

Teaching Math Effectively: Insights from LSAY and PISA on Primary Teacher Proficiency

Tashya Sathyajit (29672732)

Table of contents

1	Abstract	2
2	Background and motivation	3
3	Objectives and significance	4
4	Methodology	6
4.1	How did we identify primary school teachers?	6
4.1.1	What did these teachers study?	6
4.2	How did we explore teacher performance relative to other professionals? . . .	9

4.3	How did we complete the Mathematical Proficiency Analysis?	9
4.4	Working with Sampling Weights and Plausible Values Accurately	11
5	Results	13
5.0.1	How well did individuals who became primary school teachers perform in mathematics across the years?	13
5.0.2	How do teachers' math scores compare to the rest of the sample (non primary school teachers)?	14
5.0.3	How are they performing in comparison to other 'professionals'? . . .	17
6	Discussion, Limitations and Future Research	19
7	References	24

1 Abstract

This report investigates the mathematics performance of individuals who became primary school teachers in Australia, using the Longitudinal Surveys of Australian Youth (LSAY) and the Program For International Student Assessment (PISA) datasets. This analysis was done for the contribution to Grattan Institute's report on teacher proficiency in mathematics. Results show that teachers consistently have a higher average math proficiency than non-teachers over time, with better performance across PISA testing years. However, they tend to score lower than professionals, particularly in STEM fields. Limitations include disproportionate teacher counts in some PISA years, potential attrition bias, confounding

variables such as socioeconomic status, and the evolving nature of math proficiency over time.

2 Background and motivation

Over the past two decades, Australian students' performance in the mathematics components of international assessments like PISA have shown significant declines (Morsy et al., 2018) as seen in Figure 1. These declines persist even after adjusting for changes in family academic resources, affecting students across all states, social groups, and school types (albeit with variations). This trend has raised concerns among educators, researchers and policymakers as mathematics proficiency is not only foundational for broader academic success but crucial for Australia's future economic competitiveness.

Research indicates a positive correlation between teacher mathematics knowledge (as assessed through certification exams or subject-matter tests) and student achievement (Hill et al., 2005). This underscores the importance of qualified primary school teachers in improving children's math outcomes, supporting the notion that teachers with stronger mathematical knowledge are likely to teach it more effectively (Norton, 2019).

While research consistently suggests that a teacher's proficiency in mathematics is a crucial determinant of their effectiveness in the classroom, impacting student achievement directly (Morsy et al., 2018) the relationship between a teacher's subject knowledge and student performance is complex, and extends beyond the teacher's own content mastery. It involves the ability to translate that knowledge into effective pedagogical practices, adapt to curriculum

needs, engage in productive professional conversations and address student misconceptions. When students are taught by teachers who have little mathematical content *and* pedagogical content knowledge, learning suffers (Baumert et al., 2010; Hill et al., 2005).

Despite these nuances, the positive correlation between teachers' competence in mathematics and student achievement is clear (Hill et al., 2005). While this correlation does not strictly imply causation, understanding the mathematics proficiency of primary school teachers remains a critical step in improving student outcomes.

This analysis seeks to deepen the understanding of the mathematics proficiency of individuals who eventually become primary school teachers in Australia by leveraging longitudinal data from PISA and LSAY. By examining trends over time and comparing the math performance of primary school teachers, the broader Australian population and other professionals, this study aims to provide insights into the readiness of teachers to deliver effective math instruction. Understanding these trends is crucial, as they provide insights into the mathematics performance of individuals who later became primary school teachers, and may assist in guiding professional development initiatives focused on targeting and strengthening the mathematics proficiency of primary school teachers, ultimately leading to improved student outcomes across Australia.

3 Objectives and significance

The primary objective of this analysis was to utilise longitudinal survey data to understand the mathematics proficiency of students who eventually became primary school teachers in

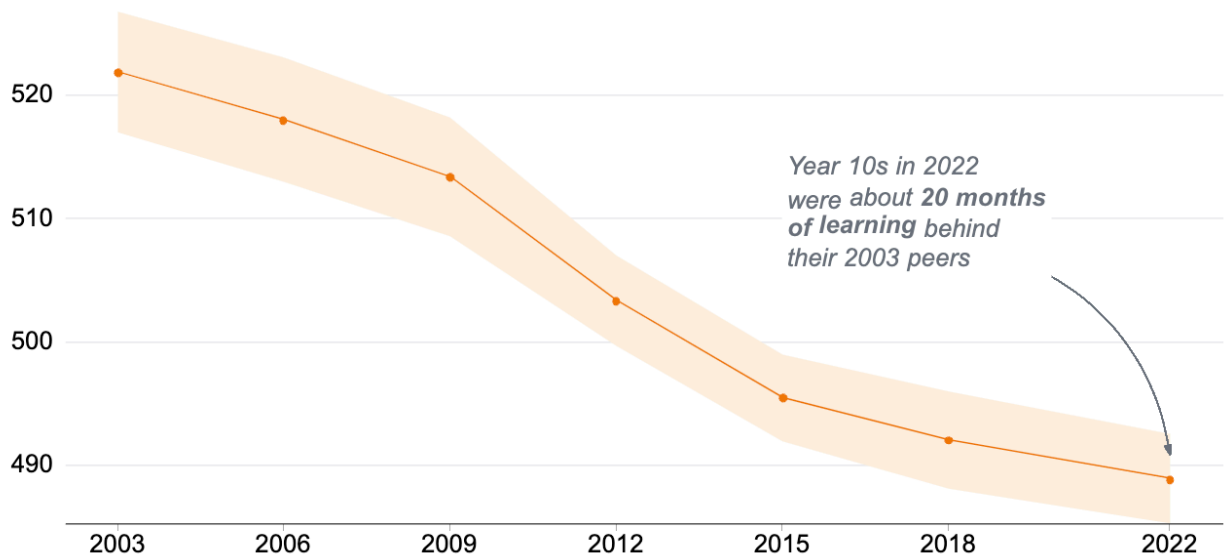


Figure 1: Decline in PISA Scores Across the Years

Australia. Using the LSAY and PISA datasets, the analysis seeks to address three main questions:

1. **Teacher performance over time:** How well did individuals who became primary school teachers perform in mathematics across the years?
2. **Comparison to other Professionals:** How are they performing in comparison to other ‘professionals’?
3. **Comparison to the General Population:** How do teachers’ math scores compare to the rest of the sample (non-primary school teachers)?

This analysis is significant as it provides critical insights into the math skills of individuals who become primary school teachers, a factor that can directly impact student outcomes. By identifying trends and proficiency gaps, the analysis can allow for a deeper understanding of the performance of primary school teachers in Australia and inform potential policy recommendations.

4 Methodology

Data was sourced from multiple LSAY cohorts: 2003, 2006, 2009 and 2015. Each dataset contains demographic data from student responses and their linked performance in PISA. These years were selected because, from 2003 onwards, LSAY participants were recruited from schools that also took part in PISA, allowing for the linkage of PISA scores to LSAY data.

4.1 How did we identify primary school teachers?

To identify primary school teachers for each cohort, we examined variables related to occupation across each LSAY cohort. The [data dictionary](#) was used to locate the relevant variables, as found in Table 1. These variables correspond to responses about the type of work the respondents did.

Individuals were flagged as primary school teachers if they reported an ANZSCO code of 2412, which represents primary school teachers. This flag distinguished those who eventually became primary school teachers from those who did not. Respondents with other occupation codes were classified as non-teachers.

4.1.1 What did these teachers study?

To understand the educational backgrounds of individuals who later became primary school teachers, fields of study were analysed using variables detailed in Table 2. The analysis

Table 1: Variables for ‘What kind of work do you do in this job?’ Questions

Year	Variables
2003	LDD025, LED025, LFD025, LGD025, LHD029, LID029, LJD029, LKD029
2006	ANZSCO, LBD011, LCD024, LDD024, LED029, LFD029, LGD029, LHD029, LID029, LJD029, LKD029
2009	ANZSCO73, LBD014, LCD029, LDD029, LED029, LFD029, LGD029, LHD029, LID029, LJD029, LKD029
2015	LSAYID, LBEM004, LCD022, LDD022, LED022, LFD025, LGD022, LHD022

Table 2: Variables for Main Area of Study Questions

Year	Variables
2003	LBCZ087, LCCZ087, LDCZ087, LEC087, LFC087, LGC087, LHC087, LIC087, LJC088, LKC088
2006	LBC087, LCC087, LDC087, LEC087, LFC087, LGC088, LHC088, LIC088, LJC088, LKC088
2009	LBC089, LCC088, LDC088, LEC088, LFC088, LGC088, LHC088, LIC088, LJC088, LKC088
2015	LCC088, LDC088, LEC088, LFC088, LGC088, LHC088

focused on identifying whether respondents pursued primary school education, coded as 7301 in accordance with the [Australian Standard Classification of Education \(ASCED\)](#).

Additionally, the analysis explored whether respondents who became teachers had previously studied fields outside of education before transitioning to teaching. This was done with a similar approach of examining ASCED scores reported across waves of those who were identified as teachers. Table 3 provides a summary of the fields of study identified and highlights cases where individuals shifted from other fields to primary education.

Table 3

Code	Field_of_Education
70101	Teacher Education: Early Childhood
70103	Teacher Education: Primary
70109	Teacher Education: Vocational Education and Training
70199	Teacher Education, n.e.c.
70115	English as a Second Language Teaching
90503	Children's Services
91703	Religious Studies
92101	Sport and Recreation Activities
100101	Music
80101	Accounting
100705	Written Communication
100103	Drama and Theatre Studies
80399	Business and Management, n.e.c.
70105	Teacher Education: Secondary
91701	Philosophy
60901	Optometry
61399	Public Health, n.e.c.
60101	General Medicine
81101	Banking and Finance
91523	Literature
90513	Counselling
90701	Psychology
79999	Education, n.e.c.
60501	Pharmacy
61799	Rehabilitation Therapies, n.e.c.

4.2 How did we explore teacher performance relative to other professionals?

This part of the analysis was driven more by curiosity than formal research aims, exploring how primary school teachers' mathematics proficiency compares to that of other professionals. Due to this, the analysis focused only on the 2003 PISA cohort which had the largest sample of teachers (view Figure 7) and of that cohort, the final wave. By examining the last wave of this cohort, the analysis captures participants' final professional roles, though this approach results in a lower teacher count than the total count for the entire 2003 cohort as seen in Table 4.

To define professional categories, the analysis used the ANZSCO codes sourced from the Australian Bureau of Statistics [website](#). The respondent's occupation code response to variable LKD029 corresponding to LSAY question relating to kind of work was used to identify whether they belonged to one of the professional groups of interest. A new variable was established which created a flag for individuals based on their occupation code (e.g. Teachers coded as '1', while other professional groups were coded '2' to '6' in accordance to their ANZSCO code).

4.3 How did we complete the Mathematical Proficiency Analysis?

The mathematics performance of primary school teachers was assessed using PISA scores. The analysis aimed to compare teachers' scores against the national proficiency standard which has been set at a [proficiency level 3](#) or [482 PISA points](#). The analysis incorporated

Table 4

Table 5 Professionals, Corresponding ANZSCO Codes, and Counts

Professionals	ANZSCO Code	Count
Primary School Teacher	2412	82
Medical Practitioners	253	48
Midwifery and Nursing Professionals	254	94
Legal Professionals	271	50
Accounts Auditors and Company Secretaries	221	92
Engineering Professionals	233	88

survey weights to ensure representativeness of the data and address differences in sample sizes across cohorts. It also had to account for the use of plausible values undertaken by PISA.

i At Level 3 Proficiency in Mathematics

“At Level 3, students can execute clearly described procedures, including those that require sequential decisions. Their interpretations are sufficiently sound to be a base for building a simple model or for selecting and applying simple problem-solving strategies. Students at this level can interpret and use representations based on different information sources and reason directly from them. They typically show some ability to handle percentages, fractions and decimal numbers, and to work with proportional relationships. Their solutions reflect that they have engaged in basic interpretation and reasoning.”

4.4 Working with Sampling Weights and Plausible Values Accurately

International surveys such as PISA report student performance using plausible values (PVs). Plausible values are a statistical technique that represents uncertainty in measurements by generating a range of possible values a student may have obtained as opposed to a single estimate. This approach assists in conveying the variability in student performance more accurately, considering that not all students answer the same questions (PISA is a large test) or have the same background information.

To derive PVs, PISA uses [Item Response Theory \(IRT\)](#), which predicts the likelihood of a student answering a question correctly by assessing the difficulty of each question and how well it distinguishes between students of different ability levels. For example, if the question is deemed a relatively conceptually ‘easy’ question, and the estimated ability of the student is high, the model predicts a high probability of a correct response. PISA can determine if a question is ‘easy’ or ‘hard’ based on the statistical analysis of response patterns from a broad group of students, using IRT to estimate how challenging each question is relative to the separate proficiency levels, as can be seen [here](#) (OECD, 2023).

Looking at Figure 2 retrieved from [Wu \(2022\)](#), the process starts with a prior distribution. This is an initial estimate of the student’s proficiency based on background variables like socioeconomic status or parental education. This prior estimation serves as a rough starting point that as the students begin to answer questions, get finetuned as PISA updates this estimate using student responses and resulting in a posterior distribution. This updating distribution combines that initial prior estimate with the observed test results.

After this, multiple imputed values are taken from this posterior distribution. These plausible values are thus a representation of the range of abilities a student may reasonably have (OECD, 2009) and are normally five to 10 plausible values per student for each domain.

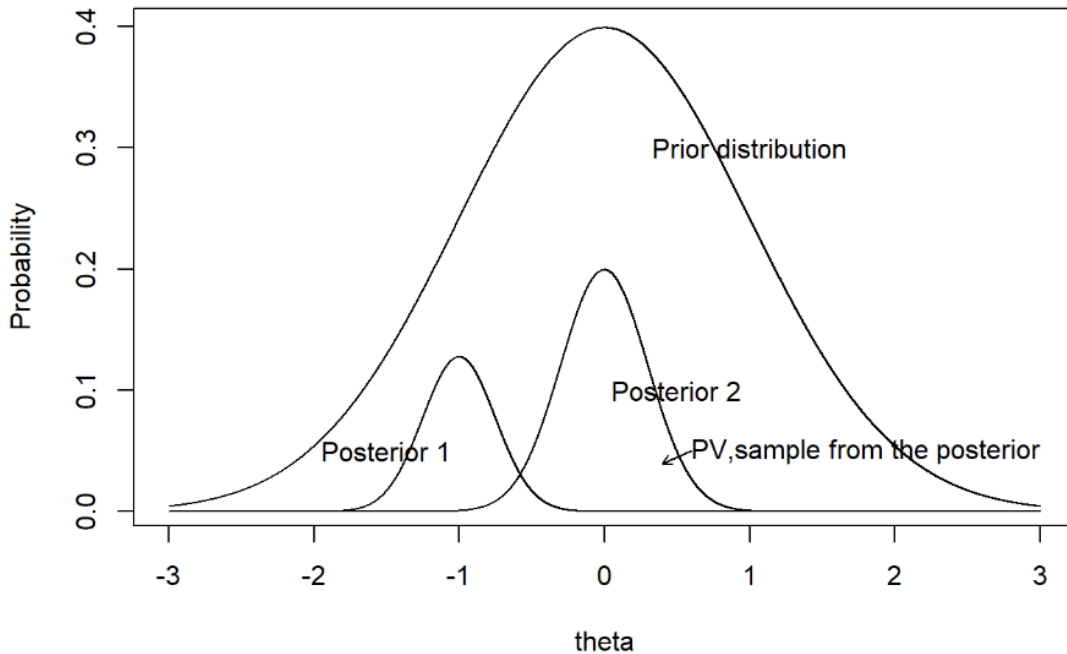


Figure 2: Prior and Posterior Distributions

The `intsvy` and `rrepest` R packages were employed to handle the complex survey design of PISA data, ensuring accurate estimation of statistics such as mean scores, confidence intervals, and standard errors. These packages allowed for the calculation of weighted averages, taking into account the total number of teachers and non-teachers in each cohort to provide a more representative estimate of the average proficiency scores.

As PISA uses sampling weights to account for its stratified, multi-staged sampling design which includes adjustments for things such as non-response, school selection and student participation, the `Rrepest` package applies these weights when calculating summary statistics.

It will iteratively apply each set of replicate weights to generate multiple estimates. These estimates are then combined to produce accurate standard errors that reflect the variability introduced by PISA's sampling structure. This approach also ensures that results are not biased by unequal probabilities of selection and attempt to provide more representativeness.

In the case of this analysis, the `Rrepest` package was utilised mostly in determining our summary statistics, although results were cross checked with the `intsvy` package. The package included functions that cycled through each plausible value, averaging the results while simultaneously applying the survey weights to maintain representativeness. This combined approach of incorporating plausible values and weights ensures that the estimates are statistically valid, accounting for uncertainty in measurement and complex survey design. It was also [recommended](#) by the OECD for handling PISA data and plausible values.

5 Results

As earlier mentioned in the report, the project aimed to explore three key questions regarding the mathematical performance of individuals who eventually became primary school teachers.

5.0.1 How well did individuals who became primary school teachers perform in mathematics across the years?

The analysis reveals that individuals who eventually became primary school teachers consistently performed better in mathematics compared to their non-teaching peers over the

years. To quantify this, the weighted average proficiency of teachers and non-teachers was calculated, considering the total number of each group to ensure more accurate average scores.

The results, as shown in Figure 3, indicate that just over 80% of teachers met or exceeded the national proficiency standard of 482 PISA points, compared to just over 60% of non-teachers. This substantial difference suggests that individuals who eventually become primary school teachers possess stronger mathematical skills than their non-teaching peers from the same cohort. These results highlight the higher mathematical readiness of future teachers, which may have implications for their effectiveness in the classroom.

However, an important limitation to consider is that the smaller sample of teachers may reduce the observed variability, potentially skewing results if high-performing individuals are overrepresented. Figure 4 shows that in the 2003 sample, there is less variability within the teacher group compared to the non-teacher group. This potential selection bias could lead to an inflated perception of average teacher math proficiency, as the sample might not fully capture the broader range of abilities likely found in the general teacher population.

5.0.2 How do teachers' math scores compare to the rest of the sample (non primary school teachers)?

Figure 5 shows that teachers tend to perform better than non-teachers across each year of testing.

This is demonstrated by a consistently higher median score for teachers (red) compared to

Table 6

Table 7PISA Teacher vs Non-Teacher Math Proficiency

Year	Teacher	Mean	SE_Mean	Std	SE_Std
2003	False	530.3	2.4	92.9	2.4
2003	True	554.0	10.5	64.8	5.3
2006	False	518.6	3.0	87.9	1.2
2006	True	550.7	10.4	67.2	8.2
2009	False	513.7	2.7	95.1	2.5
2009	True	552.7	7.2	63.9	8.4
2015	False	494.0	1.8	92.5	0.9
2015	True	519.4	20.2	61.0	19.0

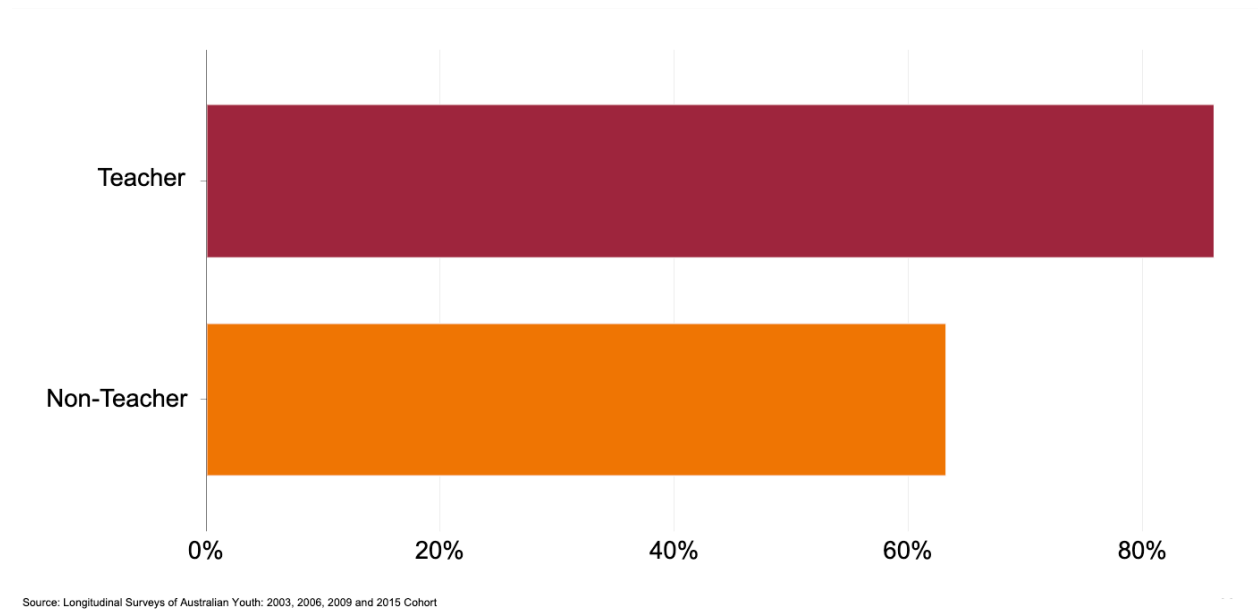


Figure 3: Weighted Average of Performance Across Years: Teachers vs Non-Teachers

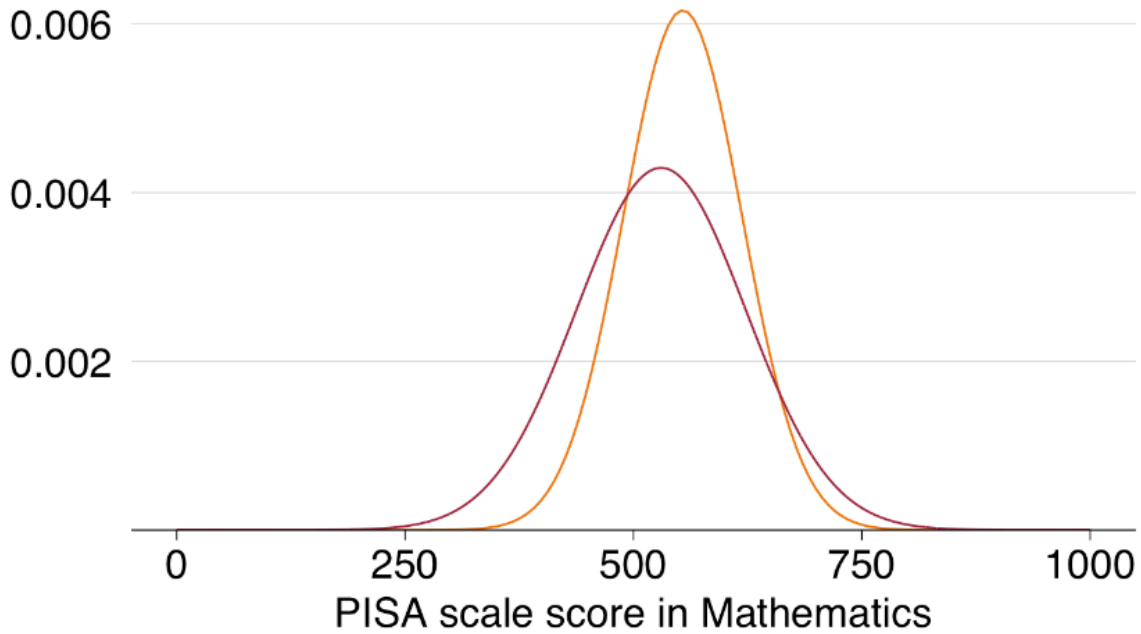


Figure 4: Probability Distribution of 2003 Mathematics PISA Scores

non-teachers (orange) over the years. The boxplots reveal differences in score distribution, with non-teachers showing longer whiskers, indicating a broader range of math scores and greater variability. This wider spread suggests that non-teachers have a more diverse range of mathematical abilities.

In contrast, the boxplots for teachers display slightly shorter whiskers, indicating more consistent scores among this group. However, as seen in Table 6, the standard error (SE) for teachers is higher, likely due to the smaller sample size. This increases the uncertainty in estimating their average score but does not imply greater variability within the teacher group.

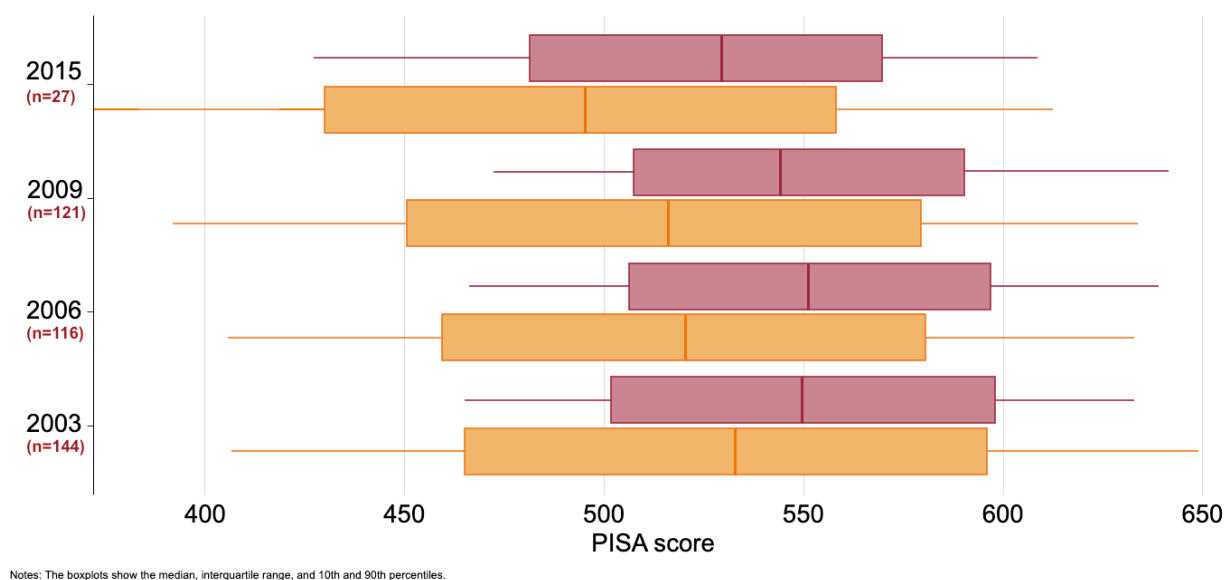


Figure 5: Proficiency of Teachers vs Non-Teachers Across Years

5.0.3 How are they performing in comparison to other ‘professionals’?

The project had a smaller focus on understanding how teachers’ mathematical proficiency compared to other professionals, driven largely by curiosity rather than formal research aims. The selection of professional categories was based on existing reports, and the analysis was conducted on the 2003 PISA cohort, the year with the largest sample of teachers. However this still had its limitations due to the smaller sample size of the relative professional groups within a total sample of 10,370 as seen in Table 4.

Figure 6 represents the average PISA scores for teachers and other professionals, along with 90 percent confidence intervals. The analysis suggested that teachers performed lower than most other professional groups in terms of average mathematical proficiency at age 15. Professions such as medical practitioners, legal professionals, engineers and auditors scored higher on average, indicating better mathematical capability among these groups. However,

teachers scored higher than the ‘Other’ category, which represents the broader population not within aforementioned professions. This suggests that while teachers were not among the highest performing professional groups, they seemed to demonstrate better mathematical proficiency than the general cohort.

Despite these findings, it is important to consider the larger confidence intervals observed for some professional groups, such as legal professionals, engineers and auditors. These wider intervals primarily reflect greater uncertainty in the estimates due to smaller sample sizes and potential sample bias within these groups. This increased uncertainty makes precise comparisons more challenging and difficult to ascertain with confidence the true average performance of these professional groups.

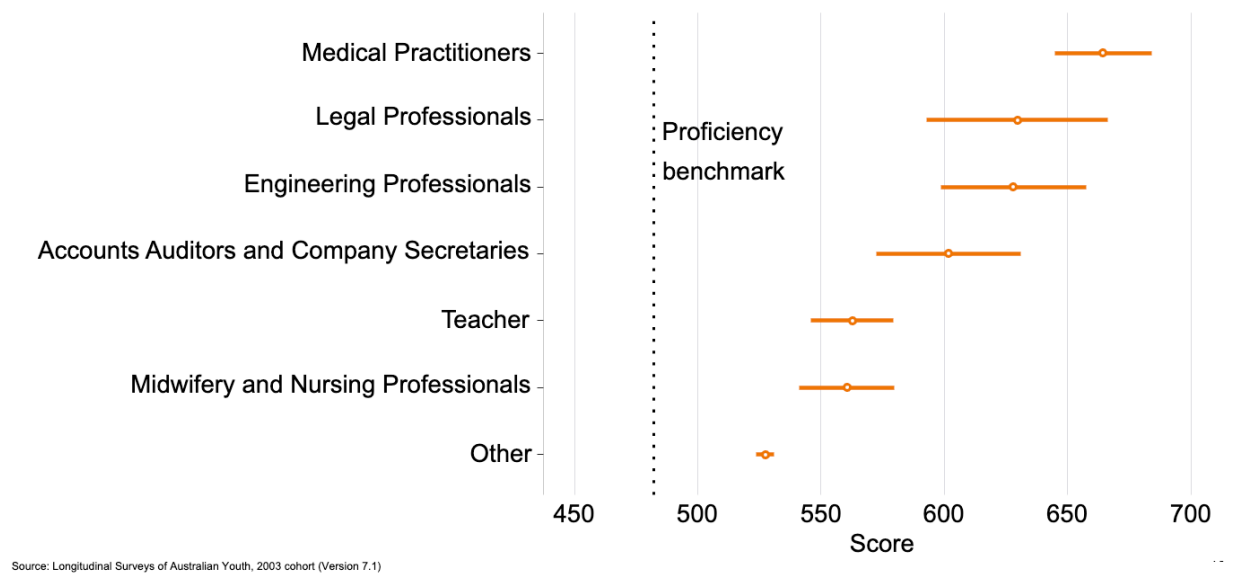


Figure 6: Teachers in Comparison to Other Professionals

6 Discussion, Limitations and Future Research

The findings of this analysis provide insights into the mathematical proficiency of individuals who eventually become primary school teachers, indicating a generally higher proficiency than their non-teaching peers across multiple years. Teachers tend to meet or exceed national proficiency standards more frequently than non-teachers, suggesting a relatively stronger foundation in mathematical skills. This trend is evident over time, as teachers consistently achieve higher median scores with less variability compared to non-teachers, as seen in Figure 5.

The more consistent performance among teachers is suggested by the slightly shorter whiskers in the boxplots, indicating a relatively uniform level of mathematical competence. In contrast, the longer whiskers for non-teachers reflect greater variability, indicating a broader range of skills. However, it is important to note that the whiskers for both groups are relatively long, suggesting a wide distribution of scores overall, with non-teachers demonstrating only slightly more spread. The higher standard error for teachers, meanwhile, is primarily due to their significantly smaller sample size, increasing uncertainty in the estimated average scores. This however, reflects the inherent limitations of sample size than actual variability in proficiency.

When comparing teachers to other professional groups, as seen in Figure 6, the analysis reveals that while future teachers performed better than the general population, they performed lower than medical practitioners, legal professionals, engineers, and auditors and performed quite similarly to midwifery and nursing professionals. This aligned with a pre-

conceived idea that STEM professions would attract students with more confidence or proficiency in mathematics. However, the large confidence intervals observed for some professional groups highlight greater uncertainty of where the true average lies due to smaller sample sizes and potential variability within these groups.

However, there are some considerations that are important to pay mind to upon the results of this analysis. Despite the relatively strong performance of teachers in mathematics, as aforementioned, effective teaching requires more than just content knowledge. Research frameworks emphasise that teachers must understand how to present content effectively, integrate it with the curriculum, and adapt it to diverse learning styles (Norton, 2018). Additionally, although we can see that individuals who became teachers in later life had a higher proficiency in mathematics when compared to the general population (save for the professions explored deeper), there are limitations to the analysis.

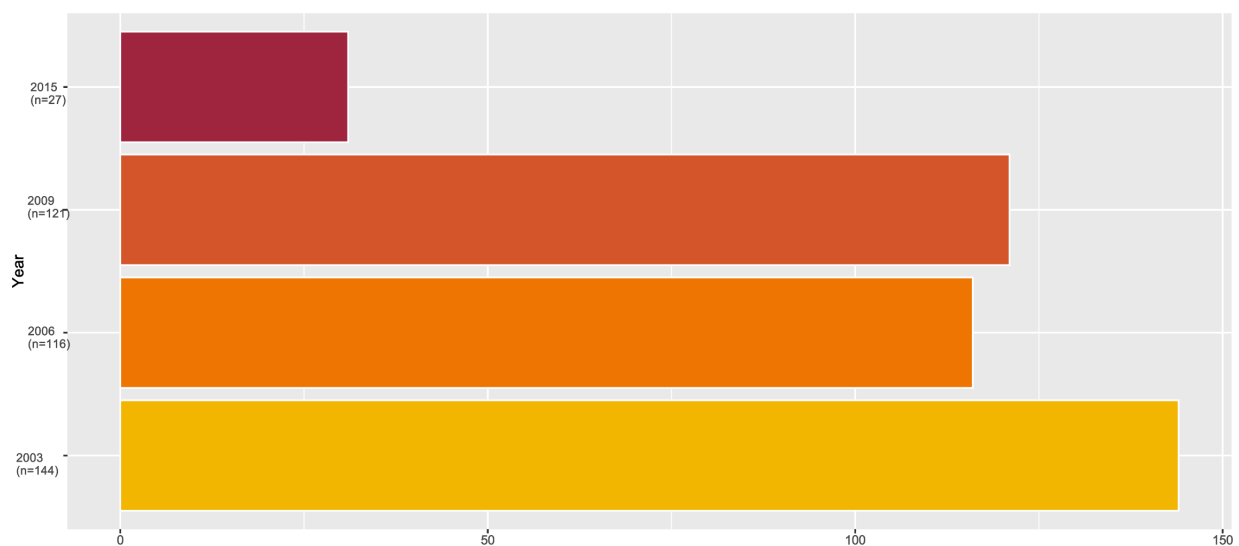


Figure 7: Count of Teachers per Year

Small sample size for teachers and professional groups:

The sample size for teachers as seen in Figure 7 and other professionals as seen in Table 4 was relatively small, which limits the robustness of the findings. The smaller sample size is also associated with larger standard errors, increasing uncertainty in the estimated averages of those groups. The lack of statistical power due to small sample sizes may have affected the precision of the estimates, potentially leading to biased results or challenges in generalising the results.

Attrition Rate in Longitudinal Data:

The analysis is based on longitudinal data, which is inherently prone to attrition over time. As participants drop out of the study, the sample size decreases which may introduce potential bias and affect the reliability of the results. Further, the later PISA cohorts haven't completed all the waves which contributes to the smaller sample size for teachers in the later years. This attrition bias may have influenced the results, particularly in the later years, and may have implications for the generalisability of the findings.

Scope of PISA:

As mentioned prior, while PISA tests part of the expected learning outcomes of for mathematics, it does not test how good an individual will be at conveying this knowledge to others or the entire range of mathematical knowledge required in teacher training or professional practice. This could suggest that although individuals who performed well in PISA who later became teachers may be good at mathematics, they may not necessarily be good at teaching it.

Additionally, PISA only tracks individuals from the age of 15 to 25, and this does not

account for the many other ways individuals go into the pathway of being teachers. For example, mathematical proficiency can be enhanced through alternative pathways or lived work experience before the individual moved into teaching. This could also suggest that the analysis may not capture the full range of individuals who become teachers and their mathematical proficiency.

High Standard Error:

The high standard errors observed in the estimates for teachers and some pro

For the project, to determine whether differences in mathematical proficiency between teachers and non-teachers statistically significant, a series of Welch's t-tests were conducted for each year. A Welch's t-test was deemed sufficient for this analysis, as it accommodates differences variances and sample sizes between groups. The t-tests generally showed statistically significant differences in average scores, with teachers consistently outperforming non-teachers in the years analysed. However, this should be treated as preliminary statistical analysis due to limitations prior mentioned and methodological considerations behind the reasoning to not include the results of these t-tests include:

- **Small sample size:** There are relatively very small sample sizes for teachers, especially in latter years due to not all waves being released for the cohort (e.g. there are only 27 teachers in the 2015 sample).
- **Normality assumption:** The t-test assumes a normal distribution of scores within each group, but this assumption may not hold, especially given the small teacher samples. This could affect the validity of the test results. As seen in Figure 4, which

represents *only* the 2003 sample, the distribution of teacher scores (orange) shows slight skewness and a narrower spread, indicating potential deviations from normality. The reduced sample size may also underrepresent true variability among teacher scores, further complicating interpretation.

Next steps could build upon this with the potential use of non-parametric tests or bootstrapping techniques.

This analysis reveals that individuals who later became primary school teachers generally exhibit higher mathematical proficiency than non-teaching peers, consistently meeting or exceeding national standards. However, limitations such as smaller sample sizes, attrition rates, and PISA's narrow scope in terms of age and content coverage must be acknowledged. Despite these constraints, meaningful trends can be seen. In particular, teachers achieve stronger median scores and less variability. Yet, they still fall behind STEM professionals.

Although these findings suggest a strong basis for the individuals in this sample that eventually became teachers, teacher effectiveness extends beyond proficiency, emphasising the need for future research to include pedagogical skills and statistical tests that can determine with more confidence whether the difference between teachers and non-teachers are statistically significant. Thus, while the findings offer insights into teachers' mathematical strengths, they remain preliminary and should be used in conjunction with assessments that indicate how well these individuals would be able to teach and communicate mathematical concepts effectively.

7 References

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., Klusmann, U., Krauss, S., Neubrand, M., & Tsai, Y.-M. (2010). Teachers' Mathematical Knowledge, Cognitive Activation in the Classroom, and Student Progress. *American Educational Research Journal*, 47(1), 133-180. <https://doi.org/10.3102/0002831209345157>

Caro DH, Biecek P (2017). "intsvy: An R Package for Analyzing International Large-Scale Assessment Data." *Journal of Statistical Software*, 81(7), 1-44. doi:10.18637/jss.v081.i07.

Cowgill M (2024). ggdirectlabel: Make it Easier to Directly Label ggplot2 Charts Rather Than Using Legends. R package version 0.1.0.901.

Cowgill M, Mackey W (2024). grattantheme: Create Graphs in the Grattan Institute Style. R package version 1.2.0.

Gotsis, T. (2020). NSW school education: PISA 2018, socioeconomic background and proposals for reform. Parliament of New South Wales. <https://www.parliament.nsw.gov.au/researchpapers/Pages/school-education-PISA-2018.aspx>

Grolemund G, Wickham H (2011). "Dates and Times Made Easy with lubridate." *Journal of Statistical Software*, 40(3), 1-25. URL: <https://www.jstatsoft.org/v40/i03/>.

Hester J, Bryan J (2024). glue: Interpreted String Literals. R package version 1.7.0, <https://CRAN.R-project.org/package=glue>.

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of Teachers' Mathematical Knowledge for Teaching on Student Achievement. *American Educational Research Journal*, 42(2), 371-406. <https://doi.org/10.3102/00028312042002371>

Ilizaliturri R, Avvisati F, Keslair F (2023). Rrepest: An Analyzer of International Large Scale Assessments in Education. R package version 1.3.0, <https://CRAN.R-project.org/package=Rrepest>.

Kay M (2024). "ggdist: Visualizations of Distributions and Uncertainty in the Grammar of Graphics." *IEEE Transactions on Visualization and Computer Graphics*, 30(1), 414-424. [doi:10.1109/TVCG.2023.3327195](https://doi.org/10.1109/TVCG.2023.3327195).

Lumley T (2004). "Analysis of Complex Survey Samples." *Journal of Statistical Software*, 9(1), 1-19.

Lumley T (2010). *Complex Surveys: A Guide to Analysis Using R*. John Wiley and Sons.

Lumley T (2024). *survey: analysis of complex survey samples*. R package version 4.4.

Morsy, L., Khavenson, T., & Carnoy, M. (2018). How international tests fail to inform policy: The unsolved mystery of Australia's steady decline in PISA scores. *International Journal of Educational Development*, 60, 60-79. <https://doi.org/10.1016/j.ijedudev.2017.10.018>

Norton, S. (2019). The relationship between mathematical content knowledge and mathematical pedagogical content knowledge of prospective primary teachers. *Journal of Mathematics Teacher Education*, 22(5), 489-514. <https://doi.org/10.1007/s10857-018-9401-y>

- OECD. (2009). “Plausible Values”, in OCED (Ed.), PISA Data Analysis Manual: SPSS (2nd ed., pp. 93-101). OECD Publishing. <https://doi.org/10.1787/9789264056275-7-en>
- OECD. (2019). What can students do in mathematics? In OCED (Ed.), PISA 2018 Results (pp. 103-110). OECD Publishing. <https://doi.org/10.1787/f649d3c2-en>
- OECD. (2023). What can students do in mathematics, reading and science? In OCED (Ed.), PISA 2022 Results (pp. 88-108). OECD Publishing. <https://doi.org/10.1787/02f44b44-en>
- Okubo, T. (2022). “Theoretical considerations on scaling methodology in PISA.” OECD Education Working Papers (Working Paper No. 282). OECD Publishing. <https://doi.org/10.1787/c224dbeb-en>
- O’Hara-Wild M, Kay M, Hayes A (2024). distributional: Vectorised Probability Distributions. R package version 0.4.0, <https://CRAN.R-project.org/package=distributional>.
- Sievert C (2020). Interactive Web-Based Data Visualization with R, plotly, and shiny. Chapman and Hall/CRC Florida.
- Slowikowski K (2023). ggrepel: Automatically Position Non-Overlapping Text Labels with ggplot2. R package version 0.9.3, <https://CRAN.R-project.org/package=ggrepel>.
- Wickham H (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.
- Wickham H, Averick M, Bryan J, et al. (2019). “Welcome to the tidyverse.” Journal of Open Source Software, 4(43), 1686. doi:10.21105/joss.01686.

Wickham H, Bryan J (2023). readxl: Read Excel Files. R package version 1.4.3, <https://CRAN.R-project.org/package=readxl>.

Wickham H, François R, Henry L, Müller K, Vaughan D (2023). dplyr: A Grammar of Data Manipulation. R package version 1.1.3, <https://CRAN.R-project.org/package=dplyr>.

Wickham H, Henry L (2023). purrr: Functional Programming Tools. R package version 1.0.2, <https://CRAN.R-project.org/package=purrr>.

Wickham H, Miller E, Smith D (2023). haven: Import and Export ‘SPSS’, ‘Stata’ and ‘SAS’ Files. R package version 2.5.4, <https://CRAN.R-project.org/package=haven>.

Wickham H, Pedersen T, Seidel D (2023). scales: Scale Functions for Visualization. R package version 1.3.0, <https://CRAN.R-project.org/package=scales>.

Wilke C, Wiernik B (2022). ggtext: Improved Text Rendering Support for ‘ggplot2’. R package version 0.1.2, <https://CRAN.R-project.org/package=ggtext>.

Wu, M. (2022, June 7). Chapter 14 Estimating Population Characteristics - Part II: Plausible Values. A Course on Test and Item Analyses. Retrieved October 28, 2024, from <https://www.edmeasurementsurveys.com/Ch14.html>

Zhu H (2021). kableExtra: Construct Complex Table with ‘kable’ and Pipe Syntax. R package version 1.3.4, <https://CRAN.R-project.org/package=kableExtra>.