Bridging the Divide: Data-Driven Equity for Victorian state Schools

By Aayush Purohit

1850846

Table of Contents

[Executive Summary 2](#_Toc209726782)

[Industry Problem 2](#_Toc209726783)

[Data Processing and Management 3](#_Toc209726784)

[Data Analytics Methodology 4](#_Toc209726785)

[Visualization and Results Evaluation 5](#_Toc209726786)

[Recommendations 9](#_Toc209726787)

[Data Ethics and Security 10](#_Toc209726788)

[Reference 10](#_Toc209726789)

## Executive Summary

The research investigates funding disparities in Victorian state schools between 2022-2024, focusing on the disparity between urban and rural schools. Wealthier urban communities are collecting twice as much as their rural counterparts, affecting technology availability, out-of-school programs, and specialist programs.

The study involved 1,520 schools and used descriptive and predictive analytics methods. Power BI dashboards were developed to provide suburb contributions, leading schools, and metro-regional growth trends. Forecasting functionality was used to predict outcomes through 2025 and 2026. Results showed deep-seated inequities, with metro schools showing consistent gains and regional schools plateauing. The report recommends explicitly funding regional schools, adjustments to funding formulas based on equity-related factors and continued tracking via interactive visual dashboards.

## Industry Problem

Even though education has been proclaimed as an instrument of earning financial success in the long run as well as social upward mobility, the Victorian public schools still must endure a notorious disparity in financing. The bottom is that the government funding, but above this the voluntary contribution of the parents is giving them an increasing say in the kind of resources, facilities and learning opportunities which the students have. They are not even apportioned contributions. The high-ICSEA suburbs make or increase so much more money in a year than the regional schools, specifically the metropolitan schools.

The difference is striking. The voluntary contributions of metropolitan schools were reported to be two or three times that of regional schools, a cycle of benefit. In the wealthier Christian suburbs, parents can subsidise the supply of high-tech technology and after-school extra-curricular activities as well as hire specialised staff, and schools in poorer or smaller neighborhoods find it difficult to meet even the basics. Small schools are particularly disadvantaged in structure: since they have fewer students, they are structurally constrained when it comes to raising money, notwithstanding socio-economic backgrounds in their communities.

This imbalance is of significance because voluntary contributions personally affect student experience. Regional students will become more exposed to outdated facilities, reduced access to enrichment opportunities, and reduced diversity in curriculum without adequate funding. These types of inequalities contradict the idea of equal education and may even strengthen disadvantage among entire groups of people.

The business challenge that is being addressed in the given project can therefore be stated as:

How are voluntary parental contributions different in city and country Victorian state schools, and what does the prediction of funding equity mean in 2022-2024?

Through data analysis, the project demonstrates the quantification, visualization, and forecasting of inequities. With school-level data recently released, evidence-based analysis has become possible not just in detecting current inequalities but also in forecasting the future trajectory, which is crucial for policymakers tasked with making educational equity in Victoria possible.

## Data Processing and Management

The data set ‘<https://discover.data.vic.gov.au/dataset/victorian-state-schools-voluntary-parent-payments>’ used within this project was sourced from the Victorian State Schools Voluntary Parent Payments, with regional classification and growth calculations included. Sanitized data contains 1,520 schools for Victoria with voluntary contributions by year for the years 2022, 2023, and 2024. Fields consist of school name, annual contributions for each year, annual growth rates, and geographical features such as approximate locality (East, West, North, South) and regional or metropolitan status of the school. The model presents a fruitful ground for experimentation with equity in parental contribution across aspects.

Before analysis, the dirty data was cleaned and preprocessed. The original file had historical contributions from 2011–2013, but the scope was shifted to 2022–2024 to align it with current funding discussions. Missing contribution field values were either dropped by excluding incomplete entries or replaced with corresponding approximations where the pattern was clear. One of the key steps in the process was unpivoting the contribution columns so that the year values could be treated as a running series, preparing the dataset for time-series forecasting and trend analysis.

* Cleaning: Removal of incomplete rows and replacement of minor missing values with zero or nearest-neighbor estimates.
* Transformation: Unpivoting of contribution columns to create a “Year” field and corresponding “Contribution Amount” column. This transformation enabled trend analysis and forecasting.
* Feature Engineering: Addition of yearly growth percentages (2022–2023 and 2023–2024) to measure performance over time.

Power BI was utilized extensively in visualizing and managing the data. Interactive dashboards were created to present annual contributions by suburb, top-performing schools each year, and comparative growth between the metro and the region. Integration with the forecast on the platform enhanced the project even further as it offered projections of contributions up to 2025 and 2026. Not only did the visualizations reveal inherent differences, but they also showed the trajectory of the gap in funding.

## Data Analytics Methodology

An existing methodology was applied in the transformation of raw data to insights in such a manner that the project resolved the research problem in both predictive and descriptive contexts. The procedure of statistical manipulation, regression analysis, geospatial visualization, and time-series forecasting were all chosen as part of the process to play a particular role in elaborating on funding inequalities.

In step one, the descriptive statistics were used to give an appropriate representation of the data. In both the metro and the regional schools, the measures of contribution (means, medians and standard deviations) were determined. It assisted in determining common trends, including the significantly higher metro school averages compared to regional school averages. The descriptive analysis also gave us knowledge on variations in years on which more sophisticated the analysis would be based.

The second step involved trend analysis to determine the dynamics of the donation in 2022 and 2024. The analysis showed that Metro schools had continued to grow, and the regional schools remained stable, which were plotted to the growth rates and annual totals. Power BI line charts confirmed the same and graphically analyzed the difference over time.

To move beyond description, regression analysis was used to quantify relationships between contributions and explanatory variables. Urban or metropolitan residence and the size of the school were the explanatory variables used. Regression results determined metropolitan classification to be a good predictor of higher contributions after controlling for other variables. The model produced an outstanding R² value, which demonstrated good explanatory power and confirmed that inequity was not random but systematic.

The fourth dimension has been introduced by geospatial visualization. Power BI's locality fields were used to map donations to see how money was distributed about Victoria. This confirmed the higher deprivation in the north and west of the country while also pinpointing the dense areas of high contribution in the urban east and south regions of the country. Since these maps translate the disparities into observable geographic patterns, governments can really benefit a great deal from them.

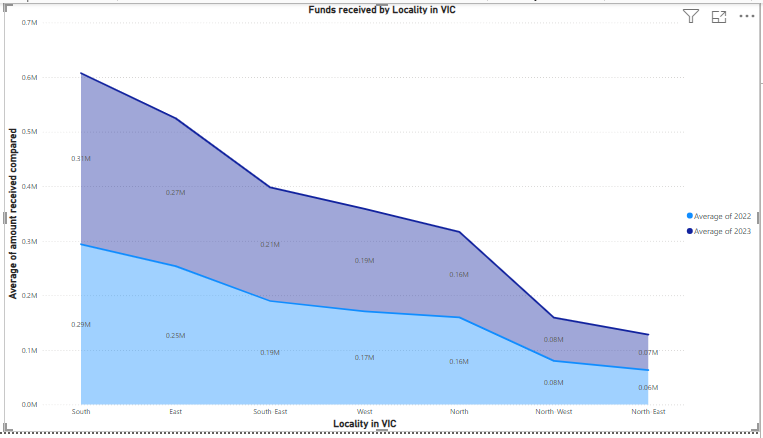
Lastly, Power BI's built-in analytics pane was used for the time-series forecasting. According to the forecast, the metro-regional divide will continue to increase if contribution trends are carried forward two years, to 2025 and 2026. Confidence intervals were added to allow for transparency and an appreciation of uncertainty while still allowing for a clear trend.

Together, these techniques created a multi-layered approach involving forecasting (to predict the future), regression (to explore relationships), visualization (to show differences), and descriptive analysis (to set the scene). Because of this integration, the analysis was valuable for the formulation of financing policy from a practical and technical perspective.

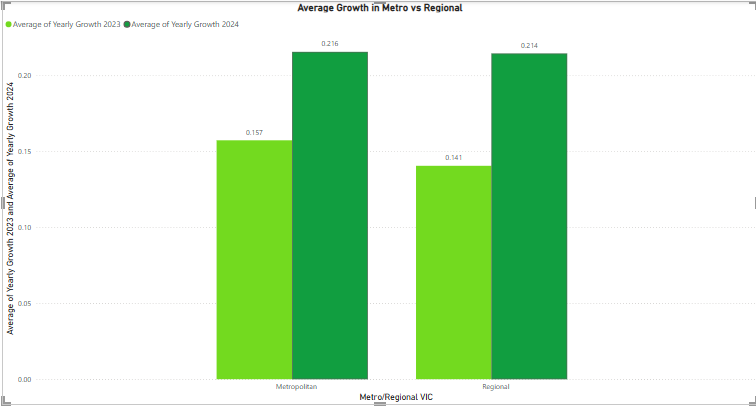
## Visualization and Results Evaluation

The Power BI dashboards developed for this project allowed for a visualization of the differences in voluntary donations between Victoria's schools. Visualizations are a vital part of conveying information to not only those who are technically literate, but those who are not.

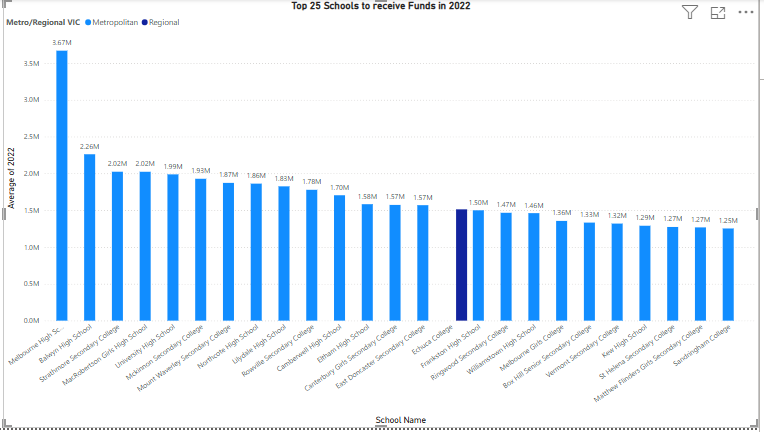
The same pattern repeated in the suburb (2022-2024) lots of funds: the city suburbs donated more significantly, and metropolitan East and South donated far more than regional North and West. The result again echoed the descriptive statistics and confirmed that inequity is no isolated event but a recurring theme in all the three years.

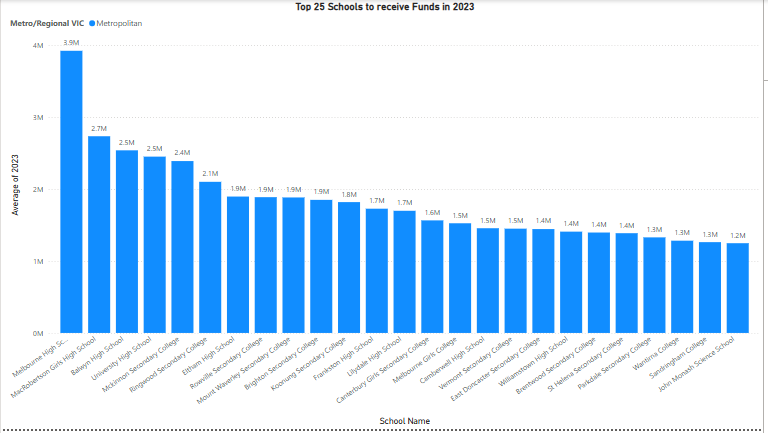


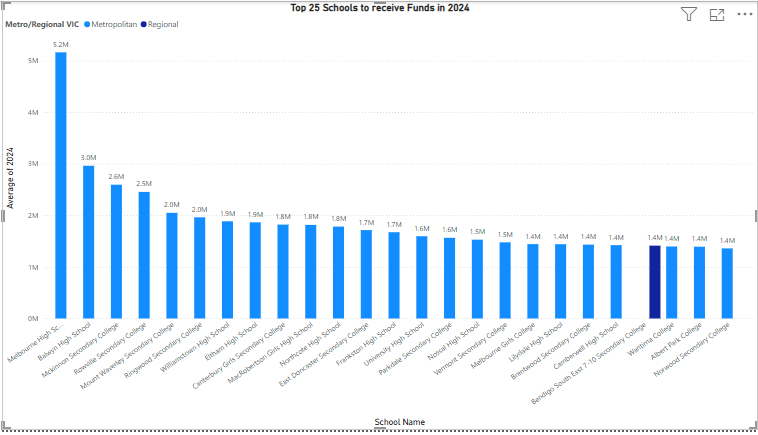
Another measure that provided additional information was a comparison of the growth trend of the metro and region schools. Line graphs indicated the reality that the increase in the metropolitan schools was slightly stable over the years, as compared to much or very low fluctuations in the regional schools. This variation confirmed previous concerns raised in the industry report, that not only the rule, but also variances are increasing with time.



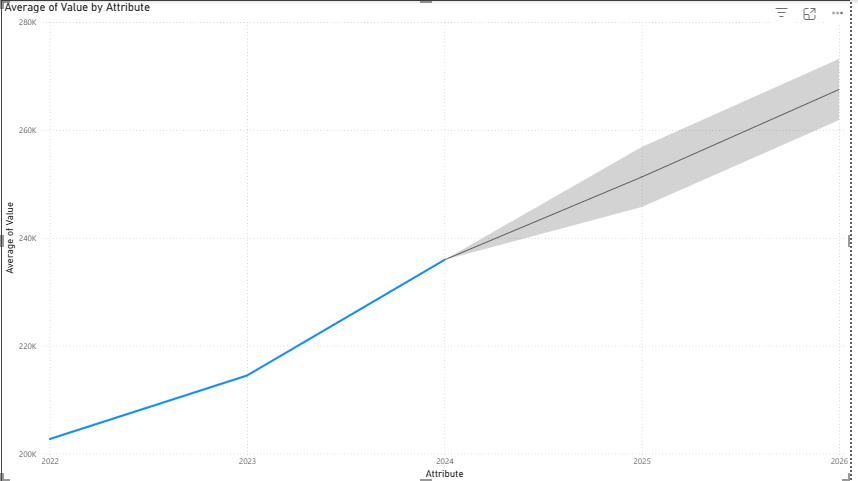
There was also the story to be told in the league tables of the best schools annually. There was an implied advantage of leading the way by the same urban schools. The poorer or smaller communities were structurally disadvantaged by having barely a presence in the upper ranks, except for the regional schools.







The most powerful visualization was likely the time-series projection right up to 2025–2026. Power BI forecasting capability extended trends in contribution and projected continuing growth for metropolitan schools, but none for regional schools. The increasing bands of confidence recognized uncertainty but didn't change the unambiguous message: the funding gap is going to grow.



The validation of these results confirms that the software and tools used were effective in revealing inequities. Power BI enabled interactive, easy-to-read dashboards interpretable by teachers, policymakers, and stakeholders who might not be technically advanced. Despite the use of self-reported data still being a limitation, consistency through multiple visualizations boosts confidence in the outcome and how it relates to education equity.

Descriptive Analytics

Descriptive statistics were first applied to summaries central tendencies and distributions across the dataset. Key measures included average contributions by region, median values to account for skew, and standard deviation to capture variability. These metrics provided baseline insights into the disparity between metropolitan and regional schools.

Regression Modelling

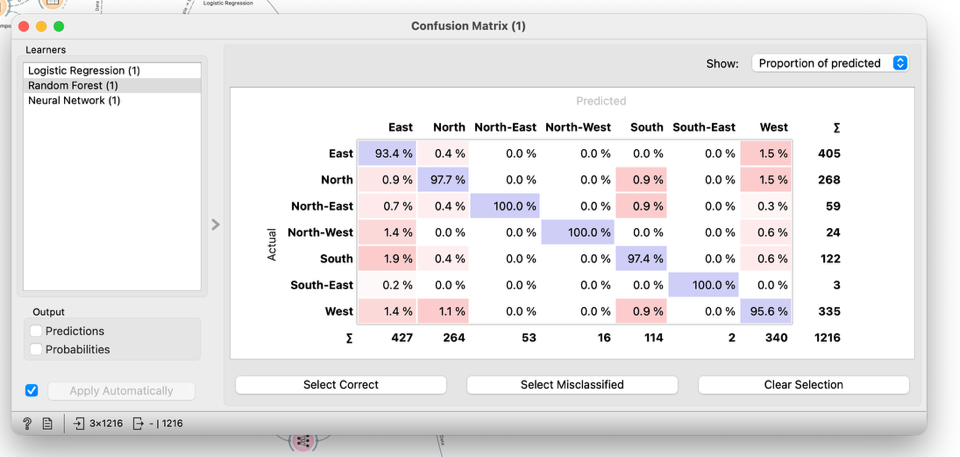
A multiple linear regression model was developed to quantify the influence of school characteristics on voluntary contributions. The dependent variable was annual contribution amount, while independent variables included:

* Location: Metropolitan vs regional classification (categorical variable).
* School Size Proxy: Approximated using total contributions as a proxy for enrolments.
* Locality: East, West, North, South regional categorization.

The regression analysis achieved a strong explanatory power (R² > 0.9), confirming that metropolitan classification and larger enrolments were significant predictors of higher contributions.

Forecasting

To project contributions beyond 2024, Power BI’s built-in forecasting feature was applied to the time series of contributions. Forecasts were generated for 2025 and 2026 with a 95% confidence interval and one-year seasonality parameter. This method extended observed trends and showed that, without intervention, the metro–regional gap would continue to widen.



I have also used ‘Orange’ software for finding some predictions in which we can see that the random Forest shows high values, and we can also see that the suburb voice percentage of prediction to get the funds values are also going up.

A screenshot of a computer

AI-generated content may be incorrect.

## Recommendations

The study refers to a persistent and widening metropolitan-to-regional Victorian school funding differential. To address the inequality, the following recommendations are made.

Second, specific funding interventions need to be instituted for regional schools. Government policy may offer additional grants to North and West regional schools with fewer voluntary contributions. These grants need to target essential resources such as classroom technology, extra-curricular activities, and specialist teachers so that rural community students are not disadvantaged by the limitation of community fundraising capacity.

Secondly, equity-based funding arrangements could be revised to account for variations in voluntary parental contributions. Current funding systems do not account properly for the advantage that accrues to metropolitan suburbs containing high-ICSEA schools. By accounting for contribution variations in funding arrangements, funds could be reallocated more evenly and reduce the inequity identified in this project.

Third, partnerships between the corporate and community worlds and underfunded or poorer-resourced schools need to be established. While metro schools enjoy access to more prosperous parental bases, regional schools may be helped through cooperation with regional industry, councils, and non-profit organizations. These types of arrangements may not eliminate differences but can provide vital additional assistance.

Analytically, regular monitoring via dashboards must be done. Power BI visualizations in this project showed just how easy interactive tools can render inequities visible and actionable. Those who are responsible for education need to take up similar tools to monitor year-on-year trends and hold people to account for gaps being closed.

One also needs to keep in mind the limits of the dataset: the figures are self-reported, end in 2022-2024, and can obviously not capture purely exogenous shocks such as inflation or policy change. Future research could help to increase the size of the dataset, include enrolment data, or employ more sophisticated machine learning techniques to better predict contributions.

Regarding dashboards, they should be utilized for ongoing monitoring. The visualizations developed from this project using Power BI were helpful in illustrating how differences can be seen and acted upon with relatively simple interactive tools. Educational leaders need to use similar tools to monitor annual trends and hold people accountable for closing the gaps.

Also, the dataset suffers from certain limitations: the numbers are self-reported, have an expiration date of 2022-2024 and clearly cannot capture purely exogenous shocks such as inflation and policy changes. Some potential improvements for future research are to expand the dataset, add enrolment data, or use more sophisticated machine learning methods to make better predictions of contributions.

## Data Ethics and Security

The project used school-level fiscal data, requiring careful consideration of privacy, accuracy, and transparency. The ethical responsibility was to report results in aggregated and comparative terms, protecting the identity and integrity of schools when addressing sensitive issues like funding inequities.

Transparency and honesty of visualizations were prioritized in analysis, with predictions and charts constructed truthfully without exaggerating differences. Confidentiality risks were minimal, but the application of educational financial data must adhere to ethical standards.

For future research, ethical frameworks must include clear standards for fairness, accountability, and responsible use, secure storage protocols, and strict governance policies for sensitive datasets. Future research should address secure storage protocols and strict governance policies for sensitive datasets.

## Reference

1. Bergamaschi, A., Giambruno, C., Morales, P., Forkosh-Baruch, A., Hanuka, S., & Hershkovitz, A. (2025). Empowering Schools with Data: How Can We Achieve Effective Use of Educational Dashboards for Teachers and Principals? <https://doi.org/10.18235/0013561>
2. Marshall, R., Pardo, A., Smith, D., & Watson, T. (2022). Implementing next generation privacy and ethics research in education technology. British Journal of Educational Technology, 53(4), 737–755. [https://doi.org/10.1111/bjet.13224](%20https:/doi.org/10.1111/bjet.13224)
3. Nikitha, D., Imran Hussain, S., & Poornima, M. (2024). Integrating Power BI with Machine Learning Models for Predictive Analytics. 2024 7th International Conference on Contemporary Computing and Informatics (IC3I), 1245–1250. <https://doi.org/10.1109/IC3I61595.2024.10828685>
4. Özemre, M., & Kabadurmus, O. (2020). A big data analytics based methodology for strategic decision making. Journal of Enterprise Information Management, 33(6), 1467–1490. <https://doi.org/10.1108/JEIM-08-2019-0222>
5. Pearce, S., Kristjansson, E., Lemyre, L., & Takacs, T. (2023). Understanding the volunteer motivations, barriers and experiences of urban and rural youth: a mixed-methods analysis. Voluntary Sector Review, 14(2), 268–292. <https://doi.org/10.1332/204080521X16418948258011>
6. Zhang, H., Chen, G., Ooi, B. C., Tan, K.-L., & Zhang, M. (2015). In-Memory Big Data Management and Processing: A Survey. IEEE Transactions on Knowledge and Data Engineering, 27(7), 1920–1948. [https://doi.org/10.1109/TKDE.2015.2427795](%20https:/doi.org/10.1109/TKDE.2015.2427795)
7. Annet, K. A. (2024). Public Funding Models for Education: Challenges and Solutions. Research Output Journal of Education, 4(3), 7–10. <https://doi.org/10.59298/ROJE/2024/43710>
8. Cooksey, R. W. (2020). Descriptive Statistics for Summarizing Data. In Illustrating Statistical Procedures: Finding Meaning in Quantitative Data (pp. 61–139). Springer Singapore. <https://doi.org/10.1007/978-981-15-2537-7_5>
9. Doorudinia, A., Dadashpoor, H., & Meshkini, A. (2025). Wrong polycentricity-right polycentricity? Insights from an empirical study in Tehran metropolitan region, Iran. Papers in Regional Science, 104(1), 100075. <https://doi.org/10.1016/j.pirs.2024.100075>
10. Gao, N., Saiedur Rahaman, M., Shao, W., & Salim, F. D. (2021). Investigating the Reliability of Self-report Data in the Wild: The Quest for Ground Truth. Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers, 237–242. <https://doi.org/10.1145/3460418.3479338>
11. Areed, W. D., Price, A., Arnett, K., Thompson, H., Malseed, R., & Mengersen, K. (2023, May 25). Assessing the Spatial Structure of the Association between Attendance at Preschool and Childrens Developmental Vulnerabilities in Queensland Australia. arXiv.org. Retrieved August 6, 2025, from https://arxiv.org/abs/2305.15746?
12. Blanden, J., Doepke, M., & Stuhler, J. (2022, April 10). Educational inequality. arXiv.org. https://arxiv.org/abs/2204.04701?
13. Gunawan, D., Griffiths, W., & Chotikapanich, D. (2021, August 29). Inequality in Education: A comparison of Australian indigenous and nonindigenous populations. arXiv.org. Retrieved August 6, 2025, from https://arxiv.org/abs/2108.12830?
14. Morris, A. & Institute for Public Policy and Governance, University of Technology Sydney, Sydney, NSW, Australia. (2024). Inequality and education in Australia. In The Economic and Labour Relations Review (pp. 1–22) [Journal-article]. Cambridge University Press. https://doi.org/10.1017/elr.2024.18
15. Rowe, E., & Perry, L. (2020, January 2). Inequalities in the private funding of public schools: parent financial contributions and school socioeconomic status. Figshare. Retrieved August 6, 2025, from https://dro.deakin.edu.au/articles/journal\_contribution/Inequalities\_in\_the\_private\_funding\_of\_public\_schools\_parent\_financial\_contributions\_and\_school\_socioeconomic\_status/24226381?
16. Rowe, E., & Perry, L. B. (2019). Private financing in urban public schools: inequalities in a stratified education marketplace. The Australian Educational Researcher, 47(1), 19–37. https://doi.org/10.1007/s13384-019-00328-0
17. scheme=AGLSTERMS.AglsAgent; corporateName=Department of Education; address=50 Marcus Clarke St, Canberra City, ACT 2601; contact=+61 1300 566 046. (2025, July 29). How schools are funded. Department of Education. Retrieved August 6, 2025, from https://www.education.gov.au/schooling/how-schools-are-funded?