Australia's Educational Divide

Edwin Mellett

December 8, 2017

Abstract

This report is a comparative study of the effect of student dispositions on mathematics literacy in Australia between public and private schools. There is a noticeable divide between the performance of public and private school students. Supervised machine learning is used to identify which parts of the students' dispositions are important in determining math literacy. Based on the machine learning generated models, student dispositions of self-efficacy and math anxiety have the greatest effect on math scores. Self-efficacy is more favorable among private school students and could be one of the major reasons private schools outperform public schools in Australia.

Introduction

The data for this report comes from the Programme for International Student Assessment (PISA) 2012 results. PISA occurs every three years and tests 15 year-old students around the world in math, science, and reading (OECD 2017). The data collected includes test scores, school information, and student and parent information. Part of PISA includes a questionnaire for students with questions ranging from demographics to psychological dispositions.

I am interested in looking at the effect of student disposition on mathematics literacy in Australia between public and private schools. In this case, student disposition can be thought of as the way a student feels about mathematics and his or her mindset and confidence towards solving math problems. There have been a few interesting findings relating to the effects of student dispositions that show a student's self-efficacy has a noticeable effect on math scores (Gabriel, Signolet, and Westwell 2017). Another study conducted in Australia found "private schools achieve better results than public schools" (Buckingham 2000). Buckingham's measures of achievement included test scores and high school graduation percentages. Since PISA contains a plethora of useful information, we can expand on previous research to see if differences in psychological dispositions between public and private school students could help explain the performance gap that Buckingham and others have talked about.

Methods

The statistical analyses in this paper were all done using R (R Core Team 2017). The following R packages were used: h2o (team 2017), data.table (Dowle and Srinivasan 2017), ggplot2 (Wickham 2009), kableExtra (Zhu 2017), and magrittr (Bache and Wickham 2014).

The machine learning approach in this report is heavily influenced by a paper produced by my professor Jason Signolet and his colleagues (Gabriel, Signolet, and Westwell 2017). A gradient boosting machine (GBM) was used to determine possible variables of importance and k-folds was used for cross-validation. 22 of the survey questions relating to student disposition were used as features for the GBM. The survey questions chosen fell into 3 categories: math self-efficacy, math anxiety, and perceived control. The mathematics literacy scores were calculated as the aggregate of the 5 plausible values (PV1MATH, PV2MATH, PV3MATH, PV4MATH, PV5MATH) weighted by the final student weight (W_FSTUWT).

Results

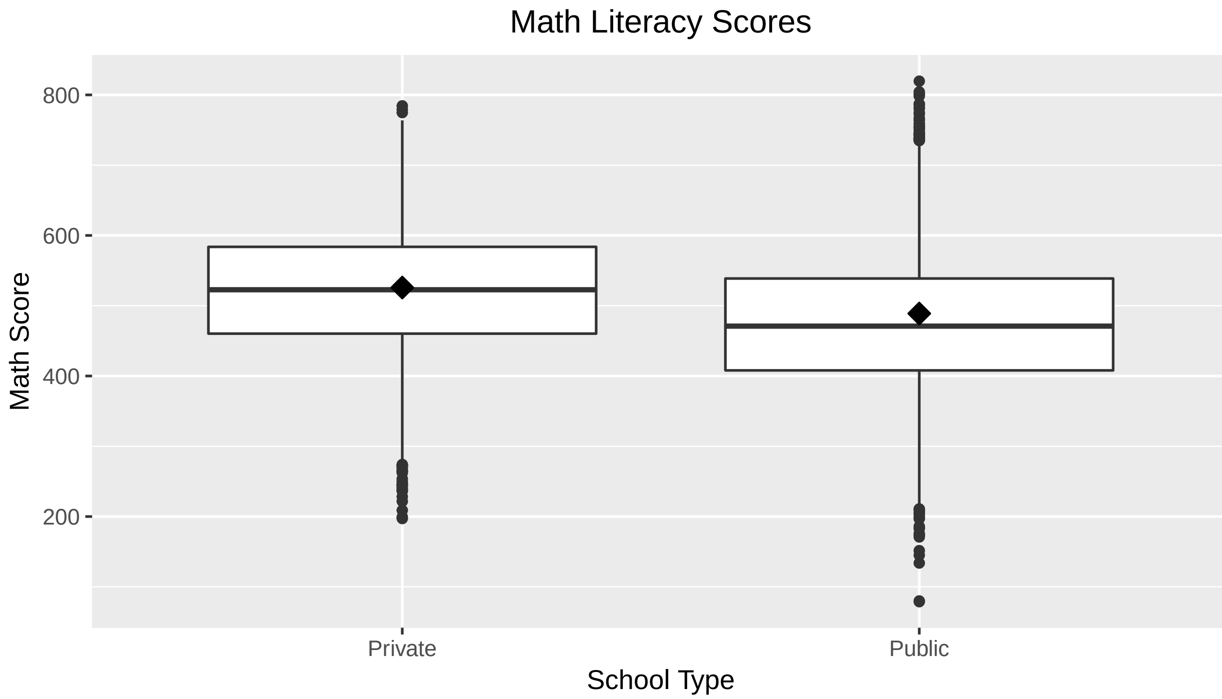

Figure 1: Breakdown of mathematics literacy scores based on school type. The diamond represents the weighted average.

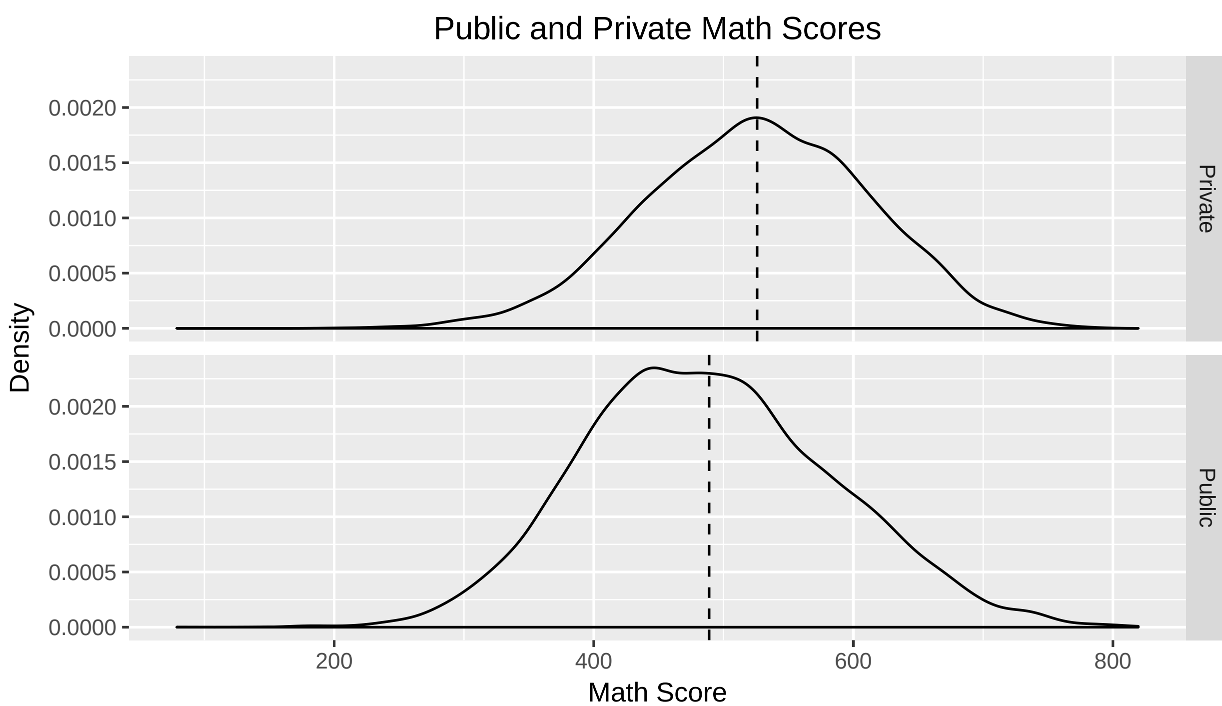

Figure 2: A density plot for public and private schools. The dashed line represents the weighted average.

Private School Variables of Importance

variable	relative_importance	scaled_importance	percentage
ST37Q05	52953232	1.0000000	0.2845634
ST42Q02	22098394	0.4173191	0.1187537
ST37Q06	17623804	0.3328183	0.0947079
ST43Q06	14371999	0.2714093	0.0772331
ST37Q07	13307177	0.2513006	0.0715109
ST37Q04	8040667	0.1518447	0.0432094

Table 1: The table consists of the variables ranked by their importance based on the best model from the GBM for private schools. ST37 questions are relating to self-efficacy. ST42 questions are relating to math anxiety. ST43 questions focus on the student's perception of their level of control to do well in mathematics.

Public School Variables of Importance

variable	relative_importance	scaled_importance	percentage
ST37Q05	88113232	1.0000000	0.2597299
ST37Q04	59034352	0.6699828	0.1740145
ST42Q02	33932372	0.3850996	0.1000219
ST43Q06	24762620	0.2810318	0.0729923
ST37Q06	16903930	0.1918433	0.0498274
ST42Q04	13466584	0.1528327	0.0396952

Table 2: The table is also derived from the best model produced by a GBM. This time, the model was trained using the public school data instead. Note the similarities and differences between the two tables.

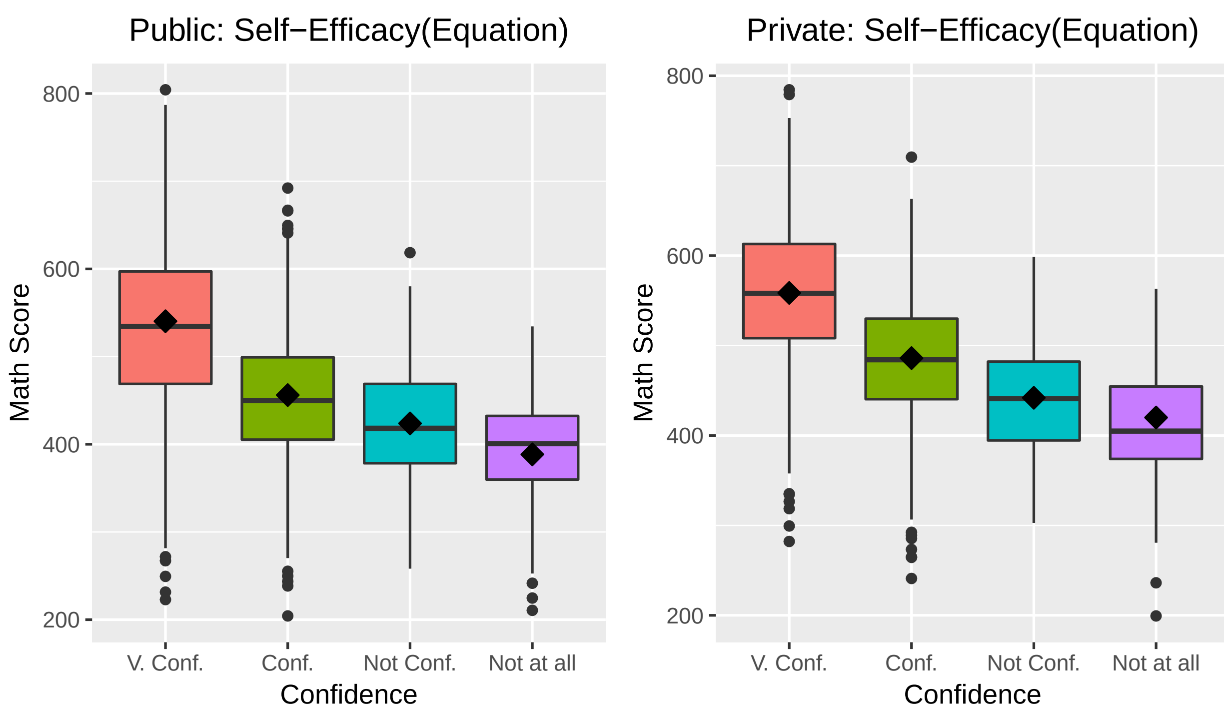

Figure 3: The two plots show a comparison between public and private school scores based on a single survey question: "How confident are you about: Solving an equation like 3x+5= 17". The choices were "Very

Confident", "Confident", "Not Confident", or "Not at all Confident". The diamonds represent the weighted average.

Public: Confidence in Solving Equation

Self_Efficacy	percent
Very confident	0.4530841
Not very confident	0.1446729
Confident	0.3708411
Not at all confident	0.0314019

Private: Confidence in Solving Equation

Self_Efficacy	percent
Very confident	0.6070196
Confident	0.2934407
Not very confident	0.0834292
Not at all confident	0.0161105

Table 3: The two tables show the percentage of students for each answer to the self-efficacy question regarding their confidence in being able to solve equation "3x+5= 17".

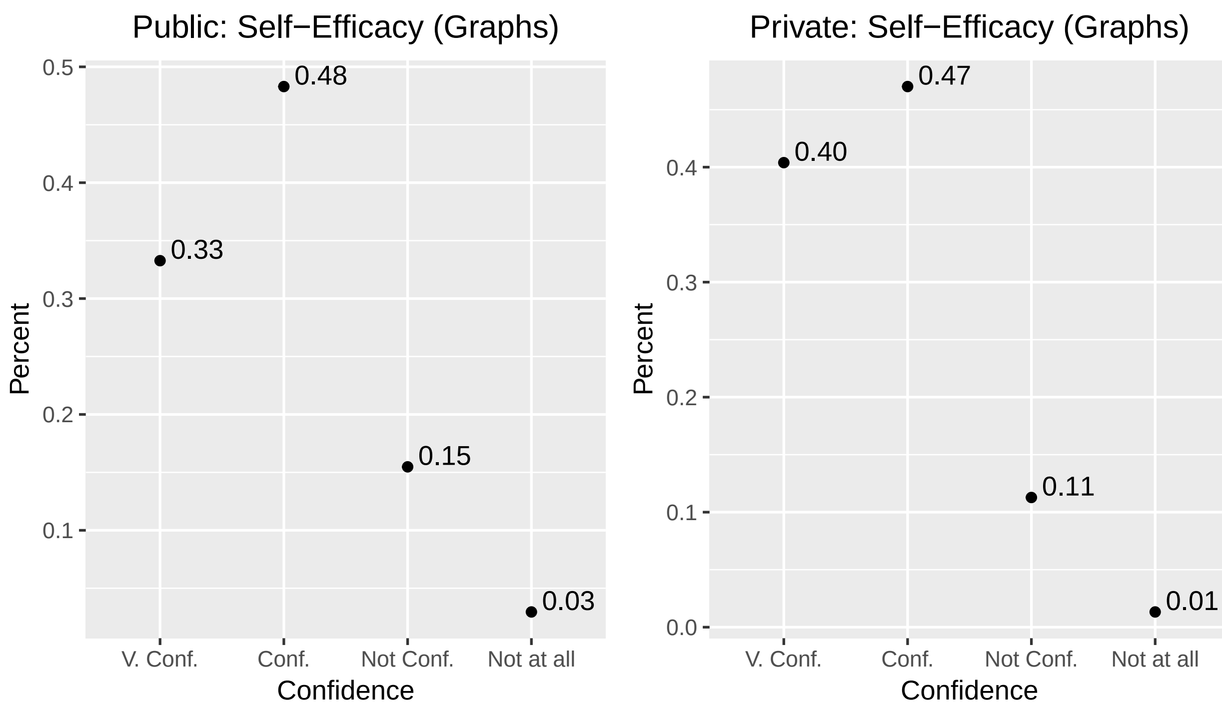

Figure 4: The two plots compare the percent of students who are confident in "Understanding graphs presented in newspapers".

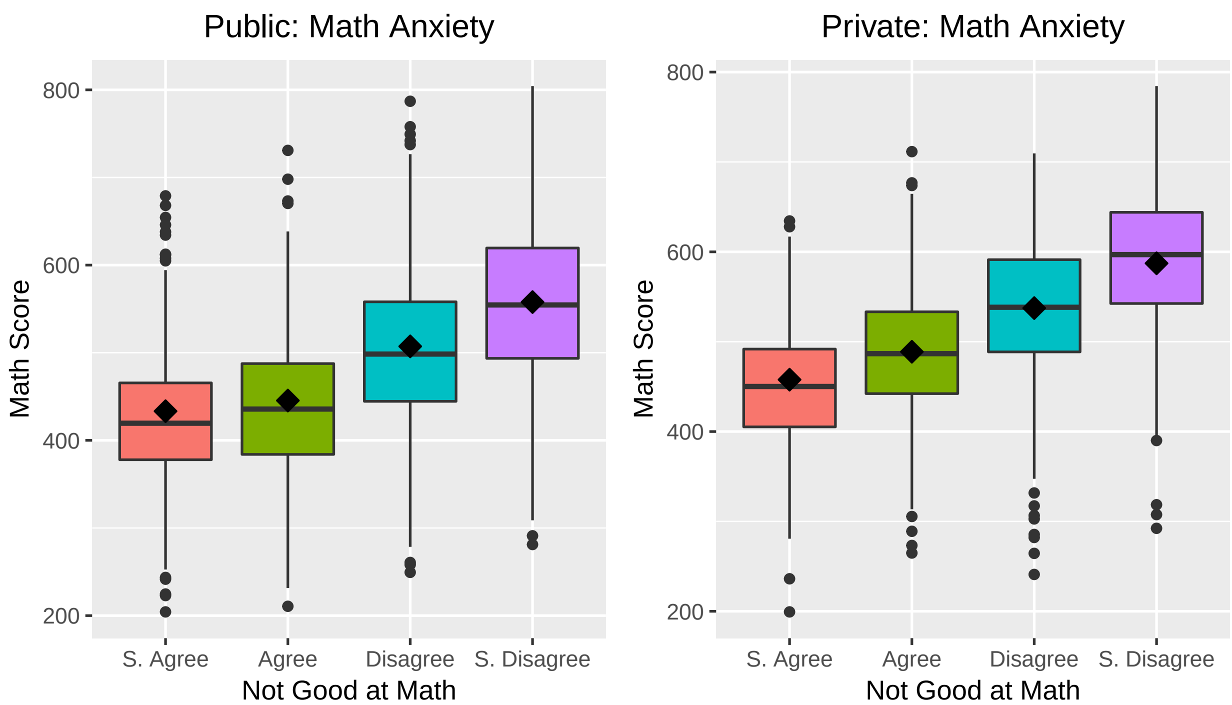

Figure 5: The plots above show the breakdown of math literacy levels for public and private school students based on the student's agreement with the following statement: "I am just not good at mathematics." The diamonds represent the weighted average.

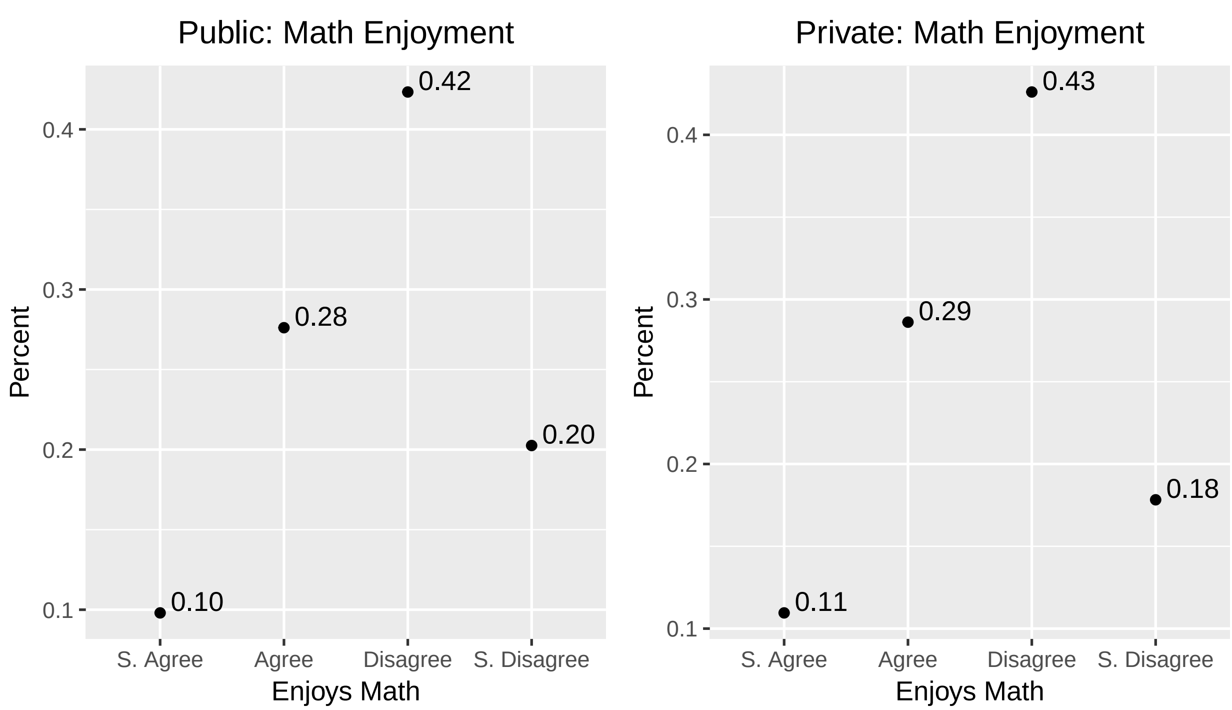

Figure 6: These plots look at the student's response to the following statement: "I do mathematics because I enjoy it."

Discussion

The comparative study started by analyzing the mathematical literacy scores for private and public schools. Australia's national average for math literacy was 504. Public schools had an average math score of 489, while private schools had a substantially higher average score of 526. Figures 1 and 2 further show how the scores differ between public and private schools. We can see from the outliers in figure 1 that public schools may have a wider range of skill levels to deal with. The highest and lowest scores are from public schools, which could be explained by higher diversity or the simple fact that there are more students in public schools. Figure 2 similarly breaks down the scores for Australian public and private schools. Here we can see the math scores for the two groups follow a mostly normal distribution. It is quite clear that there is a difference among these two groups in terms of math literacy, so it is worth exploring further if student disposition can help explain this difference.

Due to the large number of survey questions regarding student disposition, I chose to use supervised machine learning to determine which questions had the highest relevance in relation to student math scores. Tables 1 and 2 show the top 6 student disposition questions for private and public schools, respectively. In both tables, the variable of the most importance was ST37Q05. The gradient boosting machine rated this feature with a much higher relative importance than any other variable for either of the models. This question asked students how confident they were to solve the equation $3x + 5 = 17$. Questions of this nature deal with the part of a person's psychological disposition known as self-efficacy. Self-efficacy − belief in oneself − made up at least 3 of the top 6 variables in terms of importance for public and private schools. The similarities between the two tables continue with 5 variables appearing twice. This is encouraging, because it allows us to compare these similar variables and the effects they have on math literacy.

Since the self-efficacy question on confidence of solving an equation had the highest level of importance for both models, it is a logical place to start in comparing student dispositions and their effects between Australia's public and private schools. Figure 3 shows the math literacy breakdown for public and private schools for each of the possible answers to the survey question. While public school scores are slightly lower, the two plots look similar. Students who are very confident in their belief to solve such a problem score quite well regardless of which type of school they are from. And on the other end, if a student says he or she is not at all confident in solving such a problem then his or her score is quite low. This plot shows a correlation between self-efficacy and math scores, but is missing a key piece of the puzzle. Table 3 is a further analysis of student confidence in solving the equation. The two tables show a noticeable difference in self-efficacy between public and private schools. Over 60% of private school students are very confident in their ability to solve this problem, while only 45% of public school students feel this way. This helps explain why the public and private school math literacy scores are substantially different even though the plots in figure 3 looked so similar. Figure 4 continues to look at self-efficacy. Instead of assessing students' confidence levels in solving equations this question focused on understanding graphs presented in newspapers. Figure 4 plots the percent of students that fall into each of the four confidence levels. This plot also highlights the difference in self-efficacy between public and private school students with only 33% of public school students being very confident in comparison to 40% from private schools.

Apart from self-efficacy, math anxiety ranked fairly high as an importance variable for both models. The specific survey question was asking for a level of agreement with the statement, "I am just not good at mathematics". The results in figure 5 show math anxiety as another psychological disposition that effects students' math scores from public and private schools in a similar manner.

The final figure, figure 6, focuses on the student's enjoyment level of mathematics. Note, this was not one of the more important variables for either model. Only about 10% of students do mathematics because they enjoy it. In fact, a majority of students disagree or strongly disagree with this statement. The breakdown of math anxiety for public and private schools is nearly identical in this figure. This is an interesting contrast to self-efficacy where there was a noticeable contrast between the two groups.

Conclusion

The comparative study of student dispositions effect on mathematics literacy has yielded some intriguing results. Student disposition seems to have the same effect on students regardless of the school system. However, private school students appear to have higher beliefs in their ability to solve and understand math problems. The reason for this higher self-efficacy is unknown and could warrant further research of its own. One report on Australian PISA data stated, "students in the Catholic or independent school sectors bring with them an advantage from their socioeconomic background that is not as strongly characteristic of students in the government school sector. In previous cycles of PISA, the OECD has noted that the differences between public and private schools disappear once similar adjustments are made in most OECD countries" (Thomson, De Bortoli, and Buckley 2013). In other words, private school students perform better, because a higher percentage of them have a socioeconomic advantage compared to public school students and this explains the difference in test scores. Further research could be conducted to see if student disposition, specifically self-efficacy, is directly related to socioeconomic status. This statistical analysis has garnered a few fresh insights into what could be causing the gap in academic performance between public and private schools.

References

Bache, Stefan Milton, and Hadley Wickham. 2014. *Magrittr: A Forward-Pipe Operator for R*. https://CRAN.R-project.org/package=magrittr.

Buckingham, Jennifer. 2000. *The Truth About Private Schools in Australia*. https://www.cis.org.au/app/uploads/2015/07/ia13.pdf.

Dowle, Matt, and Arun Srinivasan. 2017. *Data.table: Extension of 'Data.frame'*. https://CRAN.R-project.org/package=data.table.

Gabriel, Florence, Jason Signolet, and Martin Westwell. 2017. "A Machine Learning Approach to Investigating the Effects of Mathematics Dispositions on Mathematical Literacy." *International Journal of Research & Method in Education* 0 (0). Routledge: 1–22. doi:10.1080/1743727X.2017.1301916.

OECD. 2017. "About Pisa." *Oecd.org*. http://www.oecd.org/pisa/aboutpisa/.

R Core Team. 2017. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

team, The H2O.ai. 2017. *H2o: R Interface for H2o*. https://CRAN.R-project.org/package=h2o.

Thomson, Sue, Lisa De Bortoli, and Sarah Buckley. 2013. *PISA 2012: How Australia Measures up*. Australian Council for Educational Research.

Wickham, Hadley. 2009. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. http://ggplot2.org.

Zhu, Hao. 2017. *KableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.