An analytics approach to investigate teacher turnover

Bruno Elias Penteado, Seiji Isotani

Instituto de Ciências Matemática e Computação – Universidade de São Paulo (USP)
13566-590 – São Carlos – SP – Brasil

brunopenteado@alumni.usp.br, sisotani@icmc.usp.br

***Abstract**. Some educational problems embed spatial and temporal complexities, and the aggregation of these data may cause informational loss. One example regards teacher turnover, which impacts the students' learning processes. In this work, we adopted a cross-sectional study, using visual analytics techniques to identify complex patterns in the mobility data of teachers in public schools. We used education census data from the Brazilian government, which maps which teachers teach in which schools. In addition, we sought to understand which are the main factors that influence this sort of decision. We used synthetic indicators developed by INEP to identify different motivation clusters that may influence teachers' decision. As result, we identified different geographical patterns varying according to their contract type. The clusters identified as the main factors: school performance, school climate, and management complexity.*

***Resumo**. Certos problemas educacionais embutem complexidades espaciais e temporais e a agregação desses dados pode causar perda informacional. Um exemplo é a rotatividade docente, que impacta a aprendizagem dos alunos. Neste trabalho, adotamos uma metodologia transversal, usando técnicas de análise visual para identificar padrões complexos nos dados das movimentações dos docentes entre escolas públicas. Utilizamos dados do censo educacional do governo brasileiro, que mapeiam onde os professores estão lotados anualmente. Além disso, buscamos compreender quais os principais fatores que influenciam essa decisão. Usamos indicadores sintéticos desenvolvidos pelo INEP para identificar diferentes clusters de motivação que podem influenciar a decisão dos professores. Como resultado, identificamos diferentes padrões geográficos variando de acordo com seu tipo de contrato. Os clusters também identificaram como principais fatores: desempenho e clima escolar e complexidade da gestão.*

1. Introduction

The continuity of a teacher's pedagogical work is one of the most impacting factors in high stakes exams performance [Sorensen & Ladd, 2020; Hanushek et al., 2016; Ronfeldt, 2013; Moriconi, 2012]. It is desirable that the same teacher develops his work in the same educational units throughout time to hold the consistency of his pedagogical planning and the collaborative work among teachers [Allensworth et al. 2009]. Thus, teachers' turnover is a problem that impacts students' learning, particularly for schools that serve low-

DOI: 10.5753/sbie.2021.218738

income regions and students, which may increase the inequalities of opportunities among schools [Clotfelter et al., 2020].

As pointed by Allensworth et al. [2009]: "While some teacher mobility is normal and expected, high turnover rates can produce a range of organizational problems at schools, such as discontinuity in professional development, shortages in key subjects, and loss of teacher leadership. Previous research also indicates that schools with high turnover are more likely to have inexperienced, ineffective teachers". The lack of studies on teacher mobility hampers the proper development of effective public policies [Carvalho, 2019]. Only a few works presented a quantitative analysis on this matter; however, they show the severity of the problem. The survey made by Carvalho [2019] shows that, on average, between 2008 and 2016, 39.8% of teachers in the Brazilian educational system changed their schools per year, with a sharp distortion between regions and with a significant difference between the public and private systems.

The adequate comprehension of this phenomenon is essential for its correct diagnostics and possible interventions [Barry & Hirsh, 2005]. A limitation of these studies concerns the presentation of the information collected about this phenomenon. It is usually approached in aggregated form and descriptive tables, which may not capture subtleties of a problem with such complexity or even in an aggregated view - by school units and not individuals. In this work, we propose a new approach, based on visual analytics principles, to provide support for the comprehension of teachers' turnover phenomenon, in a finer-grained scale and enable the visualization of complex data by public administrators. Two objectives are set: i) how do these mobility patterns look like? and ii) which factors contribute to this mobility? As such, we intend to answer the reason why teachers may leave a school and what do they look for in a new place of work.

2. Background and related works

Visual analytics can be defined as 'the science of analytical reasoning facilitated by interactive visual interfaces' [Thomas & Cook, 2005] and it is used to handle large scale and complex data, which may lose information when transformed to aggregated descriptive representations. It also considers the human in the loop, formulating and testing hypotheses, codifying, and identifying interesting features in the data. In this work, we highlight 3 characteristics that are not properly handled with descriptive tables and aggregated information: i) temporal information, although present in tables, they do not convey the passage of time; ii) spatial information, by geolocating schools and contextualizing them according to their neighborhoods; iii) graph information, setting the relations of mobility across two different school units. These characteristics make the interpretation of these data more complex and with higher informational overload [Fisher, 2005]. The use of learning analytics at the macro-level – as in national data – is relatively little explored by the literature [Fischer et al. 2020] and visual analytics is still underexplored in learning analytics in general [Vieira et al., 2018]. In Vieira et al. [2018] the authors bring a systematic review with 52 papers that use visual analytics in educational data. Different contexts, data sources, techniques, and applications were identified but none of them deal with this sort of context. In this work, we explore visual analytics techniques for complex data at the macro-level.

The Brazilian Institute of Educational Research (INEP) has elaborated an indicator of this construct in order for Brazilian schools to have a measure on this matter – the Teacher Regularity Index (IRD, in Portuguese) [INEP, 2015], which monitors the permanence of teachers in their respective schools in the last five years. Although a valuable indicator of the degree of teacher turnover in schools, it provides a perspective on schools and not on teachers. Quantitative works such as Carvalho [2019] compile information in table form with descriptive information and in a higher aggregation level, considering the entire primary public education. In his work, the author found a bigger turnover in public schools, with teachers leaving more than entering the public system, as well as considerable regional differences. The author also points that schools with lower academic results in IDEB and the principal turnover as the main characteristics for teacher to consider moving to other schools. For this reason, the amount of information displayed is usually low, with a few dozens of rows, aggregated in high level administrative regions. Indicators such as the IRD [INEP, 2015] compile this information up to the school level, enabling to track the movements over time. However, these approaches fail to represent individual mobility, unable to map the origins and destinations and understand mobility patterns between school units and regions within a city.

No papers were found concerning the problem of teacher turnover and its informational complexity. By complexity, we mean the union of different dimensions of the problem: temporal (between years), geolocated (schools' locations), and graph (mobility trajectories). Another gap found was the application in macro-level educational data [Penteado et al., 2019] – open data from organizations such as governments, essentially longitudinal and less frequent than the other levels.

3. Methodology

This work uses a retrospective observational study design, with data collected by the Brazilian national K-12 school census in 2016 and 2017, having as the unit of analysis a municipality and its regions. We considered only a single unit of analysis - the city of São Paulo - given its complexity, its geographical organization, and the number of teachers in the public system (approx. 40k teachers in 1.5k schools). The following filters were applied: mobility only across public school units, only in the city of São Paulo, and with schools that were not created or shut down in the years considered.

From the problem presented, we derived three visualization requirements: i) reach the level of individual teachers; ii) trace the mobility trajectory; iii) geolocate schools in the map of the municipality. By meeting these requirements, we argue that we can augment the perception of the problem at hand. The flow map technique [Dent, 1999] meets these requirements and it is widely used in applications that show the trajectory of objects from one location to another.

To understand the contextual factors that may motivate these mobility patterns, we elicited from the literature some indicators that influence these trajectories, considering school characteristics. Table 1 details the variables considered in this study, according to their relevance evidenced in the educational literature.

Table 1: Variables selected for the analysis.

Indicator	Description	Ref
IDEB [AI & AF]	School performance in the national large-scale exam, for Math and Portuguese and mediated by approval rate. Two measures are taken: the score for elementary (AI) and for middle levels (AF)	Alves & Soares, 2013
NSE	Students' socioeconomic level	Alves et al., 2014
Management complexity (ICG)	Measure aspects related to school complexity: number of students, their ages, educational levels and periods (morning, afternoon and night)	Gobbi et al. 2020
School climate (CLIMA)	Teacher engagement in pedagogical tasks along with their colleagues	Oliveira & Waldhlem, 2016
School violence (VIOLENCIA)	Occurrence of different forms of violence in the school (theft, threats, verbal or physical aggression)	Teixeira & Kassouf, 2015
Principal's leadership (DIRETOR)	Principal's respect, motivation towards school and teachers' matters	Oliveira & Waldhlem, 2016
Infrastructure (INFRA)	Existence of basic infrastructure (classroom facilities, court, labs, library, internet, etc.)	Neto et al., 2013

The variables from Table 1 are extracted from the Brazilian government open data [Penteado & Isotani, 2017] and are represented by Item Response Theory (IRT) scores. For each of them, it was calculated the difference of the scores: score destination – score origin, to capture the gain of that aspect when moving to a new school. Schools that did not present any of the indicators for the years considered were also discarded, resulting in 2209 individual trajectories. Thus, each trajectory is characterized by an 8-dimensional vector, with the gain in each of the variables.

To find similar clusters in these dimensions, we applied the unsupervised self-organizing maps (SOM) algorithm [Kohonen, 2001] on this dataset, suitable for multidimensional visualization, which maps a multidimensional input vector to a lower-dimensional (usually two dimensional) output vector, while preserving topological properties, i.e., close vectors form the input space are also close to one another in the output space. The elbow method was used to determine the optimal number of clusters, resulting in 3 main clusters to maximize the variability.

Figure 1 illustrates the procedures adopted in this work. From the original problem of teacher turnover, we derived the two research questions, which adopt different approaches to be answered. First, we sought to understand spatial patterns in the displacement of teachers, and we specified some visual analytics requirements to be met and the application of the flow map technique, using the school census dataset from 2016 and 2017, which contained individual data from the teachers and their places of work. Second, we compiled contextual factors from the literature related to this problem, collected school-level open data indicators from the Brazilian government, and performed data clustering so that we could understand what motivates this change.

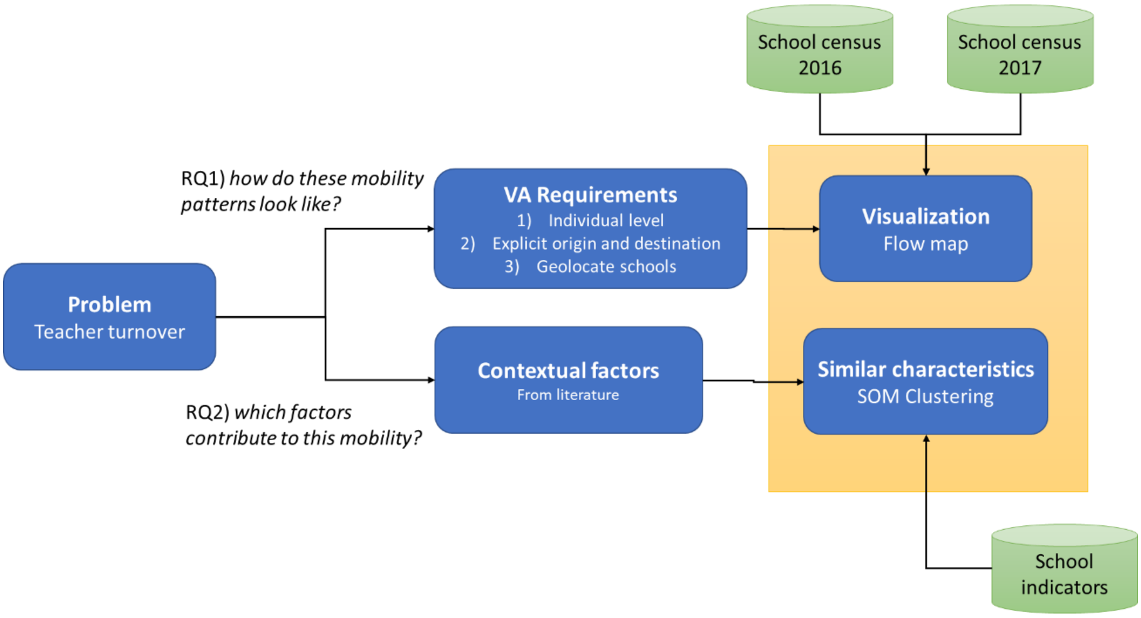

Figure 1. General procedure adopted in this work

4. Results

The resulting flow map is illustrated in Figure 2. Each trajectory represents an individual teacher (requirement #1) is mapped with an edge connecting origin schools to destination schools (requirement #2), located on a map (requirement #3). Three different results are displayed, according to the professional teachers' contract (first the whole set; the permanent teachers (66,6%) and the temporary teachers (30.2%), respectively; other types of employment contracts were not shown here because they are very few). One can see different patterns, having the temporary teachers moving towards schools closer to their original units while the permanent ones cross larger distances across the city. Most of the displacements occur in peripherical regions, where lower-income students live and there are more geographical vulnerabilities (violent regions, crowded public transportation, poor infrastructure, etc.).

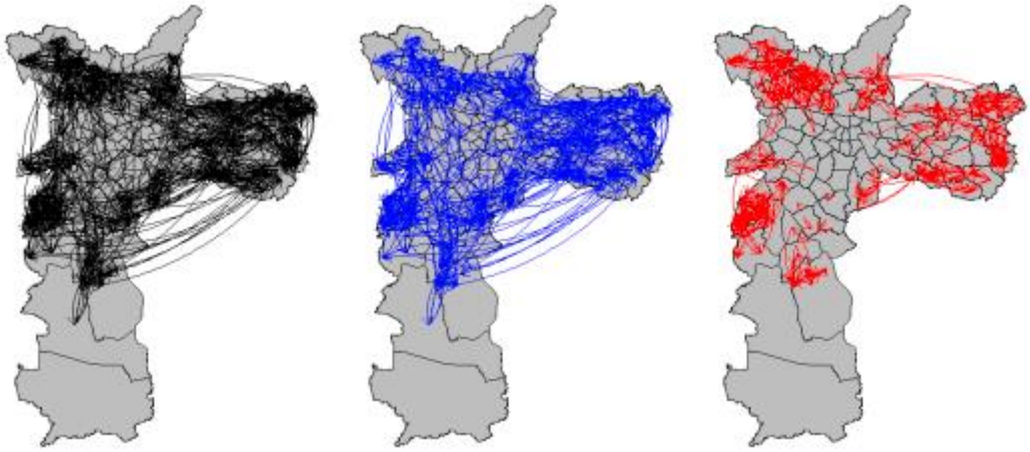

Figure 2: Teachers' mobility in public schools of the São Paulo municipality (2016-2017). From left to right: overall mobility; permanent teachers and temporary teachers.

To determine the dimension of the SOM it was adopted the same criterion that in Tian et al. (2014), which calculates the number of vertices as a function of the number of records in the dataset. Figure 3 shows how each of the factors has its values distributed for each node. Each trajectory was classified in one of the nodes of the grid, along similar vectors (trajectories) and it is in the same position across the different maps. The redder, the higher the positive variation (gain in scores), while the bluer, the higher the negative magnitude.

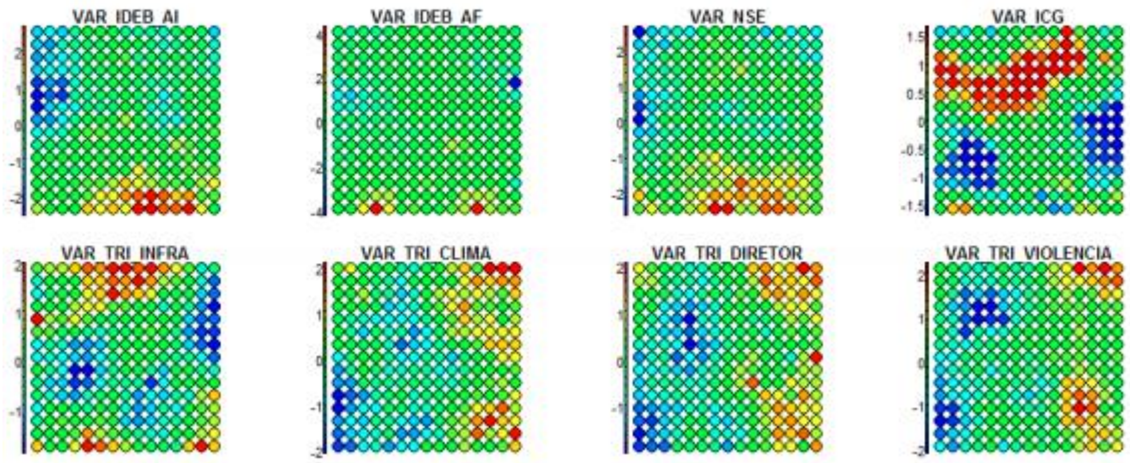

Figure 3: Self-organizing maps demonstrating the variation in each node of the grid for the 8 variables. In each node is shown the average variation for the schools contained in them.

Figure 4 shows the differences among the averages of each cluster detected, for each of the 8 dimensions. Cluster 1 points to displacements whose teachers sought schools with a higher score for school climate, better relationship with the school principal, and lower violence (in this scale, the higher the score, the lower the violence). Cluster 2 shows displacements of teachers who sought schools with better performance scores in the large-scale exam (IDEB), both for elementary and middle levels, as well as a higher average socioeconomic level of their students. Lastly, Cluster 3 shows displacements whose teachers went to schools with higher management complexity (ICG), i.e., bigger schools, with more enrolments, levels, periods, and older students. Table 2 shows the average value for each of the clusters.

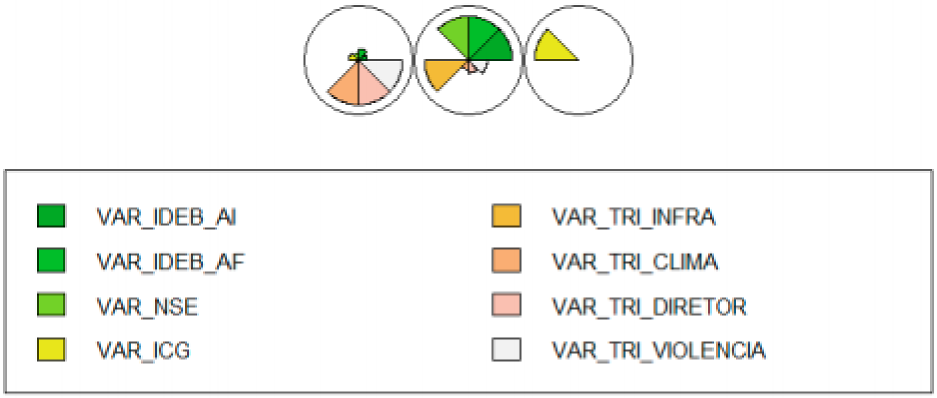

Figure 4. Representation of the average values for each variable in each cluster.

Table 2: Average values for each dimension in the clusters.

Indicator	Cluster 1 (n=857)	Cluster 2 (n=444)	Cluster 3 (n=908)
IDEB AI	-0.03	0.69	-0.24
IDEB AF	0.04	0.92	-0.29
NSE	-0.06	2.85	-0.82
ICG	-0.29	-0.27	0.42
INFRA	-0.10	0.17	-0.07
CLIM	0.54	-0.29	-0.37
PL	0.67	-0.30	-0.52
VIOL	0.48	-0.07	-0.40

5. Discussion

In this work, we presented a relevant educational problem that has a big impact on school performance (a proxy for the average students' performances). The problem was characterized, and different indicators were selected to describe it. The visual analytics techniques met the requirements elicited, providing a rich overall perspective on this complex, multidimensional phenomenon, hidden in traditional forms of visualization. Through the figures, it is possible to identify different patterns of displacement in function of their work contract. Besides, one can also note the absence of displacements downtown – even if it has fewer public schools. This result seems to suggest that the number of displacements is positively correlated to social vulnerabilities indicators, what – to the best of our knowledge – is not reported in the literature. This needs to be confirmed in further iterations of the research.

Although this work does not present a traditional approach of digital learning environments or other forms of student-teacher or student-student interaction [Vieira et al. 2018], we argue that this is an opportunity for the learning analytics field, since this problem may impact more on large-scale exams – the most reliable form for comparing large scale cohorts - than small, controlled interventions.

All the aspects elicited from the literature presented differentiation among the clusters, indicating their importance in relation to the phenomenon. Although they are Brazilian instruments, measured for all school units in the country, they can be extended to other contexts. As it is a practice for INEP, each of the indicators is created based on multiple measurements and they are combined into a single IRT scale for each construct, which permits to rank different results in an interpretable manner. The application of the SOM algorithm identified three factors as the most influential: school performance, school climate, and management complexity. School performance and school climate are in consonance with the results of Carvalho [2019], which considered a decade of data with all the Brazilian municipalities and administrative dependencies. By adopting a single city, we controlled for differences in municipalities' academic outcomes, that may also vary according to socioeconomic indicators [Penteado, 2016]. Although there are obviously other criteria that determine the preference for displacements, we believe that these variables have a strong weight on the teacher's choice when considering a new place to work.

Berry and Hirsch [2005] argue that public administrators when developing educational policies and allocate resources should prioritize schools that present the most difficulty retaining its teachers. And the method presented here may help them to discover what are these schools and other schools they lose their staff and what are their motivations.

These results may also be related to the United Nation Organization Sustainable Development Goals (https://sdgs.un.org/goals), in two different aspects: Goal #4-Quality Education, by helping public administration to identify relevant factors to keep teachers at their workplaces and to provide better learning results, with safer, more inclusive and more efficient environments; and Goal #11-Sustainable cities and communities, by identifying regions which need additional public transportation or to plan social interventions in the neighborhoods of school which present atypical displacements.

As future work, we intend to design a case study with different public administrators, in order to evaluate how many insights are provided with these analytics and how novel they are from a pragmatic point of view. Other data should be included, such as the individual differences (such as age, background, professional experience, etc.) and, also, other more robust analysis techniques. We only considered two consecutive years to analyze; thus, additional years should also be included, what may make the visualization more challenging.

References

Allensworth, E.; Ponisciak, S. M.; Mazzeo, C. (2009). The schools teachers leave: teacher mobility in Chicago public schools. Research Report. Consortium on Chicago School Research at the University of Chicago Urban Education Institute.

Alves, M. T. G.; Soares, J. F.; Xavier, F. P. (2014). Índice socioeconômico das escolas de educação básica brasileiras. Ensaio: Avaliação e Políticas Públicas em Educação, 22, p. 671– 703.

Alves, M. T. G; Soares, J. F. (2013). Contexto escolar e indicadores educacionais: condições desiguais para a efetivação de uma política de avaliação educacional. Educação e Pesquisa, 39(1), p. 177–194.

Berry, B.; Hirsch, E. (2005). Recruiting and retaining teachers for hard-to-staff schools, from: https://files.eric.ed.gov/fulltext/ED489221.pdf.

Carvalho, J. C. (2019). Rotatividade dos professores na educação básica brasileira: caracterização, fatores associados e impactos sobre os resultados dos estudantes (2007-2016). PhD Dissertation, Federal University of Paraná, Brazil, 203 p.

Clotfelter C. T.; Ladd, H. F.; Vigdor, J. L (2010). Teacher mobility, school segregation and pay-based policies to level the playing field. Washington, DC: National Center for Analysis of Longitudinal Data in Education Research, 2010. (Working Paper, 44). Available at: https://caldercenter.org/sites/default/files/CALDERWorkPaper_44.pdf.

Dent, B. D. (1999). Cartography: Thematic Map Design. New York: McGraw-Hill. ISBN 0-697-38495-0.

Fisher, B. D. (2005). Visual Representations and Interaction Technologies. In: J Thomas and K Cook (Eds), Illuminating the Path: A Research and Development Agenda for Visual Analytics, p. 69-104, IEEE Press.

Fischer, C.; Pardos; Z. A; Baker, R. S.; Williams, J. J.; Smyth, P; Yu, R.; Slater, S.; Baker, R.; Warschauer, M. (2020). Mining Big Data in Education: Affordances and Challenges. Review of Research in Education, 44(1), p. 130–160.

Gobbi, B. C.; Lacruz, A. J.; Américo, B. L.; Filho, H. Z. (2020). Uma boa gestão melhora o desempenho da escola, mas o que sabemos acerca do efeito da complexidade da gestão nessa relação? Ensaio: Avaliação e Políticas Públicas em Educação, 28 (106), p.198-220. https://doi.org/10.1590/s0104-40362019002701786.

Hanushek, E. A.; Rivkin, S. G.; Schiman, J. C. (2016). Dynamic Effects of Teacher Turnover on The Quality Of Instruction. National Bureau of Economic Research, Working paper 22472. Disponível em: https://www.nber.org/papers/w22472.

INEP. (2015). Indicador de regularidade do docente da Educação Básica, from: http://download.inep.gov.br/informacoes_estatisticas/indicadores_educacionais/2014/docente_regularidade_vinculo/nota_tecnica_indicador_regularidade_2015.pdf.

Kohonen, T. (2001). Self-organizing maps. Springer Science & Business Media, 2001 - 501 p.

Moriconi, G. M. (2012). Medindo a eficácia dos professores: o uso de modelos de valor agregado para estimar o efeito do professor sobre o desempenho dos alunos. PhD Dissertation Fundação Getúlio Vargas, Brazil, 2012. http://hdl.handle.net/10438/9620.

Neto, J. J. S.; Jesus, G. R.; Karino, C. A.; Andrade, D. F. (2013). Uma escala para medir a infraestrutura escolar. Estudos em Avaliação Educacional, São Paulo, v. 24, n. 54, p. 78-99.

Oliveira, A. C. P.; Waldhlem, A. P. S. (2016). Liderança do diretor, clima escolar e desempenho dos alunos: qual a relação? Ensaio: avaliações de políticas públicas em Educação (24) 93, pp.824-844. https://doi.org/10.1590/S0104-40362016000400003.

Penteado, B. E. (2016). Correlational Analysis Between School Performance and

Municipal Indicators in Brazil Supported by Linked Open Data. Proceedings of the 25th International Conference Companion on World Wide Web, p. 507-512. https://doi.org/10.1145/2872518.2890459.

Penteado, B. E.; Isotani, S. (2017). Dados abertos educacionais: que informações temos disponíveis? Anais do VI Congresso Brasileiro de Educação, vol. 4. Bauru, Julho de 2017, ISBN 978-85-5444-002-2, p. 1933-1938. Disponível em: http://www.cbeunesp.com.br/2017/pages/anais_cbe_v01.pdf.

Penteado, B. E.; Bittencourt, I. I.; Isotani, S. (2019). Modelo de referência para dados abertos educacionais em nível macro. Anais do XXX Simpósio Brasileiro de Informática na Educação. p. 1808, nov. 2019. ISSN 2316-6533. Disponível em: http://www.br-ie.org/pub/index.php/sbie/article/view/891.

Teixeira, E. C.; Kassouf, A. L. (2015). Impacto da violência nas escolas paulistas sobre o desempenho acadêmico dos alunos. Economia Aplicada, vol 19 (2). https://doi.org/10.1590/1413-8050/ea124436.

Thomas J. J.; Cook, K. A. (2005). Illuminating the path: the R&D agenda for visual analytics. Department of Homeland Security, USA, from: https://www.hsdl.org/?view&did=485291.

Ronfeldt, M., Loeb, S., & Wyckoff, J. (2013). How Teacher Turnover Harms Student Achievement. American Educational Research Journal, 50(1), 4–36. https://doi.org/10.3102/0002831212463813.

Sorensen, L. C.; Ladd, H. F. (2020). The Hidden Costs of Teacher Turnover. AERA Open. https://doi.org/10.1177/2332858420905812.

Tian, J.; Azarian, M. H; Pecht, M. (2014). Anomaly Detection Using Self-Organizing Maps-Based K-Nearest Neighbor Algorithm. PHM Society European Conference, 2(1). https://doi.org/10.36001/phme.2014.v2i1.1554.

Vieira, C.; Parsons, P; Byrd, V. (2018). Visual learning analytics of educational data: a systematic literature review and research agenda. Computers & Education, 122, p. 119-135. https://doi.org/10.1016/j.compedu.2018.03.018.