Latent Profiles of School Trajectory Risk and Sociodemographic Patterning in Brazilian Public Education

Ana Luisa Gomes De Barros Freitas[1], Luan Filipy Freire Torres[1], Julio Cezar Albuquerque Da Costa[2], Leogildo Freires[1], Ane Mayra Melo Silva[1], Heitor Araújo[3]

[1]Universidade Federal de Alagoas, Brazil
[2]Universidade Federal de Minas Gerais, Brazil
[3]Universidade Federal da Bahia, Brazil

Abstract

School dropout and educational disengagement are increasingly conceptualized as cumulative and relational processes shaped by the interplay of institutional, familial, relational, and structural factors, requiring multidimensional approaches that account for both the configurations of risk and their social patterning. This study aimed to identify latent profiles of school trajectory risk among Brazilian public school students and to examine how these profiles are sociodemographically patterned across gender, race/ethnicity, socioeconomic status, and student labor participation. Accordingly, a sample of 10,000 synthetic data was used, based on valid responses/answers from students of four Brazilian states (Rondônia, Minas Gerais, Mato Grosso, and Maranhão) who completed the Relational Factors for the Risk of School Dropout Scale – Alternative version (IAFREE-A).The synthetic data was calculated using the Gaussian Copula technique, which considers the association between the observed variables in the dataset to estimate new responses, generating a new and bigger dataset. This approach was considered due to the General Law of Data Protection (LGPD) and the need to ensure the protection of participating children and adolescents. Latent Profile Analysis (LPA) was conducted using the five aggregated dimensions of the IAFREE-A — student–student, student–school, student–school-professionals, student–family, and student–community. Model [1] selection followed an analytic hierarchy process based on multiple fit indices (Akogul & Erisoglu, 2017). The best-fitting solution was a three-class model (BIC = 59,332; entropy = 0.79; minimum classification probability = 0.88), with profiles interpreted as low (27.2%), medium (53.9%), and high risk (18.9%), showing monotonic separation across all five dimensions, with high-risk students presenting elevated scores in the student–school (M = 2.98) and student–community (M = 2.91) domains. Subsequently, chi-square analyses examined associations between profile membership and sociodemographic variables. A statistically significant association emerged for gender ($\chi^2 = 118.47$; $p < .001$; $V = 0.077$): female students were over-represented in the high-risk profile (adjusted residual = +8.91), whereas male students concentrated in the low-risk profile (+8.02). Associations with socioeconomic level ($V = 0.033$), student employment ($V = 0.022$), and race/ethnicity ($V = 0.007$) were weak and non-significant. Findings indicate clear, well-separated multidimensional profiles of school trajectory risk and a robust gendered patterning of vulnerability, supporting a process-oriented understanding of school trajectory risk under LGPD-compliant research conditions.

Keywords: *latent profile analysis; school trajectories; educational inequality; Gaussian copula*

Introduction

School dropout is currently understood as a processual and multifactorial phenomenon, although debates persist regarding which factors influence this process and the extent of their effects (Rumberger, 2011; Sabates et al., 2010). The literature also presents conceptual divergences concerning the meaning of the term "school dropout." In the present study, the definition adopted follows the criteria established by the National Institute for Educational Studies and Research (INEP), which distinguishes between dropout (*evasão escolar*) and school abandonment (*abandono escolar*). According to INEP, dropout refers to a transition rate indicating the absence of enrollment in any school in the year following the school census, whereas school abandonment refers to a school-flow indicator that captures the interruption of attendance after the census reference date (INEP, 2017). These phenomena should not be interpreted as definitive conditions by nature, since they may represent temporary interruptions within educational trajectories and may therefore be reversible.

Although this distinction introduces analytic nuances, dropout and abandonment still correspond to specific moments within broader educational trajectories. In this context, several Latin American countries have increasingly incorporated the notion of school trajectories into public educational policies, conceptualizing dropout and abandonment as possible events within these trajectories rather than as their totality (Terigi & Briscioli, 2020). This perspective is associated with the concept of trajectory protection, which shifts the focus from the prevention of isolated events toward the promotion of conditions that sustain students' educational continuity through institutional, personal, and contextual support mechanisms during the schooling process (Fuentes, 2024).

Studies indicate that weak school belonging, negative peer interactions, low perceived support from teachers, family disengagement, and community vulnerability are associated with indicators of school disengagement, absenteeism, low academic investment, and dropout intention (Archambault et al., 2009; Wang & Degol, 2013; Lessard et al., 2008; Henry et al., 2012). These findings support the proposition that school trajectory risk emerges through the articulation of multiple relational dimensions rather than through isolated predictors (Bronfenbrenner, 1979; Rumberger, 2011).

A partir disso, o Instrument for Assessing Risk Factors for School Dropout - Alternative version (IAFREE-A) foi criado com base no Modelo Relacional (Vasconcelos et al., 2023) , integrando dimensões de proteção e fatores de atenção de vulnerabilidade ao abandono e à evasão escolar em uma única ferramenta, permitindo a mensuração e previsão do intricado sistema de relações que impactam estes fenômenos, proporcionando uma visão abrangente das experiências dos(as) estudantes no contexto educacional (Wood et al., 2017; Kearney, 2021). O Instrumento de Avaliação dos Fatores de Risco à Evasão Escolar - Alternativo (IAFREE-A) engloba 14 fatores de atenção distribuídos em cinco dimensões essenciais: Estudante-Escola, Estudante-Profissionais da escola, Estudante-Família, Estudante-Comunidade e Estudante-Estudantes (Vasconcelos et al., 2023).

In the Brazilian public education context, school trajectories are shaped by structural inequalities associated with territorial disparities, socioeconomic precariousness, institutional instability, and unequal access to educational resources (Ribeiro, 2011; Ferrão, 2022). Although dropout rates have declined in recent decades, indicators of school disengagement and interrupted trajectories remain concentrated among students exposed to conditions of social vulnerability (UNICEF, 2021; Ramos et al., 2026). Research in this field has also examined the role of sociodemographic variables, such as gender, race/ethnicity, socioeconomic level, and adolescent labor participation, in the distribution of educational risk (Barbosa et al., 2024; Menezes, 2023). However, findings remain inconsistent, particularly when relational and contextual dimensions are incorporated into the analytic models (Fortes et al., 2024).

Most studies investigating school dropout employ variable-centered approaches, such as regression models and correlational analyses, which estimate the independent contribution of predictors to educational outcomes (Morin et al., 2018; Wang & Degol, 2013). Although these approaches provide evidence regarding associations between variables, they are limited in identifying heterogeneous subgroups characterized by distinct configurations of risk (Bergman & Magnusson, 1997; Howard & Hoffman, 2018). Person-centered approaches, such as Latent Profile Analysis (LPA), enable the identification of latent subpopulations based on response patterns across multiple indicators (Oberski, 2016; Spurk et al., 2020). In the context of school trajectory research, LPA allows the estimation of qualitatively distinct profiles of relational vulnerability, contributing to the understanding of how different dimensions co-occur within individuals (Wang et al., 2019; Archambault et al., 2009).

Therefore, the aim of this study is to identify latent profiles of school trajectory risk among Brazilian public school students and to examine how these profiles are sociodemographically patterned across gender, race/ethnicity, socioeconomic status, and student labor participation.

Method

Design

This is a quantitative, cross-sectional, non-experimental study conducted on a synthetic dataset. The research combined a person-centered approach, through Latent Profile Analysis (LPA), with a group-comparison approach, through chi-square association tests, with the aim of identifying latent profiles of risk in school trajectories and examining how these profiles are distributed across students' sociodemographic characteristics.

Ethical Considerations and Generation of the Synthetic Dataset

Given the sensitive nature of educational data involving children and adolescents, and in compliance with the Brazilian General Data Protection Law (LGPD — Law no. 13,709/2018), the analyses were conducted on a synthetic dataset, constructed from the original responses of students to the IAFREE-A instrument. As análises de dados foram realizadas através da linguagem R de programação (R Core Team, 2025). Utilizou-se da técnica de Gaussian Copula para a geração de dados sintéticos. Trata-se de uma técnica de geração de dados sintéticos baseada em distribuições normais multivariadas, por meio de uma transformação integral de probabilidade com os dados reais (Li, Zhao & Fu, 2020). Uma das vantagens dessa técnica é que sua utilização leva em conta a dependência (ou associação) entre as variáveis (Ahmadian et al., 2024) de um banco de dados, algo necessário para a utilização de dados sintéticos para o uso com variáveis psicométricas, uma vez que os fatores latentes derivados dessas medidas são agrupados através das correlações entre os itens (Brown, 2015). Além disso, ressalta-se que a metodologia da Gaussian Copula tem apresentado resultados satisfatórios em reproduzir os dados reais, seja no campo da medicina (Ahmadian et al., 2024; Sichani et al., 2024) ou até para reproduzir padrões comportamentais (Savran & Karpat, 2024).

Para avaliar a reprodutibilidade dos dados sintéticos para com os dados reais, optou-se pela investigação do indicador de distância de Hellinger, que varia entre 0 e 1 e indica o quão próximas ou diferentes são duas distribuições de probabilidade. Para a interpretação da semelhança entre os dados reais, valores mais próximos de 0 indicam maior semelhança, enquanto valores mais próximos de 1 indicam diferenças entre os dados reais e gerados (Mosquera et al., 2023). Nesse estudo, utilizou-se os valores até 0,20 como adequados.

Sociodemographic Variables

The synthetic dataset comprised 10,000 records simulating public school students from four Brazilian states from the 9th grade of Elementary School to the 3rd year of Upper Secondary Education, generated from a real sample of 3536 students. Regarding gender, the sample was balanced between male (n = 4,898; 49.0%) and female students (n = 5,082; 50.8%), with a small proportion identifying as other (n = 20; 0.2%). In terms of race/ethnicity, 56.2% identified as Indigenous or other (n = 5,620), 23.0% as Black or Brown (n = 2,299), and 20.8% as White (n = 2,081). Most students reported not engaging in paid work (n = 7,933; 79.3%), while 20.7% indicated current labor participation (n = 2,067). The distribution across socioeconomic strata (INSE), ranging from I (lowest) to VIII (highest), concentrated in the middle levels (strata III to VI accounted for 73.3% of the sample), with smaller proportions in the extreme strata (3.5% combined in I and VIII).

Instrument

The study used the Instrument for the Assessment of Relational Factors for the Risk of School Dropout - Alternative version (IAFREE-A), a multidimensional scale developed to assess protective and risk factors in school trajectories across five relational dimensions: (a) Student–Student (SSt), which captures interpersonal relationships, social skills, educational expectations, and sense of belonging among peers; (b) Student–School (SSc), which encompasses the perception of school as a safe place and institutional responses; (c) Student–School Professionals (SSP), referring to the relationship with teachers and pedagogical staff; (d) Student–Family (SF), which assesses family support and engagement with the student's school trajectory; and (e) Student–Community (SC), referring to the student's relationship with the surrounding community. Each item is answered on a four-point Likert scale, with higher scores indicating that a higher level of attention is required for this school trajectory.

Analytical Procedures

Analyses were conducted in R software, version 4.4 (R Core Team, 2024), using the tidyLPA package (Rosenberg et al., 2018) for latent profile estimation. Latent Profile Analysis (LPA) was conducted using the mean scores of the five aggregated dimensions of the IAFREE-A (SSt, SSc, SSP, SF, SC) as indicators. Solutions ranging from two to five classes were compared, all estimated under the specification of equal variances between classes and zero covariances (Model 1 of tidyLPA). The selection of the optimal number of classes followed an analytic hierarchy process based on multiple fit indices (Akogul & Erisoglu, 2017), including AIC, AWE, BIC, CAIC, CLC, KIC, and SABIC, complemented by the examination of entropy, minimum and maximum classification probabilities, and the conceptual interpretability of the solutions. The random seed was fixed (set.seed = 2024) to ensure analytic reproducibility.

After selecting the final solution, each participant was assigned to the latent class with the highest posterior probability, and the profiles were labeled as low, medium, and high risk based on the aggregated means of the five dimensions

Subsequently, Pearson's chi-square association tests were conducted to examine the relationship between latent risk profile and each sociodemographic variable. Effect size was estimated using Cramér's V, and post hoc analysis was performed through adjusted standardized residuals, with cells presenting $|z| > 1.96$ ($\alpha = 0.05$) considered statistically significant. Across all contingency tables, more than 80% of cells presented expected frequencies above 5, meeting the classical Cochran criterion for the validity of the chi-square asymptotic approximation. The only exception was the cell corresponding to the intersection between the high-risk profile and the "other" gender category (expected frequency = 3.78), which represents 0.2% of the sample and does not substantively affect the inferences of the test.

Results

Selection of the Number of Latent Classes

Table 1 presents the fit indices for solutions ranging from two to five classes. The three-class solution was selected as the most adequate, based on the combination of statistical fit and conceptual interpretability. Although the BIC continued to decrease in solutions with more classes, entropy peaked at three classes (0.785), and the minimum classification probability remained high (0.880), indicating good separation between profiles. The substantive interpretation was also clearer with three classes, with profiles monotonically ordered along increasing levels of risk.

Table 1. Fit indices for solutions from two to five classes (Model 1: equal variances, zero covariances)

Classes	LogLik	AIC	BIC	SABIC	Entropy	Min. Prob.
2	−31,849	63,730	63,845	63,795	0.775	0.923
3	−29,565	59,173	59,332	59,262	0.785	0.880
4	−28,751	57,558	57,760	57,671	0.768	0.826
5	−28,425	56,918	57,163	57,055	0.749	0.774

Characterization of the Latent Profiles

The three-class solution yielded clearly differentiated and monotonically ordered profiles across the five dimensions of the IAFREE-A. The low-risk profile comprised 27.2% of the sample (n = 2,715), with means below 1.80 across all dimensions. The medium-risk profile was the most populous, comprising 53.9% of the sample (n = 5,393), with means between 2.07 and 2.32. The high-risk profile, comprising 18.9% of the sample (n = 1,892), presented the highest means, particularly in the Student–School (M = 2.98) and Student–Community (M = 2.91) dimensions. Table 2 presents the profile means in each dimension.

Table 2. Means of the IAFREE-A dimensional scores by latent profile

Profile	n (%)	SSc	SSP	SF	SC	SSt	Overall mean
Low risk	2,715 (27.2%)	1.66	1.73	1.65	1.79	1.54	1.67
Medium risk	5,393 (53.9%)	2.31	2.25	2.09	2.32	2.07	2.21
High risk	1,892 (18.9%)	2.98	2.82	2.55	2.91	2.64	2.78

The pattern shown in Figure 1 reveals monotonic separation between the three profiles across all dimensions, with progressively greater distances between scores as the level of risk increases. The

Student–Family dimension presented the lowest mean scores across all profiles, while Student–School and Student–Community presented the highest mean scores, suggesting that institutional and community dimensions contribute proportionally more to the configuration of risk across all levels.

Figure 1. Latent profile means across the five IAFREE-A dimensions

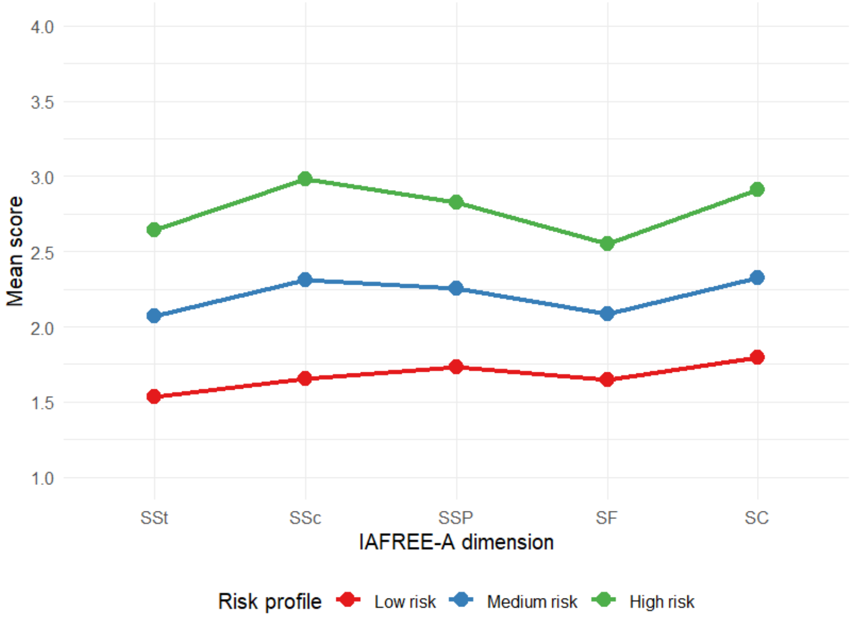
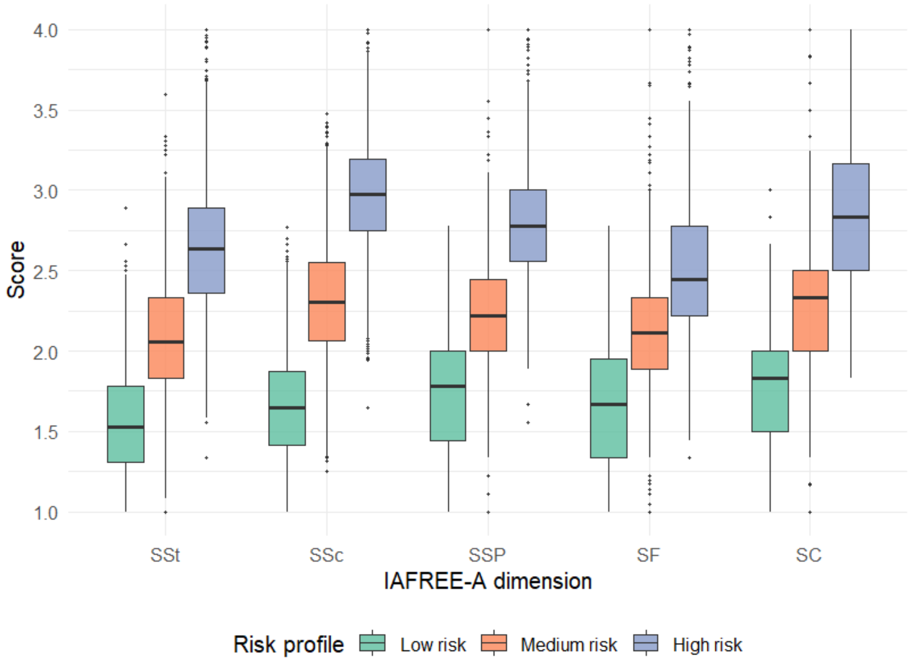


Figure 2. Distribution of IAFREE-A dimensions by latent risk profile



Associations between Latent Profiles and Sociodemographic Variables

Pearson's chi-square tests revealed only one statistically significant association between the risk profile and the sociodemographic variables examined. Table 3 summarizes the results.

Table 3. Chi-square tests between latent profile and sociodemographic variables

Variable	χ^2	df	p	Cramér's V	N
Gender	118.47	4	< 0.001	0.077	10,000
Labor participation	4.90	2	0.087	0.022	10,000
Race/ethnicity	0.98	4	0.913	0.007	10,000
Socioeconomic level (INSE)	21.25	14	0.095	0.033	10,000

Gender

The analysis revealed a significant association between gender and risk profile ($\chi^2(4) = 118.47$; $p < 0.001$; $V = 0.077$). Although Cramér's V indicates a small effect size (which is expected given the large sample size), the adjusted standardized residuals revealed a clear pattern of differential distribution. Male students were significantly over-represented in the low-risk profile ($z = +8.02$) and under-represented in the high-risk profile ($z = -9.13$). Conversely, female students were over-represented in the high-risk profile ($z = +8.91$) and under-represented in the low-risk profile ($z = -7.91$). Table 4 details the frequencies and percentages by category.

Table 4. Percentage distribution of gender by risk profile

Profile	Male	Female	Other
Low risk	55.5%	44.4%	0.1%
Medium risk	49.0%	50.8%	0.2%
High risk	39.5%	60.0%	0.4%

Labor Participation, Race/Ethnicity, and Socioeconomic Level

The associations between risk profile and labor participation ($\chi^2(2) = 4.90$; $p = 0.087$; $V = 0.022$), race/ethnicity ($\chi^2(4) = 0.98$; $p = 0.913$; $V = 0.007$), and socioeconomic level ($\chi^2(14) = 21.25$; $p = 0.095$; $V = 0.033$) did not reach statistical significance. Cramér's V values were all below 0.05, indicating associations of negligible magnitude. Inspection of the contingency tables confirms the near-equivalence of distributions across profiles: for example, the proportion of students who work varies minimally across profiles (19.8% in the low-risk profile, 20.5% in the medium-risk profile, and 22.4% in the high-risk profile), as does the distribution by race/ethnicity (≈21% White, ≈23% Black/Brown, and ≈56% Indigenous/Other across all profiles) and by socioeconomic stratum.

The absence of substantive associations between these sociodemographic variables and the latent profiles, in contrast with the robust finding for gender, suggests that, in the synthetic dataset analyzed, the relational risk structures captured by the five dimensions of the IAFREE-A are distributed relatively homogeneously across socioeconomic, racial, and labor strata, but display a clear pattern of differentiation by gender.

Discussion

The Latent Profile Analysis identified three profiles that were subsequently categorized according to risk level. The low-risk profile presented more homogeneous scores across the IAFREE-A dimensions, whereas higher-risk profiles maintained a similar dimensional configuration but with greater score dispersion. This pattern suggests a cumulative and processual organization of school trajectory vulnerability, indicating that school trajectory risk should not be conceptualized as a dichotomous phenomenon, but rather as a continuum in which the relative contribution and intensity of relational dimensions vary according to the latent profile configuration (Bergman & Magnusson, 1997; Rumberger, 2011). The monotonic separation observed across the profiles also supports the interpretation that different relational domains co-occur in the constitution of educational vulnerability rather than operating as isolated predictors.

Within this configuration, the prominence of the Student–School and Student–Community dimensions in the high-risk profile is notable, indicating greater fragility in students' relationships with the school institution and with the communities in which they are embedded. From an intervention perspective, these findings suggest the relevance of investigating institutional and territorial conditions associated with school trajectories. Based on the literature and on the relational indicators assessed by the instrument, these vulnerabilities may involve perceptions of insecurity within school and community contexts, institutional exclusion, precarious infrastructure, and limited articulation between educational systems and health and social assistance services (Bronfenbrenner, 1979; Wang & Degol, 2013; Rumberger, 2011). In contrast, the Student–Family dimension presented the lowest scores across profiles, suggesting that family relationships remained comparatively more preserved even among high-risk students. This result indicates that family support may function as a protective factor contributing to school permanence up to the moment of data collection.

The chi-square and post hoc analyses also identified a small but systematic overrepresentation of male students in the low-risk profile and female students in the high-risk profile. In conjunction with the relational profile structure, these findings suggest that gendered experiences may differentially shape school trajectories and perceptions of institutional and contextual vulnerability. The observed gender pattern differs from studies reporting higher prevalence of school dropout among male students, particularly in contexts involving labor participation and school interruption. However, the present study operationalized school trajectory risk through relational indicators rather than completed dropout events. In this sense, female students may report higher levels of relational strain, institutional dissatisfaction, emotional burden, or perceived social vulnerability without necessarily presenting higher rates of school abandonment. The findings therefore reinforce the distinction between dropout occurrence and latent configurations of educational vulnerability.

Some limitations should be considered. The cross-sectional design restricts causal inference regarding the development of school trajectory risk over time. The use of synthetic data, although justified by ethical and legal requirements, may attenuate specific distributional characteristics present in the original sample. In addition, the analyses were restricted to students from four Brazilian states, limiting the generalizability of the findings to other educational contexts. Future studies may investigate longitudinal transitions between latent profiles, examine measurement invariance across sociodemographic groups, and incorporate additional indicators such as academic performance, mental health, and school attendance.

Conclusion

The present study identified three latent profiles of school trajectory risk among Brazilian public school students based on the relational dimensions assessed by the IAFREE-A. The profiles demonstrated monotonic separation across the five domains, indicating differentiated configurations of relational vulnerability. The findings support the conceptualization of school trajectory risk as a multidimensional construct involving institutional, interpersonal, familial, and community dimensions.

The results also indicated that gender was the only sociodemographic variable associated with latent profile membership. Female students were disproportionately represented in the high-risk profile, whereas male students concentrated in the low-risk profile. In contrast, race/ethnicity, socioeconomic level, and labor participation did not present significant associations with relational risk configurations in the analyzed dataset.

From a psychometric and methodological standpoint, the study demonstrates the applicability of Latent Profile Analysis for identifying heterogeneous patterns of educational vulnerability and supports the use of multidimensional relational indicators in the assessment of school trajectory risk. The use of synthetic data generated through Gaussian Copula procedures also contributes to the discussion regarding ethical data management in educational research involving minors under LGPD regulations.

The findings reinforce the relevance of multidimensional approaches to school trajectory analysis and indicate that prevention strategies should incorporate relational and contextual dimensions associated with school permanence. Interventions centered exclusively on academic performance indicators may not capture latent configurations of vulnerability associated with educational disengagement. Research is already being conducted on a wider sample encompassing students from all Brazilian states and can help to investigate how the tendencies observed in the present study vary on a more diverse and representative sample.

REFERENCES

[1] Ahmadian, M., Bodalal, Z., van der Hulst, H. J., Vens, C., Karssemakers, L. H., Bogveradze, N., ... & Castelijns, J. A. (2024). Overcoming data scarcity in radiomics/radiogenomics using synthetic radiomic features. Computers in Biology and Medicine, 174, 108389.

[2] Akogul, S., & Erisoglu, M. (2017). An approach for determining the number of clusters in a model-based cluster analysis. Entropy, 19(9), Article 452. https://doi.org/10.3390/e19090452

[3] Archambault, I., Janosz, M., Fallu, J.-S., & Pagani, L. S. (2009). Student engagement and its relationship with early high school dropout. *Journal of Adolescence, 32*(3), 651–670. https://doi.org/10.1016/j.adolescence.2008.06.007

[4] Barbosa, G., França, M., & Duque, D. (2024). Challenges in Brazilian education: Are Black students falling behind? ANPEC Nordeste.

[5] Bergman, L. R., & Magnusson, D. (1997). A person-oriented approach in research on developmental psychopathology. *Development and Psychopathology, 9*(2), 291–319. https://doi.org/10.1017/S095457949700206X

[6] Bronfenbrenner, U. (1979). The ecology of human development: Experiments by nature and design. Harvard University Press.

[7] Ferrão, M. E. (2022). The evaluation of students' progression in lower secondary public education in Brazil. *Evaluation and Program Planning, 93*, 102096.

[8] Fortes, G., Nunes , A., Albuquerque , J. C., Lopes , G. D., Freires , L. F., Soares , L., … Isotani, S. (2024). Exploring the Unexpected: the Relationship Between Higher Family Income and Dropout Risk. Praxis & Saber, 15(41), 1–16. https://doi.org/10.19053/uptc.22160159.v15.n41.2024.16848

[9] Fuentes, S. (2024). *Trayectorias educativas y políticas de protección escolar en América Latina*. Buenos Aires: CLACSO.

[10] Henry, K. L., Knight, K. E., & Thornberry, T. P. (2012). School disengagement as a predictor of dropout, delinquency, and problem substance use during adolescence and early adulthood. Journal of Youth and Adolescence, 41(2), 156–166. https://doi.org/10.1007/s10964-011-9665-3

[11] Howard, M. C., & Hoffman, M. E. (2018). Variable-centered, person-centered, and person-specific approaches: Where theory meets the method. *Organizational Research Methods, 21*(4), 846–876. https://doi.org/10.1177/1094428117744021

[12] INEP. (2017). *Notas estatísticas do Censo Escolar da Educação Básica*. Brasília: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira.

[13] Lessard, A., Poirier, M., & Fortin, L. (2008). Student-teacher relationship and teacher support in students at risk of dropping out of school. Canadian Journal of Education, 31(4), 893–919.

[14] Li, Z., Zhao, Y., & Fu, J. (2020). Sync: A copula based framework for generating synthetic data from aggregated sources. In 2020 International Conference on Data Mining Workshops (ICDMW) (pp. 571-578). IEEE. http://doi.org/10.1109/ICDMW51313.2020.0008

[15] Menezes, V. M. O. (2023). Juventude, educação e trabalho no Brasil (2012–2022). *Trabalho & Sociedade*.

[16] Morin, A. J. S., Bujacz, A., & Gagné, M. (2018). Person-centered methodologies in the organizational sciences. *Organizational Research Methods, 21*(4), 803–813. https://doi.org/10.1177/1094428118773856

[17] Mosquera, L., El Emam, K., Ding, L., Sharma, V., Zhang, X. H., Kababji, S. E., ... & Eurich, D. T. (2023). A method for generating synthetic longitudinal health data. BMC Medical Research Methodology, 23(1), 67.

[18] Oberski, D. (2016). Mixture models: Latent profile and latent class analysis. In J. Robertson & M. Kaptein (Eds.), *Modern statistical methods for HCI* (pp. 275–287). Springer. https://doi.org/10.1007/978-3-319-26633-6_12

[19] R Core Team. (2024). R: A language and environment for statistical computing (Version 4.4) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

[20] Ramos, M., Cordeiro, T., Sobrinho, Á., & Pimentel, B. A. (2026). *The challenge of school dropout in Brazil: An analysis of over two decades*. Revista Brasileira de Informática na Educação.

[21] Ribeiro, C. A. C. (2011). Trends in educational inequality in Brazil. *Dados*, 54(1), 41–87.

[22] Rosenberg, J. M., Beymer, P. N., Anderson, D. J., Van Lissa, C. J., & Schmidt, J. A. (2018). tidyLPA: An R package to easily carry out latent profile analysis (LPA) using open-source or

commercial software. Journal of Open Source Software, 3(30), Article 978. https://doi.org/10.21105/joss.00978

[23] Rumberger, R. W. (2011). Dropping out: Why students drop out of high school and what can be done about it. Harvard University Press.

[24] Sabates, R., Hossain, A., & Lewin, K. M. (2010). School dropout in Bangladesh: New insights from longitudinal evidence. Compare: A Journal of Comparative and International Education, 40(1), 33–47. https://doi.org/10.1080/03057920903125664

[25] Sichani, E. K., Smith, A., El Emam, K., & Mosquera, L. (2024). Creating high-quality synthetic health data: Framework for model development and validation. JMIR formative research, 8(1), e53241.

[26] Savran, E., & Karpat, F. (2024). Synthetic data generation using Copula model and driving behavior analysis. Ain Shams Engineering Journal, 15(12), 103060.

[27] Spurk, D., Hirschi, A., Wang, M., Valero, D., & Kauffeld, S. (2020). Latent profile analysis: A review and "how to" guide of its application within vocational behavior research. *Journal of Vocational Behavior, 120*, 103445. https://doi.org/10.1016/j.jvb.2020.103445

[28] Terigi, F., & Briscioli, B. (2020). Políticas de escolarización y protección de trayectorias educativas en América Latina. Revista Latinoamericana de Educación Comparada, 11(17), 45–62

[29] UNICEF. (2021). *Out-of-school children in Brazil: A warning about the impacts of the COVID-19 pandemic on education*. UNICEF Brazil.

[30] Wang, M.-T., & Degol, J. L. (2013). School climate: A review of the construct, measurement, and impact on student outcomes. *Educational Psychology Review, 25*(3), 357–385. https://doi.org/10.1007/s10648-013-9230-9

[31] Wang, M.-T., Kiuru, N., Degol, J. L., & Salmela-Aro, K. (2019). Friends, academic achievement, and school engagement during adolescence: A social network approach to peer influence and selection effects. *Learning and Instruction, 58*, 148–160. https://doi.org/10.1016/j.learninstruc.2018.06.003