

# PRISM-EDU: Mining ENADE Microdata to Diagnose Learning Deficiencies in Brazilian Higher Education

Ivanylson Honorio Gonçalves<sup>1</sup>, Victor Ströele<sup>1</sup>, Regina Braga<sup>1</sup>, José Maria David<sup>1</sup>

<sup>1</sup>Postgraduate Program in Computer Science – Federal University of Juiz de Fora (UFJF) Caixa Postal 20.010 – 36.016-970 – Juiz de Fora – MG – Brazil

ivanylson.honorio@estudante.ufjf.br, victor.stroele@ufjf.br,  
regina.braga@ufjf.br, jose.david@ufjf.br

**Abstract. Introduction:** ENADE microdata are underused. **Objective:** The aggregate institutional score blends different learning patterns into a single grade, offering no specific signals for content intervention. **Methodology:** PRISM-EDU employs exploratory Latent Class Analysis on item responses from each Higher Education Institution (HEI), matching profiles with INEP's item–Knowledge Object (KO) mapping as an implicit Q-matrix, and translating results into four actionable zones and two diagnostic categories. **Results:** Using ENADE 2023 data (230,088 students, 1,880 HEIs, 28 programs), the pipeline converges, showing mostly horizontal internal heterogeneity. KO-level diagnosis for Civil Engineering, Medicine, and Nursing highlights repeatedly deficient KOs, exposing structural training gaps beyond the aggregate score.

**Keywords** Intelligent Agents, Teacher Training, Education Management, Learning Analytics, Educational Data Mining.

## 1. Introduction

The National Student Performance Examination (ENADE) is Brazil's largest source of standardized evaluation data on Brazilian undergraduate education, with microdata covering hundreds of thousands of graduating students each cycle [Leão et al. 2021], [Barbosa et al. 2023]. Each ENADE item is officially mapped by INEP to a specific Knowledge Object (KO), forming, in essence, a Q-matrix: a curricular grid linking observable responses to the content a graduating student is expected to master. In principle, this design makes ENADE a rich diagnostic instrument; in practice, the exam is consumed almost exclusively through its aggregate institutional concept, which conflates very different learning patterns into a single grade and provides course coordinators with no actionable signal about which content requires intervention.

Prior work on ENADE microdata reflects this limitation. Student-preparation platforms address individual study rather than class-level diagnosis [Damazio and Carvalho 2023], [Chagas Júnior et al. 2023], [Brito et al. 2025], [Ribeiro et al. 2022]. Curricular and item analyses operate on per-item average success rates, ignoring how gaps co-occur within students [Lima et al. 2018], [Charão et al. 2020], [Cunha et al. 2021], [Souza et al. 2024]. Machine-learning approaches predict the institutional concept or model socioeconomic effects based on the aggregate score [Rezende et al. 2022], [Gondran et al. 2023], [Sakashita et al. 2023]. When clustering is used, it relies on the aggregate score itself, yielding a vertical reading of the class as "strong" vs. "weak" students [Vieira et al. 2022]. More recent work has used Large Language Models (LLMs) as test-takers or curricular-analysis assistants [Carvalho et al. 2026], [Mendonça 2024],

[Silva et al. 2025]. Three gaps cut across these efforts: (i) diagnoses are computed on aggregates rather than on respondents' latent response patterns; (ii) the official INEP item-KO mapping is rarely exploited as a Q-matrix for linking statistical findings with curricular vocabulary; and (iii) coverage mainly focuses on Computing-related programs.

The closest line of work outside the Brazilian literature is the family of Cognitive Diagnosis Models (CDMs) [Rupp et al. 2010], [de la Torre 2011], which classify examinees into mastery profiles using a Q-matrix. CDMs such as DINA, DINO, LCDM, and G-DINA, however, require an instrument designed for diagnosis from the outset and a Q-matrix validated item by item by experts, conditions that ENADE, as a summative exam, does not meet. For settings like this, methodological literature recommends exploratory Latent Class Analysis (LCA) as a principled alternative that allows profiles to emerge from the data without imposing conjunctive, disjunctive, or compensatory structures [Hickendorff et al. 2018], [Weller et al. 2020]. Building on this gap, we investigate the following Research Question: *For students at a given Higher Education Institution (HEI) enrolled in a specific program, which ENADE Knowledge Objects (KO) represent the most relevant learning deficiencies, and how do these deficiencies distribute across student profiles within the program?* We address this RQ by proposing PRISM-EDU (Prescriptive Insight System for Mining Educational Data). This architecture integrates exploratory LCA over item-level response vectors with INEP's official item-KO classification. It translates the resulting statistical profiles into four pedagogically actionable zones (mastery, intermediate, deficient, critical) and two diagnostic categories (chronically deficient KOs and profile-marker KOs). Profile stability is validated by bootstrap resampling [Hennig 2007] and cross-checked against an independent K-Means + Multiple Correspondence Analysis pipeline via the Adjusted Rand Index. Inter-institutional regularities are surfaced by Ward's hierarchical clustering over the standardized gap space. This paper makes four main contributions: (1) a person-centered diagnostic pipeline for ENADE microdata that combines exploratory LCA with INEP's official item-KO matrix used as an implicit Q-matrix, enabling actionable curricular diagnoses rather than aggregate-score readings; (2) a reproducible methodological protocol for selecting and validating latent profiles on summative-exam data (BIC + normalized entropy + bootstrap + ARI cross-validation), aligned with best practice in person-centered educational research [Nylund et al. 2007], [Weller et al. 2020]; (3) a nationwide empirical study covering 230,088 graduating students across 1,880 HEIs and 28 undergraduate programs going well beyond the Computing-focused scope that dominates the Brazilian literature; and (4) an empirical finding with direct pedagogical implications: internal class heterogeneity in Brazilian higher education is predominantly horizontal in 92.0% of HEIs with two-group solutions, profiles differ by specialization (complementary KO mastery) rather than by overall proficiency level. Moreover, these profiles recur as national archetypes across nearly all HEIs in the analyzed courses, and several KOs emerge as chronically deficient, signaling structural rather than local training gaps.

The article is organized as follows. Section 2 reviews related work on cognitive diagnosis models, person-centered educational research, and ENADE microdata analysis. Section 3 details the PRISM-EDU architecture and methodology. Section 4 reports the results of an instantiation study on ENADE 2023. Section 5 discusses pedagogical implications and threats to validity. Section 6 concludes and outlines future work.

## 2. Related Work

PRISM-EDU is an Educational Data Mining architecture that extracts actionable knowledge from educational data to support teaching and learning [Romero and Ventura 2020], [Siemens 2013]. The related research includes Cognitive Diagnosis Models (CDMs) or Diagnostic Classification Models (DCMs) [Rupp et al. 2010], [de la Torre 2011], which are specialized versions of Latent Class Analysis (LCA) used to classify examinees into mastery profiles based on a Q-matrix. Models like DINA [Junker and Sijtsma 2001], DINO [Templin and Henson 2006], LCDM [Henson et al. 2009], and G-DINA [de la Torre 2011] have been applied to contexts such as PISA [Hsu et al. 2020] and vocational exams [Tjoe and de la Torre 2022]. We used LCA because the INEP item-KO matrix functions as a Q-matrix, and the profiles align with mastery profiles. Parametric CDMs need a validated Q-matrix and item-specific adjustments, which requires an instrument designed for diagnosis, something ENADE is not. Thus, exploratory LCA is used to identify profiles directly from data, allowing emergence of response patterns without restrictive assumptions, as in PRISM-EDU. LCA detects unobserved subgroups with distinct response patterns, indicating whether students form qualitatively different groups concerning KO. For  $J$  items and  $K$  groups, the probability of response vector  $y$  is  $P(Y=y) = \sum_k \pi_k \prod_j p_{jk}^{y_j} (1 - p_{jk})^{1-y_j}$ , where  $\pi_k$  is group  $k$ 's size and  $p_{jk}$  is the probability of a correct response. Each group reflects a student profile, with  $p_{jk}$  indicating mastery or deficiency. The best number of groups ( $K$ ) is guided by BIC, entropy, and group sizes to avoid overfitting. Profile stability was validated with bootstrap resampling, and the profiles were cross-verified with K-Means clustering on Multiple Correspondence Analysis, confirming high agreement via the Adjusted Rand Index.

Regardless of the psychometric tradition, LCA has established itself as a person-centered tool in educational research [Hickendorff et al. 2018], [Weller et al. 2020]. Unlike variable-centered approaches, which treat the population as homogeneous, the person-centered approach identifies qualitatively distinct subgroups in the joint pattern of responses. Recent applications to higher education [Fryer 2017] and to learning strategies in mathematics [Hickendorff et al. 2010] have demonstrated a recurring finding: students often differ not in their level of proficiency but in their mastery patterns, mastering distinct subsets of the content. The horizontal heterogeneity observed in this research, in which ENADE student archetypes predominate in complementary KOs, is consistent with this structural pattern. In this work, we selected KOs using information criteria, normalized entropy, bootstrap validation, and assessed cross-sample replicability, consistent with the systematized best practices outlined in [Nylund et al. 2007], [Nylund-Gibson and Choi 2018], and [Weller et al. 2020]. ENADE microdata has been a recurring subject of investigation in Brazil. [Damazio and Carvalho, 2023], [Chagas Júnior et al. 2023], [Brito et al. 2025], and [Ribeiro et al. 2022] aim to train students individually rather than provide a collective diagnosis of the class. In contrast, [Lima et al. 2018], [Charão et al. 2020], [Cunha et al. 2021], and [Souza et al. 2024] automate curricular analyses and question classification by operating on aggregated item-level averages, thereby failing to capture latent heterogeneity. Similarly, [Rezende et al. 2022], [Gondran et al. 2023], and [Sakashita et al. 2023] apply machine learning to socioeconomic diagnosis and prediction of the institutional concept. When grouping is included, [Vieira et al. 2022] base it on the aggregate score, producing a vertical reading of the class ("strong" and "weak" students) which, as our results show, obscures the predominant horizontal heterogeneity. Meanwhile, [Carvalho et al. 2026] and [Mendonça 2024] use

language models as test candidates, whereas [Silva et al. 2025] use them to support curricular analysis. The problem with these approaches is that the diagnosis operates at the aggregate level rather than at the level of respondents' latent profiles. In addition, most studies are restricted to computer science courses and related areas. PRISM-EDU proposes an approach that combines the person-centered style of educational research with the official curricular structure of ENADE: it applies exploratory LCA to the complete pattern of responses and cross-references the profiles with INEP's implicit Q-matrix (the item-KO classification). Furthermore, it adopts consolidated methodological best practices for this class of models (K selection by BIC and entropy, validation by bootstrap and independent methods), analyzing 1,880 HEIs across 28 courses. Finally, it translates statistical profiles into curricular vocabulary, classifying KOs into four pedagogical zones (mastery, intermediate, deficient, critical) and flagging two diagnostic categories (chronically deficient and profile-marker KOs) that directly guide the appropriate intervention in each case. To facilitate understanding of the results, it also uses an explainable dashboard that supports analysis by coordinators and teachers.

### 3. PRISM-EDU

The PRISM-EDU architecture is organized into services (Figure 1) orchestrated into a pipeline. PRISM-EDU collects microdata, extracts competencies (via Data Collector and NLP), models learning profiles and deficiencies (via unsupervised algorithms such as LCA and K-Means), and delivers the results in an interactive control panel that generates pedagogical plans. The central objective is to move beyond purely descriptive analysis of ENADE by automating the identification of learning gaps and generating data-driven pedagogical action plans. As illustrated in Figure 1, the architecture comprises five services. Although instantiated here on ENADE, the pipeline is designed for any competency-based large-scale exam with an item–competency mapping.

#### 3.1. Knowledge Modeling Service

This service cross-references official information to define the relationship between test items and competencies, making it clear which skills are assessed. Using Natural Language Processing (NLP), it structures the INEP competency matrix and retrieves the official classification for each of the 28 courses covered by this study. This allows educators, administrators, and policymakers to easily identify competency gaps and align educational strategies. A master table was constructed to associate each of the 38 evaluated items (8 from General Education and 30 from the Specific Component) with its main Knowledge Object (KO), professional profile, and specific competency. Because several items may assess the same KO and the Specific Component differs across programs, the 38 items map onto a course-dependent number of distinct KOs (e.g., 20 in Civil Engineering, 13 in Medicine, and 23 in Nursing). This table defines the relationship between the statistical output and the competencies that teachers should develop.

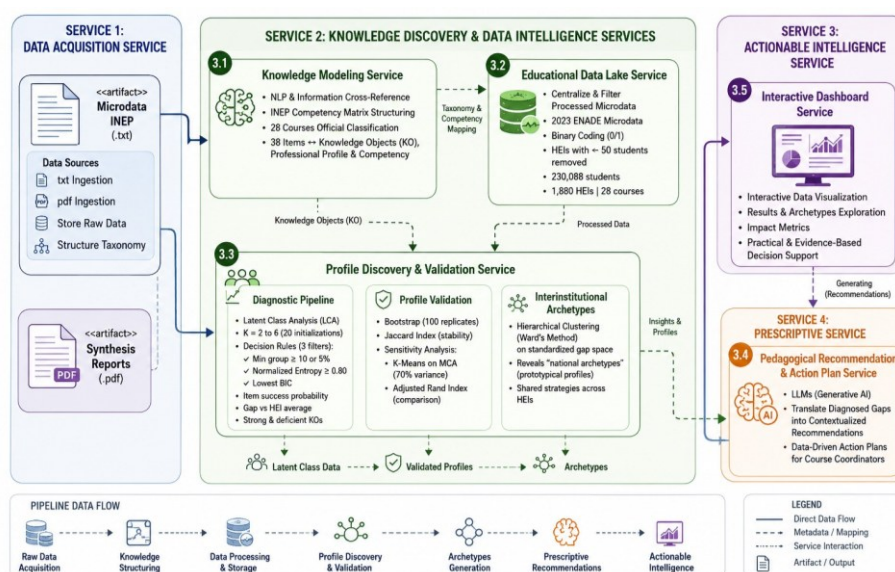


Figure 1. Architecture of the PRISM-EDU.

### 3.2. Educational Data Lake Service

This service centralizes and filters the processed public microdata. In this article, we processed the 2023 ENADE microdata. The item-by-item responses of each graduating student were coded as binary: correct (1) or incorrect (0), with blank and annulled responses treated as incorrect, following INEP's convention. To ensure statistical stability of the estimates, HEIs with fewer than 50 students were excluded. The final dataset comprises 230,088 students from 1,880 HEIs across 28 undergraduate courses.

### 3.3. Profile Discovery and Validation Service

This service applies the diagnostic pipeline, validation, and interinstitutional grouping independently by HEI and course: (i) **Diagnostic Pipeline**: Latent Class Analysis (LCA) models were fitted with K varying from 2 to 6 groups, using 20 random initializations per model to avoid local maxima. The final K is chosen by a priori decision rules with three filters: (1) The smallest group must have at least 10 students or 5% of the sample to eliminate spurious groups; (2) Normalized entropy must be  $\geq 0.80$  [Celeux and Soromenho 1996] to ensure clear profile separation; (3) The solution with the lowest BIC is chosen. The pipeline then calculates the estimated probability of success for each item to determine the gap relative to the HEI average and identify strong and deficient KOs per profile. (ii) **Profile Validation**: Stability was evaluated via bootstrap with 100 replicates [Hennig 2007]. This measures the coincidence between original and obtained groups using the Jaccard index. A sensitivity analysis was conducted by fitting K-Means to Multiple Correspondence Analysis coordinates (preserving 70% of the variance) and comparing assignments using the Adjusted Rand Index. For two-class solutions, the horizontal or vertical reading uses raw class-conditional probabilities, never gaps centered on the HEI mean, whose signs are opposite by construction when  $K = 2$ . Therefore, an item is informative when the between-class difference reaches 0.10 and survives a two-proportion test ( $\alpha = 0.05$ ), and an HEI is horizontal when the minority direction holds at least 25% of informative items. This criterion was calibrated against homogeneous and purely vertical null models simulated with each HEI's class sizes. (iii) **Interinstitutional Archetypes**: To investigate consistent interinstitutional patterns,

hierarchical clustering (Ward's method) [Murtagh and Legendre 2014] was applied to all identified groups within the standardized gap space. This reveals "national archetypes," prototypical profiles reproduced among HEIs, allowing professors from different institutions to benefit from shared strategies.

### 3.4. Prescriptive Service

This service bridges the statistical output of the Profile Discovery pipeline with pedagogically interpretable recommendations. For each identified profile, it aggregates the diagnosed KO gaps and feeds them along with the KO label, profile success probabilities, and course context into a structured prompt processed by a Large Language Model (LLM). The LLM's role is strictly interpretive: it translates the numerical gap structure into natural-language action suggestions oriented toward course coordinators, without altering or reinterpreting the statistical findings.

To illustrate, suppose the pipeline reports that, in a given Civil Engineering HEI, one of the two identified profiles shows a critical deficiency in a foundational KO while the other reaches mastery in the same KO. The service produces a three-part action plan: (i) *Diagnosis*: a natural-language reading of what the gap means pedagogically (e.g., "roughly half of graduates have mastered applied problem-solving but lack the scientific basis underpinning it"); (ii) *Recommended Actions*: targeted interventions such as preparatory mini-courses on the foundational content before dependent disciplines, mixed-team activities pairing the complementary profiles for peer learning, and curricular weight adjustments toward core KOs; (iii) *Monitoring*: lightweight formative instruments (short quizzes linked to the target KO) for tracking recovery over the semester. Concrete instances of this output, grounded in the empirical profiles identified in Section 4, are discussed alongside the dashboard in Section 5. The adequacy and practical utility of these LLM-generated recommendations in real curriculum-review cycles are not evaluated in this paper; a qualitative study with course coordinators is planned as future work. The present instantiation demonstrates the feasibility of the service as a component of the PRISM-EDU pipeline. All code and data are available at the Data Availability statement at the end of this article.

### 3.5. Actionable Intelligence Service

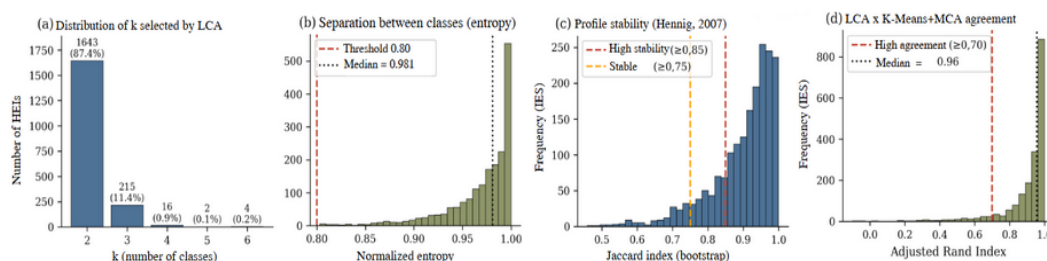
This service provides an interactive data visualization dashboard that transforms results into practical, evidence-based pedagogical interventions for higher education institutions. A screenshot is shown in Section 5.

## 4. Results

The architecture was evaluated through an instantiation study [March and Smith 1995] with ENADE 2023 microdata (230,088 students, 1,880 HEIs, 28 courses). An instantiation study [March and Smith 1995] is a type of Design Science research focused on building and evaluating a concrete IT artifact, such as a prototype, system, or tool, that brings an abstract concept into existence.

### 4.1. Technical Validation of the Pipeline

Applying the pipeline to the 1,880 HEIs led to complete convergence of the estimation algorithm. Figure 2 summarizes the quality indicators of the identified profiles.



**Figure 2. Quality indicators of the identified profiles: (a) selected K per HEI; (b) normalized entropy; (c) bootstrap stability (Jaccard); (d) LCA x K-Means+MCA agreement**

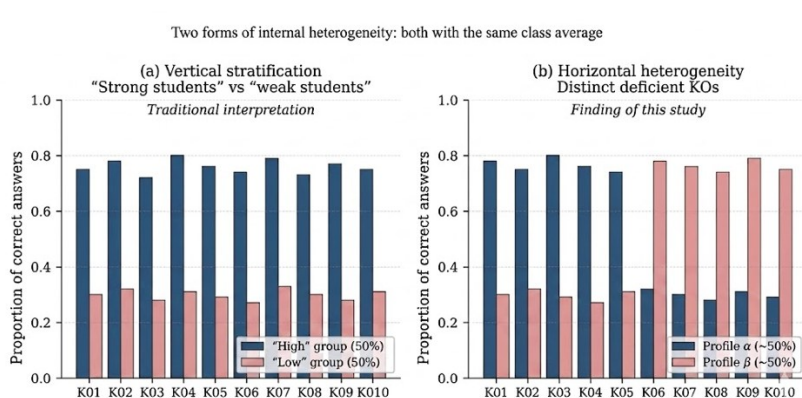
Figure 2 shows that the normalized entropy had a median of 0.98, indicating that most of the identified profiles are clear, with no ambiguity about which group they belong to. Similarly, bootstrap stability had a median of 0.93, with 92% of the HEIs showing Jaccard indices  $\geq 0.75$  (acceptable stability according to [Hennig 2007]), which indicates that the identified profiles are not artifacts of a particular sample. Furthermore, the agreement between LCA and the alternative K-Means+MCA method had a median of 0.96, reinforcing confidence that the profiles reflect the true data structure. Notably, 87.4% of HEIs chose a two-group design, suggesting that in most of the analyzed Brazilian courses, student heterogeneity is best captured by two profiles. Lastly, solutions with three or more groups occurred only in the larger HEIs, suggesting that additional profiles can be defined only with sufficiently large sample sizes to enable reliable statistical identification.

#### 4.2. Characterization of the Identified Profiles

Initially, we analyzed the results by dividing profiles (groups) for a pedagogical evaluation. This initial analysis aims to determine whether the profiles represent distinct learning levels (vertical stratification) or are complementary in some way (horizontal heterogeneity), that is, one profile has knowledge of one set of competencies, while the other has knowledge of another. Across the 1,643 two-group HEIs, 92.0% show profiles that differ in KO specialization rather than in overall proficiency level (Figure 3b). The share remains between 81.5% and 96.7% across all threshold combinations tested, and HEIs simulated as purely vertical are misread as horizontal in only a median of 6.2% of replicates (8.0% for homogeneous ones). The pattern varies by program: engineering programs are horizontal in nearly every HEI, Medicine is 93.4% horizontal, and Nursing splits between 64.0% horizontal and 36.0% vertical. Ranking students by overall skill level, therefore, misses where most classes struggle, with Nursing as the main partial exception.

#### 4.3. National Archetypes and Diagnosis by Knowledge Objects

We then asked whether the profiles identified at each HEI follow patterns shared across institutions in the same program. We applied Ward's hierarchical clustering to the standardized gap space of all per-HEI profiles. The profiles consistently align with the same two national archetypes in 127 of 127 Civil Engineering HEIs, 218 of 218 in Medicine, and 229 of 230 in Nursing. Building on this division, these two types explain between 47% and 60% of the differences in item-level probabilities.



**Figure 3. (a) Traditional reading (b) Profile-specialization reading.**

Therefore, per-HEI profiles are not local phenomena but expressions of a shared national structure. The pedagogical consequence is direct: KOs that appear chronically deficient across nearly all HEIs in a program signal a structural training gap, not local problems. This distinction helps coordinators separate "issues in my class" from "issues in training in my area". To translate this national structure into curricular vocabulary, we proceed as follows. For each archetype, we aggregate item-level success probabilities by Knowledge Object. We use INEP's official item-KO mapping (Annex IX of the course summary report) as an implicit Q-matrix. We then classify each KO into four pedagogical zones: mastery ( $\geq 0.65$ ), intermediate ( $\geq 0.50$  and  $< 0.65$ ), deficient ( $\geq 0.40$  and  $< 0.50$ ), and critical ( $< 0.40$ ). Figure 4 presents the resulting diagnostic map for the three programs analyzed.

The complementarity pattern is direct and pronounced. Labeling the two national archetypes  $\alpha$  and  $\beta$ , in Civil Engineering, archetype  $\alpha$  dominates "Statistics, physics, mathematics, and chemistry applied to Civil Engineering" (0.66), where  $\beta$  is critically deficient (0.05). Conversely,  $\beta$  leads in "Geotechnics" (0.48 vs. 0.23) and "Graphic expression and universal design" (0.45 vs. 0.002). Also, the two archetypes' success rates always add up to about the same number for each KO, showing that where one does well, the other does not. This is consistent with the presence of two distinct training pathways in Civil Engineering: one emphasizing theoretical and formal content, the other emphasizing practical and technical content. Shifting to Medicine, the complementarity is more subtle:  $\alpha$  leads in "Culture, art and communication" (0.94) and "Mental Health" (0.81), while  $\beta$  is stronger in "Accessibility and social inclusion" (0.82 vs. 0.63). In Nursing,  $\alpha$  is highest for "Palliative care" (0.89), "Care technologies" (0.65), and "Nursing process" (0.67). Since this complementarity is reproduced across nearly all HEIs, the most plausible interpretation is that it reflects national curricular structures rather than institutional choice. A second diagnostic question addresses KOs in which both archetypes score below 0.50, indicating persistent deficiencies that affect the entire training field in Brazil. Figure 5 presents these findings. Civil Engineering exhibits 11 chronically deficient KOs, suggesting broad gaps in the basic training of Brazilian civil engineers: "Electricity applied to Civil Engineering" (0.015), "Water resources and sanitation" (0.20), "Materials science" (0.18), "Topography and geoprocessing" (0.20), and even core areas such as "Civil construction" (0.32) and "Geotechnics" (0.36) fall into the critical zone. Medicine shows only two such KOs, reflecting the program's stronger overall performance, but these include central training contents: "Clinical Medicine" (0.50) and "Family and Community Medicine" (0.48). In contrast, Nursing displays

critical deficiencies in "History of nursing and legislation" (0.06) and "Nursing care in urgent and emergency situations" (0.23). When a curricular unit lacks both archetypes across most HEIs, the issue is structural, not local. It stems from factors like teaching hours from the National Curricular Guidelines, material quality, and teacher training, which need intervention beyond the classroom.

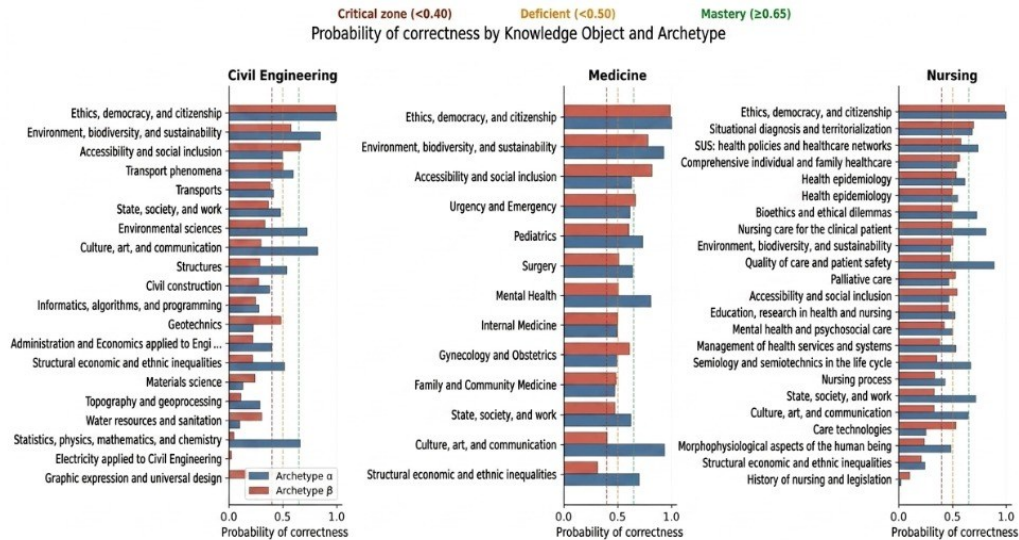
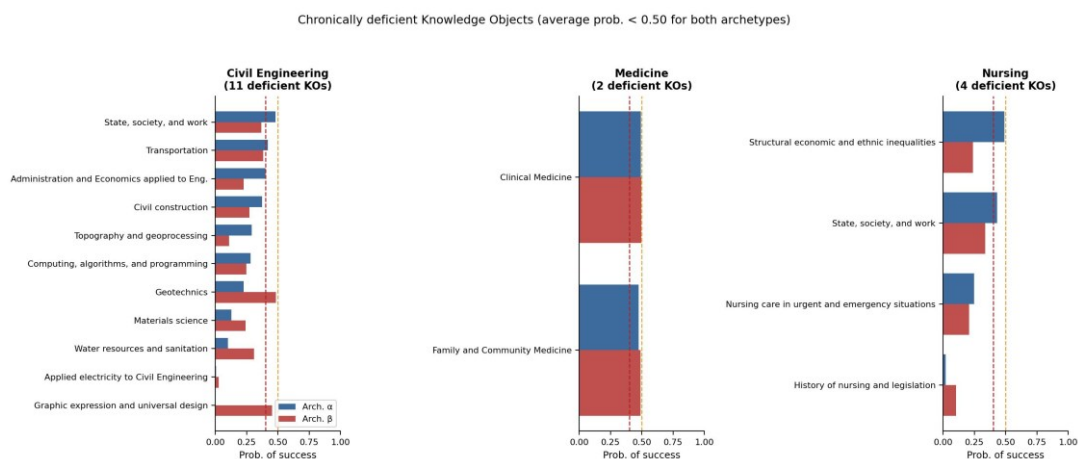


Figure 4. Success probability per Knowledge Object and archetype (blue =  $\alpha$ , red =  $\beta$ ).

Table 1. Diagnosis of KOs by course: distribution among the pedagogical zones.

| Course            | KOs evaluated | KOs chronically deficient | KOs profile markers | Avg. probability |
|-------------------|---------------|---------------------------|---------------------|------------------|
| Civil Engineering | 20            | 11                        | 9                   | 0.40             |
| Medicine          | 13            | 2                         | 3                   | 0.64             |
| Nursing           | 23            | 4                         | 8                   | 0.51             |

Table 1 summarizes diagnoses by program. Chronically deficient KOs are those with an average success probability  $< 0.50$  in both archetypes; profile-marker KOs are those in which the absolute difference between archetypes is  $\geq 0.20$ . Civil Engineering combines the highest number of chronically deficient KOs (11/20) with the lowest overall mean probability (0.40). This is the most concerning scenario among the three. Medicine sits at the opposite end (2 chronically deficient KOs, mean 0.64), and Nursing is in between (4 chronically deficient, mean 0.51). With respect to inter-archetype divergence, Civil Engineering has 9 out of 20 KOs, defined as specific assessment objectives that serve as profile markers, meaning that nearly half of the Specific Component content strongly distinguishes between student archetypes  $\alpha$  and  $\beta$ , that is, the two complementary profiles defined in Section 4.3. Nursing follows with 8 out of 23 KOs (35%), and Medicine with 3 out of 13 (23%).



**Figure 5. Chronically deficient KOs by course.**

Medical training, therefore, yields more convergent profiles, whereas Civil Engineering and Nursing yield more specialized training paths. Whether this divergence reflects legitimate curricular choice, students' vocational profiles, or unintentional fragmentation of training is an open question for further investigation. Across the three programs, this triple profile differs in the magnitude of the deficiency. Yet, the pattern remains structurally identical in its bimodal specialization. This reinforces the central finding: regardless of the course's average learning level, Brazilian undergraduate students tend to fall into two profiles of complementary KO mastery. Identifying these profiles, together with curricular cross-referencing, gives instructors pedagogically actionable information that the aggregate ENADE score cannot.

## 5. Discussion

The dominant finding, that internal class heterogeneity is structured by KO specialization rather than overall proficiency, recasts ENADE as a person-centered rather than a variable-centered signal. A vertical reading licenses leveling interventions; the horizontal pattern we observe instead licenses complementarity strategies (peer learning between archetypes) and targeted KO-specific reinforcement per profile. These imply different uses of limited classroom time. To support these uses in practice, an explanatory dashboard with LLM-assisted interpretation is made available to coordinators (Figure 6).

Our results align with and extend three lines of prior work. They confirm that exploratory LCA recovers interpretable mastery profiles in summative-exam settings where a fully specified Q-matrix is unavailable [Hickendorff et al. 2018], reproduce at the national scale the person-centered finding that students differ in patterns rather than in levels of mastery [Weller et al. 2020], [Fryer 2017], and contradict the implicit assumption of Brazilian variable-centered approaches that group students by aggregate performance [Vieira et al. 2022]: in 92.0% of two-group HEIs, aggregate proficiency is not the dominant axis of variation. This has consequences for any downstream analysis that conditions on the aggregate score, including predictive modeling and socioeconomic profiling, thereby risking the obscuration of exactly the heterogeneity that matters pedagogically. For coordinators, the distinction between chronically deficient and profile-marker KOs shifts the locus of intervention: chronically deficient KOs that appear in nearly all HEIs within a program point to structural gaps in the National Curriculum Guidelines or teacher training, while profile-marker KOs point to complementarity within

the class. For the evaluation research community, our results suggest that institutional concepts discard the latent structure carrying the most actionable pedagogical information, making item-KO releases like INEP's a resource the community should exploit more systematically.

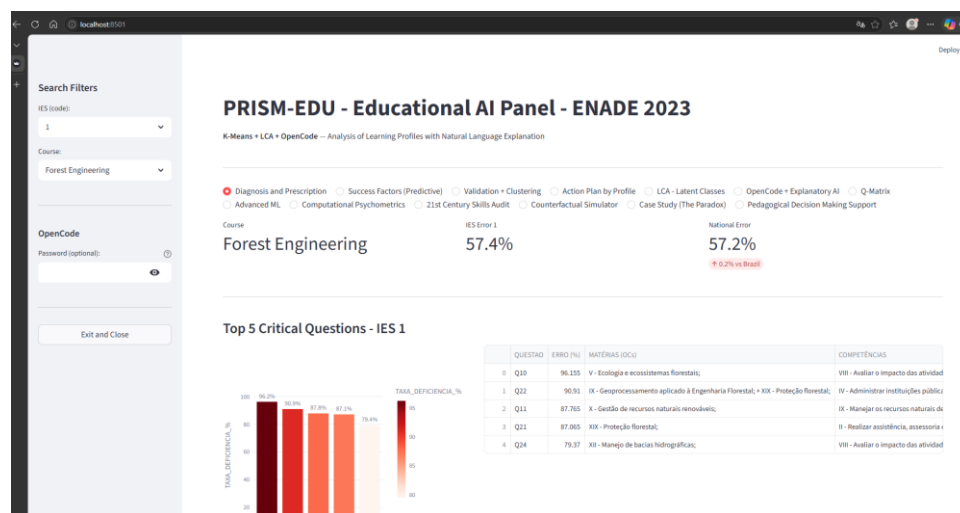


Figure 6. PRISM-EDU interactive dashboard (ENADE 2023)

### 5.1. Threats to Validity

**Construct validity:** The INEP item-KO mapping served as an implicit Q-matrix without extra expert validation at the item level, as ENADE was not made as a diagnostic tool. Item difficulty and discrimination were not explicitly modeled; blank and annulled responses were marked incorrect following INEP's rules. These limitations come from using a summative exam diagnostically. Since the exam is low stakes, blanks may signal disengagement. Recomputing KO success rates among students who answered each item shows all 17 chronically deficient KOs below 0.50 (the two Medicine KOs are at 0.498). Blank rates weakly or negatively correlate with item position ( $r = -0.22, -0.07, 0.11$  in Nursing, Civil Engineering, and Medicine), not indicating end-of-test disengagement. The two codings bound mastery probability from opposite sides, with the same deficient-KO list in both. **Internal validity:** Profile identification was strengthened through multiple methods: BIC-guided K selection with normalized entropy  $\geq 0.80$ , bootstrap validation (median Jaccard 0.93, 92% above threshold of 0.75), and cross-validation via K-Means+MCA (median ARI 0.96). While LCA may absorb response heterogeneity into classes, two data features argue against this: strong/weak KO reversals cluster in meaningful content groups, not randomly, and this pattern independently appears across 127, 218, and 229 HEIs in Civil Engineering, Medicine, and Nursing. **External validity:** The findings relate to graduating students in HEIs with  $\geq 50$  examinees in the 2023 ENADE cycle. Pre-exam dropout, inter-cycle profile stability, and generalization to smaller HEIs remain untested. Detailed KO-level diagnosis is reported for three programs (Civil Engineering, Medicine, Nursing) chosen for contrasting success rates (0.40, 0.64, 0.51); the pipeline was applied to all 28 programs, with global convergence stats reported. **Conclusion validity:** Aggregate claims summarize the analyzed HEIs, not hypothesis tests. The pedagogical thresholds (0.40 / 0.50 / 0.65) are standard; a sensitivity analysis shows shifting them by  $\pm 0.05$  can reclassify borderline KOs but does not change the pattern of deficient and marker KOs in each program.

## 6. Conclusions and Future Work

This work presented PRISM-EDU, an architecture that demonstrates how integrating curriculum mapping based on Natural Language Processing (NLP) with data mining (via LCA) can transform a competency-based large-scale evaluation, demonstrated here on ENADE, from a mere generator of 'institutional grades' into a prescriptive pedagogical compass. To evaluate its feasibility, the architecture was tested in an instantiation study based on an analysis of data from 230,088 students across 1,880 higher education institutions (HEIs). It was shown that the heterogeneity of learning in Brazilian classes is predominantly horizontal: students are not linearly divided into "strong" and "weak" students but rather group themselves into profiles with differentiated domains of Knowledge Objects (KOs). Furthermore, we demonstrated that these profiles follow national archetypes, reproducing themselves in almost 100% of the HEIs across the analyzed courses. The cross-referencing of statistical findings with the official classification of Knowledge Objects published by INEP revealed not only pedagogically nameable specializations, such as Civil Engineering, where one archetype is oriented towards theoretical-formal content and the other towards practical-technical content, but also chronically deficient KOs. Content areas such as "Water Resources and Sanitation" in Civil Engineering or "Clinical Medicine" in Medicine proved deficient in both archetypes and in almost all institutions, signaling chronic structural gaps in higher education in Brazil. Therefore, by focusing on Knowledge Objects and segmenting latently structured profiles, the tool provides course coordinators with granular, actionable, and explainable evidence for curriculum review and targeted interventions, filling a critical gap in the country's educational evaluation literature.

Therefore, we could answer the RQ: within each HEI-program pair, the most relevant learning deficiencies are the KOs we classify as "chronically deficient" (low mastery probability in both archetypes) and as "profile markers" (large mastery gap between archetypes); the former indicate structural training gaps shared across HEIs, while the latter indicate complementary specializations between coexisting student profiles within the same class. This distribution is dominated by horizontal heterogeneity rather than by a single proficiency axis and provides coordinators with a curriculum-level diagnostic signal that the aggregate ENADE concept does not. As future work, we plan to conduct a longitudinal analysis by extending it across multiple ENADE cycles to determine whether the chronically deficient KOs identified worsen or improve over time, providing evidence about the impact of curriculum reforms and the National Curriculum Guidelines on graduating-student performance. We also intend to incorporate expert validation of INEP's item-KO matrix to move from exploratory LCA toward restricted Cognitive Diagnosis Models. Moreover, we intend to instantiate PRISM-EDU on other competency-based large-scale exams ENAMED (Brazilian National Medical Students Exam), Saber Pro (Colombia), EGEL (Mexico), TECO (Italy), and OECD's PIAAC to test the methodological generality of the architecture and surface cross-country structural patterns in higher-education learning, providing the necessary service extensions; and to conduct a qualitative study with course coordinators and instructors to evaluate the actionability of the action plans produced by the Prescriptive Service, and measure adoption and perceived utility of the dashboard in real curriculum-review cycles.

## Data Availability

<https://github.com/Ivanylson/PRISM-EDU>

## Acknowledgments

This work was partially funded by the Federal University of Juiz de Fora (UFJF, Brazil), the Coordination for the Improvement of Higher Education Personnel (CAPES, Brazil), and the National Council for Scientific and Technological Development (CNPq, Brazil), under grants 307194/2022-1, 313568/2023-5, and 309129/2025-7.

## Ethical Considerations

The development and application of the PRISM-EDU architecture adhered to rigorous ethical and legal standards for the handling of educational data.

## Statement on the use of artificial intelligence

The authors declare that LLMs are used in two distinct roles in this work: as a component of the proposed architecture (the Prescriptive Service, Section 3.4) and as writing-assistance tools during manuscript preparation. All statistical findings were produced by the LCA pipeline. The LLM does not alter or generate any of the quantitative results.

## References

- Barbosa, P. L. S., Damazio, G. N. D. O., Carvalho, W. V. and Carmo, R. (2023). O sucesso não é apenas uma questão de sorte: um mapeamento sistemático sobre técnicas de análise do Enade da área de Computação. In Anais do III Simpósio Brasileiro de Educação em Computação (EDUCOMP), p. 59–68. DOI: 10.5753/educomp.2023.228963
- Brito, L. M., Nascimento, F. S., Monteiro, V. C. S. and Mourão, A. B. (2025). Prepara Enade: plataforma web para apoiar o desempenho dos concluintes de licenciatura em Computação em conteúdos de educação especial e inclusiva. In Anais do XXXI Workshop de Informática na Escola (WIE), p. 133–144. DOI: 10.5753/wie.2025.247421
- Carvalho, L. M., Carvalho Júnior, C. M. and Mendonça, N. C. (2026). Evaluating the accuracy and stability of frontier LLMs on ENADE Computer Science questions. In R. de Freitas and D. Furtado (eds.), Intelligent Systems (BRACIS 2025), Lecture Notes in Computer Science, vol. 16182. Springer. DOI: 10.1007/978-3-032-15993-9\_18
- Celeux, G. and Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification*, 13(2):195–212. DOI: 10.1007/BF01246098
- Chagas Júnior, J. M., Amora, S. S. A., Rodrigues, L. C. C. and Queiroz, P. G. G. (2023). Solucone.me: um sistema responsivo baseado em gamificação para auxiliar o processo de ensino-aprendizagem, apoiado no ENADE. In Anais do XXXIV Simpósio Brasileiro de Informática na Educação (SBIE), p. 707–717. DOI: 10.5753/sbie.2023.234064

- Charão, A. S., Wiechork, K., Rodrigues, M. L. S. and Barbosa, F. P. (2020). Explorando resultados por questão no Enade em Ciência da Computação para subsidiar revisão de projeto pedagógico de curso. In *Anais do XXVIII Workshop sobre Educação em Computação (WEI)*, p. 16–20. DOI: 10.5753/wei.2020.11119
- Collins, L. M. and Lanza, S. T. (2010). *Latent class and latent transition analysis*. Wiley, Hoboken. DOI: 10.1002/9780470567333
- Cunha, R., Sales, C. and Santos, R. (2021). Análise automática com os microdados do ENADE para melhoria do ensino dos cursos de Ciência da Computação. In *Anais do XXIX Workshop sobre Educação em Computação (WEI)*, p. 208–217. DOI: 10.5753/wei.2021.15915
- Damazio, G. N. D. O. and Carvalho, W. V. (2023). Gamificando a preparação para o Enade em cursos superiores de Computação. In *Anais do III Simpósio Brasileiro de Educação em Computação (EDUCOMP) — Workshop de Teses e Dissertações*, p. 44–45. DOI: 10.5753/educomp\_estendido.2023.229094
- de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika*, 76(2):179–199. DOI: 10.1007/s11336-011-9207-7
- Fryer, L. K. (2017). (Latent) transitions to learning at university: a latent profile transition analysis of first-year Japanese students. *Higher Education*, 73(3):519–537. DOI: 10.1007/s10734-016-0094-9
- Gondran, E., Lucca, G., Borges, E. N. and Santos, H. (2023). A evolução do perfil de estudantes de Engenharia com bom desempenho no ENADE. In *Anais do XXXIV Simpósio Brasileiro de Informática na Educação (SBIE)*, p. 1238–1248. DOI: 10.5753/sbie.2023.234252
- Greenacre, M. (2017). *Correspondence analysis in practice*. CRC Press, 3rd edition. DOI: 10.1201/9781315369983
- Hennig, C. (2007). Cluster-wise assessment of cluster stability. *Computational Statistics & Data Analysis*, 52(1):258–271. DOI: 10.1016/j.csda.2006.11.025
- Henson, R. A., Templin, J. L. and Willse, J. T. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika*, 74(2):191–210. DOI: 10.1007/s11336-008-9089-5
- Hickendorff, M., van Putten, C. M., Verhelst, N. D. and Heiser, W. J. (2010). Individual differences in strategy use on division problems: mental versus written computation. *Journal of Educational Psychology*, 102(2):438–452. DOI: 10.1037/a0018177
- Hickendorff, M., Edelsbrunner, P. A., McMullen, J., Schneider, M. and Trezise, K. (2018). Informative tools for characterizing individual differences in learning: latent class, latent profile, and latent transition analysis. *Learning and Individual Differences*, 66:4–15. DOI: 10.1016/j.lindif.2017.11.001
- Hsu, C.-L., Jin, K.-Y. and Chiu, M. M. (2020). Cognitive diagnostic models for random guessing behaviors. *Frontiers in Psychology*, 11:570365. DOI: 10.3389/fpsyg.2020.570365
- Hubert, L. and Arabie, P. (1985). Comparing partitions. *Journal of Classification*, 2(1):193–218. DOI: 10.1007/BF01908075
- INEP (2023). *Relatório-síntese ENADE 2023*. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira.

- Junker, B. W. and Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, 25(3):258–272. DOI: 10.1177/01466210122032064
- Leão, J., Araújo, A., Souza, B. M. F., Araújo, J. F., Ferreira, K., Barros, M., Lauschner, T. and Nakamura, F. G. (2021). Uma análise regional sobre o perfil de estudantes concluintes em cursos STEAM através do ENADE. In *Anais do XV Women in Information Technology (WIT)*, p. 200–209. DOI: 10.5753/wit.2021.15871
- Lima, P., Ambrósio, A., Brancher, J. and Felix, I. (2018). SysEnade: análise das questões de provas do Enade organizadas pelos temas abordados. In *Anais dos Workshops do Congresso Brasileiro de Informática na Educação (CBIE)*, v. 7, p. 419. DOI: 10.5753/cbie.wcbie.2018.419
- March, S. T. and Smith, G. F. (1995). Design and natural science research on information technology. *Decision Support Systems*, 15(4):251–266. DOI: 10.1016/0167-9236(94)00041-2
- Mendonça, N. C. (2024). Evaluating ChatGPT-4 Vision on Brazil’s national undergraduate Computer Science exam. *ACM Transactions on Computing Education*, 24(3), Article 37. DOI: 10.1145/3674832
- Murtagh, F. and Legendre, P. (2014). Ward’s hierarchical agglomerative clustering method: which algorithms implement Ward’s criterion? *Journal of Classification*, 31(3):274–295. DOI: 10.1007/s00357-014-9161-z
- Nylund, K. L., Asparouhov, T. and Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling. *Structural Equation Modeling*, 14(4):535–569. DOI: 10.1080/10705510701575396
- Nylund-Gibson, K. and Choi, A. Y. (2018). Ten frequently asked questions about latent class analysis. *Translational Issues in Psychological Science*, 4(4):440–461. DOI: 10.1037/tps0000176
- Rezende, C. C. S., Cantarino, L. A. B., Souza, P. F., Alves, T. O. M. and Campos, R. S. (2022). O impacto de aspectos socioeconômicos no desempenho de estudantes de Sistemas de Informação no Enade. *Revista Brasileira de Informática na Educação*, 30:157–181. DOI: 10.5753/rbie.2022.30.0.157
- Ribeiro, T. C., Córdova Júnior, W. F. and Castello Branco Neto, W. (2022). Plataforma para auxílio na preparação de estudantes para as avaliações do ENADE e POSCOMP. In *Anais do II Simpósio Brasileiro de Educação em Computação (EDUCOMP)*, p. 243–250. DOI: 10.5753/educomp.2022.19224
- Romero, C. and Ventura, S. (2020). Educational data mining and learning analytics: an updated survey. *WIREs Data Mining and Knowledge Discovery*, 10(3):e1355. DOI: 10.1002/widm.1355
- Rupp, A. A., Templin, J. L. and Henson, R. A. (2010). *Diagnostic measurement: theory, methods, and applications*. Guilford Press, New York.
- Sakashita, R. G. A., Bertolini, D. and Schwerz, A. L. (2023). Predição do conceito ENADE dos cursos de Computação no Brasil. In *Anais do XXXIV Simpósio Brasileiro de Informática na Educação (SBIE)*, p. 1305–1316. DOI: 10.5753/sbie.2023.234268
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464. DOI: 10.1214/aos/1176344136

- Siemens, G. (2013). Learning analytics: the emergence of a discipline. *American Behavioral Scientist*, 57(10):1380–1400. DOI: 10.1177/0002764213498851
- Silva, J. A., Pedreira, K. S. and Alves, T. O. M. (2025). Avalia: ferramenta baseada em inteligência artificial e dados do Enade para análise curricular do curso de Sistemas de Informação. In *Anais do XXI Simpósio Brasileiro de Sistemas de Informação (SBSI) — Anais Estendidos*, p. 122–131. DOI: 10.5753/sbsi\_estendido.2025.4905
- Souza, K. E., Miranda Neto, M. and Lana, C. A. (2024). Classificação curricular das questões do ENADE em Engenharia de Computação: uma mineração de texto. In *Anais do XXXV Simpósio Brasileiro de Informática na Educação (SBIE)*, p. 3009–3019. DOI: 10.5753/sbie.2024.243981
- Templin, J. L. and Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods*, 11(3):287–305. DOI: 10.1037/1082-989X.11.3.287
- Tjoe, H. and de la Torre, J. (2022). The G-DINA model framework: a review and applications in cognitive diagnostic assessment. *Educational Measurement: Issues and Practice*, 41(1):21–36. DOI: 10.1111/emip.12463
- Vieira, A. S., Bertolini, D. and Schwerz, A. L. (2022). Análise do desempenho no Enade dos concluintes de Computação usando a técnica de agrupamento. In *Anais do XXXIII Simpósio Brasileiro de Informática na Educação (SBIE)*, p. 834–845. DOI: 10.5753/sbie.2022.225528
- Weller, B. E., Bowen, N. K. and Faubert, S. J. (2020). Latent class analysis: a guide to best practice. *Journal of Black Psychology*, 46(4):287–311. DOI: 10.1177/0095798420930932
- Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B. and Wesslén, A. (2012). *Experimentation in Software Engineering*. Springer, Berlin Heidelberg. DOI: 10.1007/978-3-642-29044-2