equ. Model. A Multidiscip. J.*, vol. 17, no. 3, pp. 392–423, 2010, doi: 10.1080/10705511.2010.489003.
58. G. Rasch, *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: Danish Institute for Educational Research, 1960. doi: 10.7208/chicago/9780226705756.001.0001.
59. A. Birnbaum, "Some Latent Trait Models and Their Use in Inferring an Examinee's Ability," in *Statistical Theories of Mental Test Scores*, F. M. Lord and M. R. Novick, Eds., Reading, MA: Addison-Wesley, 1968, pp. 397–479.
60. F. M. Lord, *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1980. doi: 10.4324/9780203056615.
61. F. Nwobi and F. Akanno, "Power comparison of ANOVA and Kruskal–Wallis tests when error assumptions are violated," *Adv. Methodol. Stat. / Metod. Zv.*, vol. 18, no. 2, pp. 53–71, 2021, doi: https://doi.org/10.51936/ltgt2135.
62. O. J. Dunn, "Estimation of the Means of Dependent Variables," *Ann. Math. Stat.*, vol. 29, no. 4, pp. 1095–1111, Feb. 1958, [Online]. Available: http://www.jstor.org/stable/2236948
63. E. P. Herrera-Granda, M. J. Aza-Espinosa, M. Burbano-Pulles, J. Mina-Ortega, I. D. Herrera-Granda, and W. J. Yambay-Vallejo, "Statistical analysis of digital transformation and its incidence in reducing the use of paper in a higher education institution: A case study," *J. Technol. Sci. Educ.*, vol. 14, no. 4, p. 1041, Sep. 2024, doi: 10.3926/jotse.2242.
64. M. S. BARTLETT, "THE STATISTICAL CONCEPTION OF MENTAL FACTORS," *Br. J. Psychol. Gen. Sect.*, vol. 28, no. 1, pp. 97–104, Jul. 1937, doi: 10.1111/j.2044-8295.1937.tb00863.x.
65. B. Muthén, "A General Structural Equation Model with Dichotomous, Ordered Categorical, and Continuous Latent Variable Indicators," *Psychometrika*, vol. 49, no. 1, pp. 115–132, Mar. 1984, doi: 10.1007/BF02294210.