



Latent Profiles of Relational Risk in School Trajectories: An Analysis of Gaussian Copula-Synthesized Data from Brazilian Public School Students

Luan Filipy Freire Torres¹, Leogildo Freires¹, Ana Luisa Gomes de Barros Freitas¹,
Julio Cezar Albuquerque da Costa², Ane Mayra Melo Silva¹, Heitor Araújo³

¹Universidade Federal de Alagoas, Brazil

²Universidade Federal de Minas Gerais, Brazil

³Universidade Federal da Bahia, Brazil

Abstract

Peer relationships and school belonging are central relational dimensions shaping students' school trajectories, particularly in contexts of structural vulnerability. This study aimed to identify latent profiles of relational risk among Brazilian public school students and to examine whether these profiles represent empirically distinct groups, combining person-centered and group-comparison approaches. Given the sensitive nature of educational data involving minors and the requirements of the Brazilian General Data Protection Law (LGPD), analyses were conducted on a synthetic dataset ($N = 10,000$) generated via a Gaussian Copula procedure, simulating responses from public school students from the 9th grade to the 3rd year of upper secondary education in four Brazilian states. Latent Profile Analysis (LPA) was conducted using the three factors of the Student–Student (SSt) dimension: interpersonal relations and social skills, educational expectations, and belonging/identification. Model selection followed an analytic hierarchy process based on multiple fit indices. The best-fitting solution was a three-class model ($BIC = 52,390$; entropy = 0.65), with profiles interpreted as low (30.7%), medium (49.7%), and high relational risk (19.6%). Kruskal–Wallis analyses showed significant differences between all classes for the SSt total score ($\chi^2 = 7,702.60$; $p < .001$; $\epsilon^2 = 0.770$), interpersonal relations ($\epsilon^2 = 0.687$), and belonging ($\epsilon^2 = 0.654$), indicating very large effects. Educational expectations differed less ($\epsilon^2 = 0.081$). Profiles also differed in the student–school dimension ($\epsilon^2 = 0.359$), particularly perception of school as a safe place ($\epsilon^2 = 0.318$), and across student–school-professionals ($\epsilon^2 = 0.189$), student–community ($\epsilon^2 = 0.171$), and student–family relations ($\epsilon^2 = 0.150$). Post hoc Dunn comparisons confirmed differences between all groups. Findings indicate well-separated latent profiles of relational risk, supporting a process-oriented understanding of school trajectory risk under LGPD-compliant research conditions.

Keywords: latent profile analysis; school trajectories; peer relations; Gaussian copula; LGPD.

1. Introduction

School dropout and school leaving constitute multifaceted phenomena that transcend the mere interruption of the formal educational trajectory, resulting from the interweaving of socioeconomic, cultural, and institutional variables that go beyond purely individual and familial circumstances [25]. Beyond the pedagogical domain, these processes constitute a public health issue with substantial economic impact [3, 14, 28], since the lack of formal qualifications frequently predisposes individuals to unemployment and intensifies social inequalities [9, 27]. Accordingly, the present study aims to identify latent profiles of relational risk among Brazilian public school students based on the factors of the Student–Student dimension of the IAFREE-A, and to examine whether these profiles differentiate significantly across the remaining dimensions of the instrument.

The circumstances surrounding this phenomenon are intricate and interconnected, ranging from socioeconomic factors — such as the early entry into the labor market to supplement family income — to pedagogical issues, including unmotivating teaching methods or the absence of adequate support for struggling students [4, 16, 21, 23]. Problems in the school environment, such as discrimination, violence, or the perception that the institution offers no promising future, also contribute significantly [12]. In Brazil, dropout and school leaving are differentiated by the School Census criteria: school leaving refers to withdrawal during the school year, while dropout is identified when the student fails to enroll for the subsequent year [7] — a distinction that is essential for the formulation of more effective public policies [5].

Dropout and school leaving should therefore be understood not as isolated events, but as processes of disconnection from school, characterized by relational and multifactorial dynamics that reveal structural inequalities historically present in Brazilian society [10]. From this perspective, the notion of school trajectory protection gains relevance, proposing a broader interpretation of schooling that considers transitions, articulations, and linkages across systemic, institutional, and individual dimensions [11, 24]. Students' pathways cannot be understood solely from their individual characteristics, as they reflect broader inequalities related to familial, institutional, and educational contexts [22].

Given this complexity, there is a need for psychometrically adequate measures to assess the factors that may protect school trajectories. In this regard, the Instrument for the Assessment of Relational Factors for the Risk of School Dropout (IAFREE) [26] was developed under a relational and multidimensional approach, organized across five dimensions: Student–School, Student–School Professionals, Student–Family, Student–Community, and Student–Student. Subsequent revised and alternative versions were later developed, including the IAFREE-R [8] and the IAFREE-A-46, the latter shifting the emphasis from identifying risk factors toward the perspective of protection and continuity of school trajectories.

2. Method

2.1 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 non-parametric Kruskal–Wallis tests, with the aim of identifying latent profiles of relational risk in school trajectories and examining whether these profiles represent empirically distinct groups across the dimensions of the IAFREE-A.

2.2 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 3536 original responses of students to the IAFREE-A instrument. The Gaussian Copula technique was used for the generation of synthetic data. It is a synthetic data generation technique based on multivariate normal distributions, applied through a probability integral transformation of the real data [13]. One of the advantages of this technique is that it accounts for the dependence (or association) between the variables [2] of a dataset, a feature that is necessary for the use of synthetic data with psychometric variables, since the latent factors derived from such measures are grouped through the correlations between items [6]. Furthermore, the Gaussian Copula methodology has shown satisfactory results in reproducing real data, both in the medical field [2, 20] and in reproducing behavioral patterns [19].

To assess the reproducibility of the synthetic data in relation to the real data, the Hellinger distance indicator was used, which ranges from 0 to 1 and indicates how close or different two probability distributions are. For the interpretation of similarity between the real and synthetic data, values closer to 0 indicate greater similarity, while values closer to 1 indicate greater differences between the original and generated data [15]. In this study, values up to 0.20 were considered adequate.

2.3 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 designed to assess protective and risk-related factors in school trajectories across five relational domains: (a) Student–Student (SSt), (b) Student–School (SSc), (c) Student–School Professionals (SSP), (d) Student–Family (SF), and (e) Student–Community (SC). Each item is answered on a four-point Likert scale, with higher scores indicating higher levels of relational risk. The Student–Student dimension, which constitutes the central focus of the present study, comprises three factors: interpersonal relations and social skills (SSSt1), educational expectations (SSSt2), and belonging/identification (SSSt3).

2.4 Analytical Procedures

Analyses were conducted in R software, version 4.4 [17], using the tidyLPA package [18] for latent profile estimation. LPA was conducted using the three factors of the Student–Student dimension (SS1, SS2, SS3) 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 [1], including AIC, AWE, BIC, CAIC, CLC, KIC, and SABIC, complemented by the examination of entropy, 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 relational risk, based on the means of the three SS factors — an ordering justified by the fact that higher IAFREE-A scores indicate greater relational risk. Subsequently, non-parametric Kruskal–Wallis tests were conducted to examine differences between latent classes in the SS total score and its three factors, as well as in the remaining dimensions of the IAFREE-A (SSc, SSP, SF, SC) and their respective sub-factors. The non-parametric approach was adopted given the ordinal nature of the response scale and the non-normal distribution of the dimensional scores. Effect size was estimated using epsilon-squared (ϵ^2), with values of .01, .08, and .26 interpreted as small, moderate, and large effects, respectively. Post hoc analyses were performed using Dunn’s test with Bonferroni correction for multiple comparisons.

3. Results

3.1 Sample Characterization

The synthetic dataset comprised 10,000 records simulating public school students from four Brazilian states. 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%). Regarding race/ethnicity, the composition reflected the typical profile of Brazilian public schools: 54.9% identified as Brown (*Parda*), 21.7% as White, and 20.6% as Black, with the remaining 2.8% comprising Asian, Indigenous, and undeclared students). 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) 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).

3.2 Selection of the Number of Latent Classes

Table 1 presents the fit indices for solutions ranging from two to five classes. Although models with four and five classes presented progressively lower BIC values, indicating better statistical fit, the three-class solution was selected as the most adequate based on the combination of parsimony, conceptual interpretability, and coherence with the theoretical framework of relational risk, which is traditionally operationalized along three levels (low, medium, and high). The three-class model presented adequate classification quality (entropy = 0.652; minimum classification probability = 0.796), with profiles monotonically ordered along increasing levels of relational 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.	Max. Prob.
2	-26,692	53,403	53,476	53,444	0.669	0.885	0.913
3	-26,131	52,289	52,390	52,346	0.652	0.796	0.845
4	-25,955	51,946	52,076	52,019	0.758	0.776	0.881
5	-25,734	51,512	51,671	51,601	0.716	0.531	0.894

3.3 Characterization of the Latent Profiles

The three-class solution yielded clearly differentiated profiles, monotonically ordered across the three factors of the SS dimension. The low-risk profile comprised 30.7% of the sample ($n = 3,066$), the medium-risk profile 49.7% ($n = 4,969$), and the high-risk profile 19.6% ($n = 1,965$). Table 2 presents the means of the SS total score and its three factors by latent profile.



Table 2. Means of the Student–Student (SSt) dimension and its factors by latent profile

Profile	n (%)	SSt1	SSt2	SSt3	SSt total
Low risk	3,066 (30.7%)	1.42	1.43	1.57	1.47
Medium risk	4,969 (49.7%)	2.25	1.60	2.43	2.09
High risk	1,965 (19.6%)	3.17	1.83	3.19	2.77

The patterns illustrated in Figure 1 reveal monotonic separation between the three profiles, with progressively greater distances between profiles in SSt1 (interpersonal relations) and SSt3 (belonging/identification), while SSt2 (educational expectations) presented a more compressed distribution across profiles.

Figure 1. Latent profile means across the three SSt factors

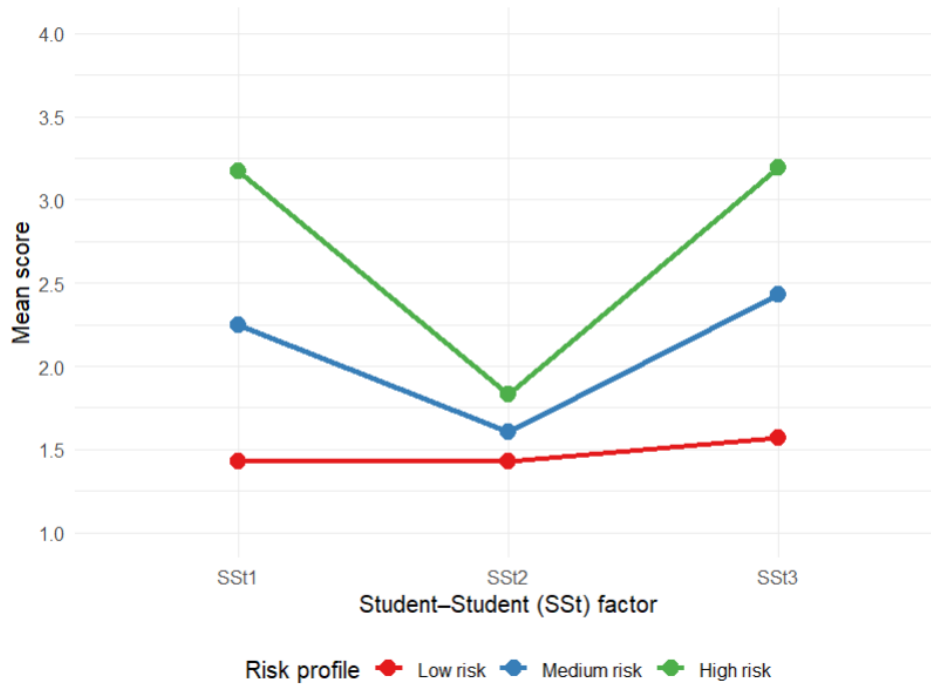


Fig. 1. Line plot of profile means across SSt factors



Figure 2. Distribution of SSt factors by latent risk profile

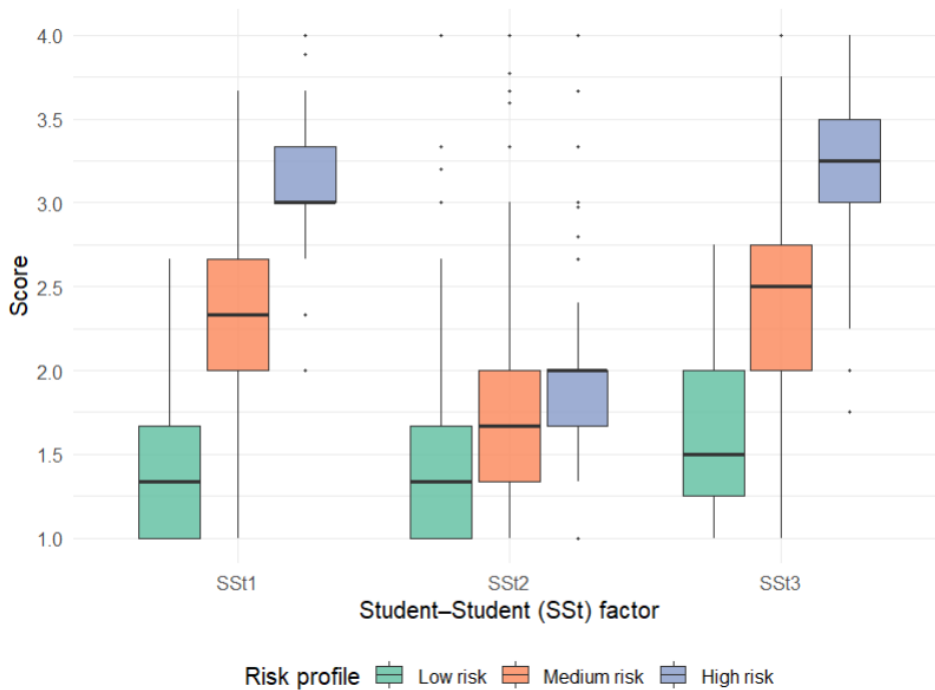


Fig. 2. Boxplot of distributions by profile

3.4 Differences Between Profiles

Kruskal–Wallis tests revealed statistically significant differences between the three latent classes for the SSt total score ($\chi^2(2) = 7,702.60$; $p < .001$; $\epsilon^2 = 0.770$), indicating a very large effect. Differences were also significant and of very large magnitude for SSt1 — interpersonal relations and social skills ($\chi^2(2) = 6,870.60$; $p < .001$; $\epsilon^2 = 0.687$) — and SSt3 — belonging/identification ($\chi^2(2) = 6,537.25$; $p < .001$; $\epsilon^2 = 0.654$). In contrast, SSt2 (educational expectations) presented a significant but smaller effect ($\chi^2(2) = 811.30$; $p < .001$; $\epsilon^2 = 0.081$). Post hoc Dunn comparisons confirmed significant pairwise differences between all three classes (low vs. medium, medium vs. high, low vs. high; all $p < .001$). The latent profiles also differed significantly across all other dimensions of the IAFREE-A. Table 3 summarizes the results.

Table 3. Kruskal–Wallis tests comparing latent profiles across IAFREE-A dimensions and sub-factors

Dimension / Factor	χ^2	df	p	ϵ^2
Student–Student total (SSt)	7,702.60	2	< .001	0.770
SSt1 — Interpersonal relations	6,870.60	2	< .001	0.687
SSt2 — Educational expectations	811.30	2	< .001	0.081
SSt3 — Belonging/identification	6,537.25	2	< .001	0.654
Student–School total (SSc)	3,587.85	2	< .001	0.359
SSc2 — Perception of safe school	3,176.01	2	< .001	0.318
Student–School Professionals (SSP)	1,886.85	2	< .001	0.189
Student–Family total (SF)	1,498.63	2	< .001	0.150
Student–Community total (SC)	1,710.22	2	< .001	0.171

The largest effect outside the SSt dimension occurred in Student–School (SSc) ($\epsilon^2 = 0.359$), particularly in the Perception of the School as a Safe Place factor (SSc2; $\epsilon^2 = 0.318$). Moderate effects were observed for SSP, SC, and SF. All post hoc Dunn comparisons confirmed significant differences between low, medium, and high risk groups across all dimensions (all $p < .001$).

4. Discussion

The present study identified three monotonically ordered profiles of relational risk (low, medium, and high) and demonstrated that these profiles extend their differentiation to all other relational dimensions of the IAFREE-A, with particularly strong effects in the institutional domain. The identification of three latent profiles with adequate classification quality supports the conceptual premise that relational risk among peers is not uniformly distributed in the student population, but organizes into qualitatively distinct configurations, aligning with process-oriented perspectives that conceptualize school disengagement as a relational and multifactorial phenomenon [10, 22].

The pattern of effect sizes across the three SSt factors revealed an important nuance. Interpersonal relations and social skills ($\epsilon^2 = 0.687$) and belonging/identification ($\epsilon^2 = 0.654$) presented very large effects, while educational expectations ($\epsilon^2 = 0.081$) showed a small effect, suggesting that students' aspirations remain relatively stable across risk profiles. This pattern reinforces the interpretation that relational risk emerges primarily in the affective and interpersonal fabric of school experience rather than in the cognitive domain of aspirations, supporting the conceptual framework of the IAFREE-A [26].

Notably, the latent profiles identified through peer relations extended their differentiation to all other IAFREE-A dimensions, with the largest effect outside the SSt dimension occurring in Student–School ($\epsilon^2 = 0.359$), particularly in the perception of school as a safe place ($\epsilon^2 = 0.318$). This convergence indicates that peer-relational risk does not operate in isolation but reverberates into how students experience the school as an institution. Such finding aligns with the broader argument that school trajectories must be understood across multiple interconnected dimensions (systemic, institutional, and individual [11, 24]) and supports the relational model underlying the IAFREE-A [26].

This study has limitations that must be acknowledged. The analyses were conducted on synthetic data generated via the Gaussian Copula technique to comply with the LGPD and to safeguard the identity of participating minors. While this approach preserves the multivariate structure of the original data, it introduces methodological constraints regarding the generalizability of magnitude estimates. Additionally, the cross-sectional design limits causal inferences.

5. Conclusion

The findings indicate that relational risk in school trajectories is organized into three qualitatively distinct profiles, characterized primarily by peer relations and school belonging, while educational expectations remain relatively preserved across risk levels. The extension of these profiles to the institutional domain, particularly to the perception of school as a safe place, supports a process-oriented understanding of school disengagement as a phenomenon that operates simultaneously in interpersonal and institutional dimensions.

Methodologically, the integration of person-centered modeling (LPA) with group-comparison testing on privacy-preserving synthetic data demonstrates the viability of conducting psychometrically rigorous research with minors under LGPD-compliant conditions, offering a pathway for ethically responsible educational research and informing protective interventions focused on the relational fabric of school experience.

Furthermore, large-scale data collection is currently underway for the nationwide application of the IAFREE-A across Brazilian public schools. This ongoing effort will allow future studies to test the replicability of the latent profile structure identified here in empirical, non-synthetic data, expanding the evidence base for the use of relational instruments in early identification of at-risk school trajectories at a national level.

REFERENCES

- [1] Akogul, S., & Erisoglu, M., "An approach for determining the number of clusters in a model-based cluster analysis", *Entropy*, 19(9), Article 452, 2017.
- [2] Ahmadian, M., Bodalal, Z., van der Hulst, H. J., Vens, C., Karssemakers, L. H., Bogveradze, N., ... & Castelijns, J. A., "Overcoming data scarcity in radiomics/radiogenomics using synthetic radiomic features", *Computers in Biology and Medicine*, 174, 108389, 2024.
- [3] American Public Health Association, "The dropout crisis: A public health problem and the role of school-based health care", Washington, DC, American Public Health Association, 2018.
- [4] Barros, R. P., "Políticas públicas para redução do abandono e evasão escolar de jovens", São Paulo, Insper, 2017.

- [5] Barroso, P. C. F., Oliveira, Í. M., Noronha-Sousa, D., Noronha, A., Mateus, C. C., Vázquez-Justo, E., & Costa-Lobo, C., "Fatores de evasão no ensino superior: Uma revisão de literatura", *Psicologia Escolar e Educacional*, 26, e228736, 2022.
- [6] Brown, T. A., "Confirmatory Factor Analysis for Applied Research", 2nd ed., New York, The Guilford Press, 2015.
- [7] Brasil, "Estatísticas do Censo Escolar da Educação Básica", Brasília, Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), 2020.
- [8] Costa, J. C. A., Torres, L. F. F., Loureto, G. D., Freires, L. A., Freitas, A. L. G. B., Teixeira Júnior, N. A., & Cordeiro, T. D., "Relational factors for the risk of school dropout scale – Revised (IAFREE-R): Validity evidence in Brazil", *Iberoamerican Journal of Health and Social Research*, in press.
- [9] Ecker-Lyster, M., & Niileksela, C., "Keeping students on track to graduate: A synthesis of school dropout trends, prevention, and intervention initiatives", *Journal of At-Risk Issues*, 19(2), pp. 24–31, 2016.
- [10] Fuentes, S., "Pesquisas sobre políticas de proteção de trajetórias: Educação secundária na América Latina", Buenos Aires, IPE UNESCO, 2024.
- [11] Kaplan, C., & Fainsod, P., "Pobreza urbana, diversidad cultural y escuela media: Notas sobre las trayectorias escolares de las adolescentes embarazadas", Buenos Aires, Universidad de Buenos Aires, 2001.
- [12] Karacabey, M. F., & Boyaci, A., "Individual and institutional factors contributing to school dropouts", *Kastamonu Education Journal*, 27(3), pp. 1047–1057, 2019.
- [13] Li, Z., Zhao, Y., & Fu, J., "Sync: A copula based framework for generating synthetic data from aggregated sources", in *2020 International Conference on Data Mining Workshops (ICDMW)*, IEEE, pp. 571–578, 2020.
- [14] Marchbanks, M. P., Blake, J. J., Smith, D., Seibert, A. L., & Carmichael, D., "More than a drop in the bucket: The social and economic costs of dropouts and grade retentions associated with exclusionary discipline", *Journal of Applied Research on Children*, 5(2), Article 17, 2014.
- [15] Mosquera, L., El Emam, K., Ding, L., Sharma, V., Zhang, X. H., Kababji, S. E., ... & Eurich, D. T., "A method for generating synthetic longitudinal health data", *BMC Medical Research Methodology*, 23(1), 67, 2023.
- [16] Neri, M. C., "Motivos da evasão escolar", Rio de Janeiro, Fundação Getúlio Vargas, 2009.
- [17] R Core Team, "R: A language and environment for statistical computing" (Version 4.4) [Computer software], Vienna, R Foundation for Statistical Computing, 2024.
- [18] Rosenberg, J. M., Beymer, P. N., Anderson, D. J., Van Lissa, C. J., & Schmidt, J. A., "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, 2018.
- [19] Savran, E., & Karpat, F., "Synthetic data generation using Copula model and driving behavior analysis", *Ain Shams Engineering Journal*, 15(12), 103060, 2024.
- [20] Sichani, E. K., Smith, A., El Emam, K., & Mosquera, L., "Creating high-quality synthetic health data: Framework for model development and validation", *JMIR Formative Research*, 8(1), e53241, 2024.
- [21] Silva Filho, R. B., & Araújo, R. M. L., "Evasão e abandono escolar na educação básica no Brasil: Fatores, causas e possíveis consequências", *Educação Por Escrito*, 8(1), pp. 35–48, 2017.
- [22] Soares, J. F., Alves, M. T. G., & Fonseca, J. A., "Trajetórias educacionais como evidência da qualidade da educação básica brasileira", *Revista Brasileira de Estudos de População*, 38, pp. 1–21, 2021.
- [23] Soares, T. M., Fernandes, N. da S., Nóbrega, M. C., & Nicolella, A. C., "Fatores associados ao abandono escolar no ensino médio público de Minas Gerais", *Educação e Pesquisa*, 41(3), pp. 757–772, 2015.
- [24] Terigi, F., & Briscioli, B., "Investigaciones producidas sobre 'trayectorias escolares' en educación secundaria (Argentina, 2003-2016)", in D. Pinkasz & N. Montes (Eds.), *Estados del arte sobre educación secundaria*, Buenos Aires, Ediciones UNGS/FLACSO, pp. 119–172, 2020.
- [25] UNICEF, "Early warning systems for students at risk of dropping out", *UNICEF Series on Education Participation and Dropout Prevention*, vol. 2, 2018.
- [26] Vasconcelos, A. N., Freires, L. A., Loureto, G. D. L., Fortes, G., Costa, J. C. A., Torres, L. F. F., Bittencourt, I. I., Cordeiro, T. D., & Isotani, S., "Advancing school dropout early warning systems: The IAFREE relational model for identifying at-risk students", *Frontiers in Psychology*, 14, Article 1189283, 2023.
- [27] Witte, K., Cabus, S. J., Thyssen, G., & Groot, W., "A critical review of the literature on school dropout", *Educational Research Review*, 10, pp. 13–28, 2013.



International Conference The Future of Education



- [28] Wood, L., Kiperman, S., Esch, R. C., Leroux, A. J., & Truscott, S. D., "Predicting dropout using student- and school-level factors: An ecological perspective", *School Psychology Quarterly*, 32(1), pp. 35–49, 2017.