Statistical analysis of Educational data of Tanzania

Consultancy for Mokoro - World Bank evaluation project

Lucas Sempé[[1]](#footnote-1)

Paul Clist[[2]](#footnote-2)

08 September 2020

# Introduction and empirical strategy

This report is part of an assessment commissioned by the World Bank (WB) Results in Education for All Children (REACH) for ‘Evaluating Results-Based Financing (RBF) in the Education Sector’ and executed by Mokoro and more specifically, the Tanzania country-level analysis which is part of the overall assessment.

This report focuses on the results of RBF answering the following evaluation questions:

* Was there evidence of improved results to which RBF contributed?
* To what degree, and in what ways?
* Has behaviour change occurred, where and to what extent? Have changes been sustained?

The analyses done here provide evidence supporting the broader assessment of the contribution of RBF – the EP4R in Tanzania – to results in two areas in which the EP4R incentivized changes towards: a) a more equitable distribution of teachers, measured by the DLI 4.2. (a ‘LGA level’ DLI); and b) learning outcomes, which in the EP4R was captured by the DLI 6.2 (a ‘national level’ DLI):

* DLI 4.2. LGAs meet annual target for schools achieving acceptable primary pupil-teacher ratios (PTR) – i.e. a PTR within a given range
* DLI 6.2. Meet annual target of improvement in average Kiswahili words per minute in 3R assessment among Standard 2 students, which is a subcomponent of the EGRA and relates to improvement in learning outcomes. The analyses focus on documenting links, if any, between RBF payments, and results on or related to these DLIs. The analysis is structured in three main sections, followed by an Appendix. The first section describes contextual and payment data. The second section focuses on descriptive and inferential analysis of data related to DLI 4.2., while the third section approaches DLI 6.2.

The analysis is structured in three main sessions, followed by an Appendix. The first section describes contextual and payment data. The second section focuses on descriptive and inferential analysis of data related to DLI 4.2, while the third section approaches DLI 6.2. There are 3 different sources of learning outcomes that will be analysed, namely, PSLE and SFNA.

# Contextual and payments dataset

The following data has been used in the analysis:

* lgacontrols: 89 LGA level variables ranging from population, poverty, government expenditure on education, to data on educational system. Data range from 2015-2019. Not all data have complete series.
* payments\_alldrs: 9 variables. Payments in dollars for all LGA DLIs between 2015 and 2019, disaggregated at council level (LGA). See Table 1 for total amounts paid each year, ranging from 9.6 millions in 2016 to 41.3 millions in 2019.
* payments\_alldlrs.c: DLI payments are aggregated at council and DLI (losing time dimension) to use in the analysis of DLI 4 - PTR (only data on 2015 and 2019, a before/after). Table 3 shows that 151 councils received payments over the period while 33 did not. Table 2 shows evolution of DLI 4.2. payments over time where year 2018 shows the highest council average: 33,880 (s.d. 65,192), where 102 councils received resources.
* The DLI 6.2 is not an LGA DLI i.e. RBF payments made for results on this DLI accrue at the national level. We approach the analysis of the potential association of RBF with improved learning outcomes at LGA level as follows:

– Considering learning outcomes are a desired impact of an educational system with a wide range of factors influencing its results, we use the aggregated value of all DLI resources received by LGAs for other DLIs, to assess its association to learning outcomes by LGA, whilst acknowledging that myriad other factors also matter.

– With regard to learning results, the sampling approach adopted in measuring results on the DLI 6.2 does not allow results to be representative at LGA level. While we do not have access to the EGRA data, we do have access to the Primary School Leaving Exam (PSLE) and the Standard Four National Assessment (SFNA).

Table 1: Total DLI Payment per year

| Year | DLI payments (US$) |
| --- | --- |
| 2016 | 9602486 |
| 2017 | 18067901 |
| 2018 | 30143431 |
| 2019 | 41321502 |

# RBF contribution to results on DLI 4.2 - Pupil teacher ratio

This section compares the PTR with EP4R payments made on the DLI 4.2; discusses whether there was an improvement in teacher deployment, and at what level/ in what senses deployment improved.

As DLR 4 payments are linked to reaching the acceptable cutoff on PTR, so there is a risk of endogeneity/reverse causality on the analysis. Due to limitations on the data collection design, we focus on association between PTR and payments rather than unveiling potential causality paths.

## Payments

DLI payments are aggregated at council and DLI (losing time dimension) to use in the analysis of DLI 4 - PTR (only data on 2015 and 2019, a before/after). Table 3 shows that 151 councils received payments over the period while 33 did not. Table 2 shows evolution of DLI 4.2. payments over time where year 2018 shows the highest council average: 33,880 (s.d. 65,192), where 102 councils received resources.

Table 2: DLI 4.2. payments per council

| Year | Average amount per council | Standard deviation amount per council | Councils receiving money for DLI 4.2 | Councils not receiving money for DLI 4.2 |
| --- | --- | --- | --- | --- |
| 2016 | 16,141 | 48,358 | 53 | 131 |
| 2017 | 11,207 | 29,187 | 51 | 133 |
| 2018 | 33,880 | 65,192 | 102 | 82 |
| 2019 | 20,676 | 47,083 | 58 | 127 |

Table 3: Councils receiving money for DLI 4.2

| Councils receiving money for DLI 4.2 | Councils not receiving money for DLI 4.2 |
| --- | --- |
| 151 | 33 |

Figure 1 shows the histogram of overall payments for DLI 4.2 across all years, aggregated by council, which shows a skewed distribution towards 0, while some exceptional LGAs received more than $200,000 over the period.

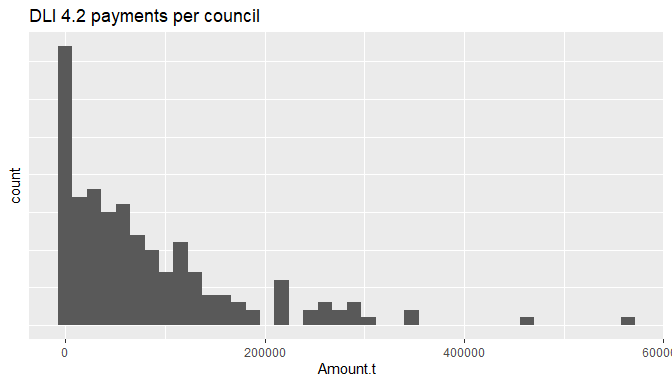


Figure 1: Histogram DLI 4.2. payments per council

## Descriptive analysis

We present in this section PTR descriptive statistics during the period 2015-2019 both analysed at council and school levels.

16,630 schools surveyed in 2015 were matched to 2019 schools, of which 173 did not have teacher counts and 137 of those did not have pupil counts. As a result 96% of the national total enrolment in 2015 is placed in one of the 16,457 matched schools. Songwe LGA was dropped due to a lack of identifiable 2015 schools. 17,352 schools are present in 2017 and 17,792 are found in 2019. Aggregated PTR (without dealing with outliers) has increased over time from 42.2 in 2015 to 52.5 in 2019, while the standard deviation, as a measure of dispersion, remains similar over time (see Table 4).

Table 4: PTR descriptive statistics

| year | Schools | Median | Average | Standard Deviation |
| --- | --- | --- | --- | --- |
| 2015 | 16,457 | 42.2 | 46.2 | 25.2 |
| 2017 | 17,352 | 47.3 | 50.7 | 24.0 |
| 2019 | 17,792 | 52.5 | 54.4 | 25.5 |

## Outliers treatment and selection of schools for analysis

### School level

Figure 2 shows school-level data for pupil-teacher ratios across years 2015, 2017 and 2019. Colors represent each region. Outliers shown when they are more than 1.5 units from the 25th and 75th percentile. It is possible to observe a wide interquartile range in all years, a significant number of outliers on the upper bound and also values close to 0. This suggests the need to address data quality systematically. Figure 3 shows school PTR distribution after outliers treatment.

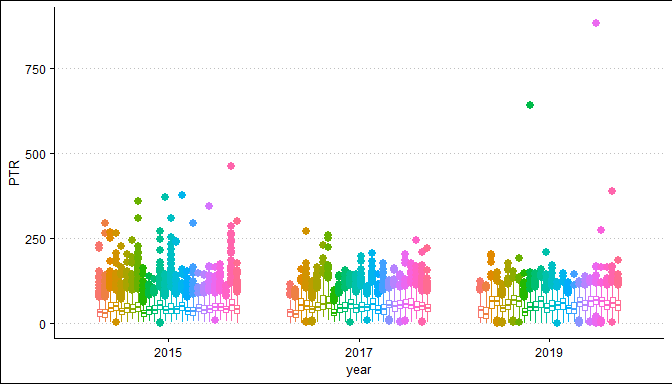


Figure 2: School level PTR 2015, 2017 and 2019

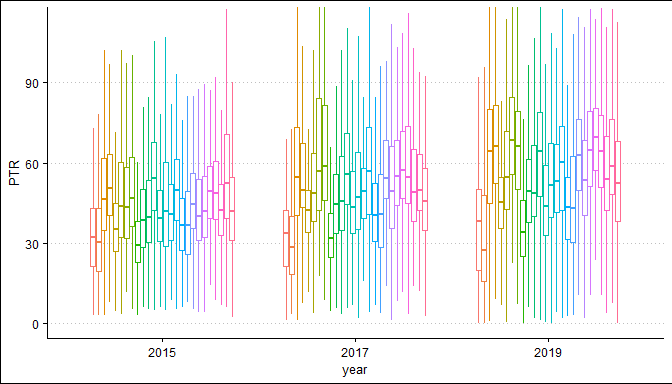


Figure 3: School level PTR 2015, 2017 and 2019 after outliers treatment

Among different alternatives of data quality analysis (see Appendix for analysis done through IQR and Mahalanobis distance), we chose to detect outliers based on z-scores distance from the mean. The criteria to exclude observations (i.e. school level PTR) rely on this outlier analysis, as well as a preference for longitudinal data. Any school falling under at least one of the criteria below was excluded from the sample, following this order:

• Negative numbers in pupils, teachers or PTR

• Plus/minus PTR 3 standard deviations (SD) assessed separately by year

• Difference within schools > 3 SD or < 3 SD

• In case of very low PTR, exclusion of PTR < - 2 SD (as -3 SD falls into the negative realm of values)

• Schools without 3 observations (2015, 2017 and 2019)

By doing this, we lose 8.3% of the observations. This will not affect statistical power (decrease from 51,786 to 47,778 observations), leaving a panel of 15,926 schools.

Table 5 displays summary statistics such as mean, median and standard deviation for PTR after discarding outliers across years, where all statistics shows a consistent growth over time. While mean values are not affected after removing schools, the median and standard deviations changed significantly. As the mean PTR is rising, the table also shows that while the number of students increased over time, the number of teachers did not increase in the same proportion.

Table 5: Summary PTR data without outliers

| year | Schools | Students | Teachers | Median | Average | Standard Deviation |
| --- | --- | --- | --- | --- | --- | --- |
| 2015 | 15,926 | 7,628,359 | 187,569 | 41.70 | 44.02 | 19.27 |
| 2017 | 15,926 | 8,595,837 | 182,762 | 47.40 | 50.00 | 20.91 |
| 2019 | 15,926 | 9,647,654 | 185,042 | 53.18 | 55.42 | 22.16 |

### Council level

School PTR across councils have deteriorated between 2015 and 2019. The overall proportion of schools with a PTR over 50 (which was the standard in 2015) grows from 31% in 2015, to 43% in 2017 and 56% in 2019. While the share dispersion measured by the standard deviation, also increases from 19% to 26% in the period (see Table 6).

Table 6: Share of schools with PTR greater than 50 - Descriptive statistics

| year.ptr | Median | Average | Standard Deviation |
| --- | --- | --- | --- |
| 2015 | 0.30 | 0.31 | 0.19 |
| 2017 | 0.47 | 0.43 | 0.24 |
| 2019 | 0.59 | 0.56 | 0.26 |

A simple measure of the surplus (or gap) of teachers per council was computed as , which gives an idea of the ability of the council to achieve equity in the distribution of teachers. A major caveat of this calculation relies on the fact that students cannot follow the same ideal distribution due to household location, school resources and population age structure – for instance, most systems aim to provide at least one teacher per grade. However, this allows to understand the magnitude of gaps to be covered and accounts for areas where schools are provided more generously.

Figure 4 shows the average PTR per council and the distribution of teacher surplus across years (calculated against a situation in which each teacher would be allocated to 50 students). Dot sizes represent total number of teachers.

• The best allocation of teachers will be across the horizontal 0 line, which represents no surplus of teachers in terms of actual PTR.

• The green area captures those councils that have PTR below 50, and diverse rates of efficiency on allocation – i.e. diverse sizes of teacher surplus.

• The red box represents councils both having a surplus and a current PTR over 50. These Councils have teachers that could be redeployed. However, this misallocation of teachers is not excessive – i.e. the surpluses are small in all concerned LGAs.

• Finally, the yellow area represents those councils that are above the PTR threshold and present a teacher’s deficit, which makes it impossible to reach an average of PTR below or equal to 50 with current resources. The number of LGAs in this area increases from 2015 to 2019. In 2019 allocation of teachers on that area presents greater inequality than previous years.

There is a strong negative linear correlation between PTR and teacher surplus (-.70, -.80 and -.87) across years (see Table 7), suggesting a direct relationship between PTR and teacher surplus. This also can be interpreted as teachers’ surplus being a strong constrain in terms of decreasing PTR over time, which becomes more stringent in 2019.

In all plots, both tails, in the negative and positive side, show different patterns. With a PTR of less than 50, found in the left upper quadrant (green), there is a larger variance between councils in terms of teacher surplus, which suggests different levels of efficiency and highly inefficient teacher allocation in a number of Councils with very large surpluses. However, the majority of councils on yellow and red areas could be considered efficient in terms of teacher allocation due to the lack (yellow) or small size (red) of teacher surplus.

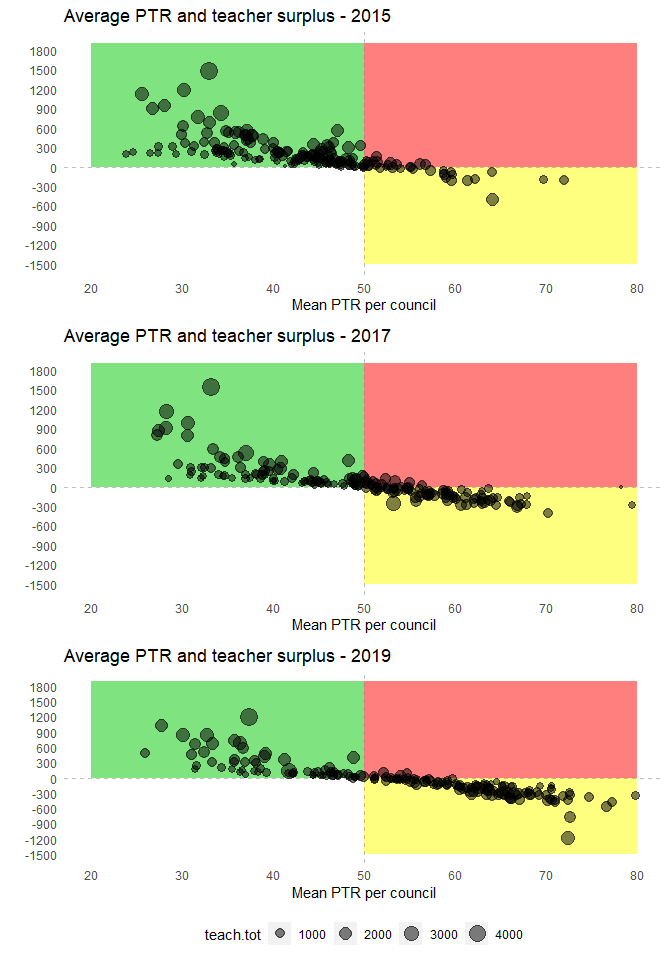


Figure 4: Average PTR and teacher surplus

Table 7: Correlation PTR and ideal teacher distribution by council

| year.ptr | Correlation |
| --- | --- |
| 2015 | -0.7 |
| 2017 | -0.8 |
| 2019 | -0.87 |

## Inferential analysis

Turning to the analysis of RBF – i.e. the link between payments made for results on the DLI 4.2, and PTR, at LGA level, we model multilevel mixed-effect regressions allowing for random intercepts and slopes vary at provinces. These models are known as conditional growth curves and have the advantage of allowing variance to occur within clusters and over time. All parameters are significant (p<.05). These models are similar to mixed-effects ANOVA to account for between- and within- clustered differences. The equation is written as:

where is the variable of interest, here the PTR, of council at year, and are the fixed intercept and slopes respectively, are the random intercept for Region, and is the residual. Finally, represents a set of predictors, our interest here being the amount of money received for the DLI 4.2 and , a dummy variable identifying districts with a PTR > 50 in 2015.

While the interaction between and is of our main interest as predicting changes on PTR, we perform our analysis starting by studying each term separately and adjusting for each other in Models (1)-(3). Models (1), (3) and (4) show a negative association between DLI amount over time, as expected due to the endogenous relationship between those variables. Model (4) does not show significant interaction between councils with high PTR in 2015 (over 50) and the total amount received, which means there are no relevant differences between the two groups and DLI amount on PTR.

##   
## ==========================================================================================  
## Dependent variable:   
## -------------------------------------------------------------------  
## PTR   
## (1) (2) (3) (4)   
## ------------------------------------------------------------------------------------------  
## Amount.t.1000 -0.023\* -0.028\*\* -0.029\*\*   
## (-0.047, 0.001) (-0.051, -0.005) (-0.056, -0.001)  
##   
## year.ptr2019 4.792\*\*\* 5.061\*\*\* 5.053\*\*\*   
## (3.054, 6.531) (3.315, 6.807) (3.302, 6.805)   
##   
## Amount.t.1000:high.ptr 0.004   
## (-0.047, 0.054)   
##   
## high.ptr 9.808\*\*\* 9.545\*\*\* 9.878\*\*\* 9.817\*\*\*   
## (7.483, 12.132) (7.270, 11.820) (7.600, 12.155) (7.377, 12.256)   
##   
## Constant 49.202\*\*\* 48.416\*\*\* 48.647\*\*\* 48.666\*\*\*   
## (46.292, 52.113) (45.366, 51.466) (45.675, 51.619) (45.682, 51.650)  
##   
## ------------------------------------------------------------------------------------------  
## Observations 366 366 366 366   
## Log Likelihood -1,342.346 -1,326.954 -1,327.759 -1,330.485   
## Akaike Inf. Crit. 2,698.692 2,667.909 2,671.517 2,678.970   
## Bayesian Inf. Crit. 2,726.010 2,695.227 2,702.738 2,714.094   
## ==========================================================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

We observe a statistically significant negative association of councils with higher PTR in 2015 (high.ptr) with PTR council differences over time (difference.council.ptr), where parameters are between -2.1 and -2.5 (95% CI -4.1, -.5). This represents a diminution on average of 2 pupils per teacher on councils with higher PTR in comparison to the others, controlling for all other variables. Descriptive statistics of the database is found in the Appendix.

##   
## ==========================================================================================  
## Dependent variable:   
## -------------------------------------------------------------------  
## difference.council.ptr   
## (5) (6) (7) (8)   
## ------------------------------------------------------------------------------------------  
## Amount.t.1000 -0.029\*\*\* -0.028\*\*\* -0.038\*\*\*   
## (-0.047, -0.010) (-0.047, -0.009) (-0.061, -0.015)  
##   
## year.ptr2019 -1.740 -1.469 -1.220   
## (-4.078, 0.598) (-3.827, 0.888) (-2.852, 0.412)   
##   
## Amount.t.1000:high.ptr 0.026   
## (-0.018, 0.070)   
##   
## high.ptr -2.163\*\* -2.416\*\*\* -2.165\*\* -2.474\*\*   
## (-3.861, -0.465) (-4.141, -0.690) (-3.863, -0.468) (-4.391, -0.556)  
##   
## Constant 6.425\*\*\* 7.106\*\*\* 7.335\*\*\* 7.250\*\*\*   
## (5.418, 7.431) (5.264, 8.948) (5.559, 9.111) (6.036, 8.465)   
##   
## ------------------------------------------------------------------------------------------  
## Observations 366 366 366 366   
## Log Likelihood -1,250.946 -1,249.274 -1,249.097 -1,259.729   
## Akaike Inf. Crit. 2,515.891 2,512.548 2,514.193 2,537.458   
## Bayesian Inf. Crit. 2,543.210 2,539.866 2,545.414 2,572.582   
## ==========================================================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

### Additional analysis

We focus on the analysis of possible linkages between good/bad teacher management (our current measure is LGA-level average in PTR) and primary completion (measured by LGA-level survival rates).

Positive impacts of pupil-teacher ratio are linked to a variety of school outcomes such as primary completion rates. As class sizes get larger, learning opportunities get smaller which drives to higher odds of grade repetition. It is known that grade repetition are significant predictors of schooling dropout. All models show an expected negative association between districts with PTR over 50 in 2015 and average survival rates. The association is statistically significant in all cases (p.<.05). There is no evidence that DLI money is associated to survival rates after controlling for differences between districts, as shown in models (9)-(12).

##   
## =================================================================  
## Dependent variable:   
## -------------------------------------------  
## surv19   
## (9) (10) (11) (12)   
## -----------------------------------------------------------------  
## high.ptr -0.079\*\*\* -0.079\*\*\* -0.081\*\*\* -0.029\*\*   
## (0.013) (0.013) (0.013) (0.014)   
##   
## Amount.t 0.000 -0.000 -0.000   
## (0.000) (0.000) (0.000)   
##   
## PTR -0.000 -0.000   
## (0.00000) (0.00000)   
##   
## year.ptr2019 0.000   
## (0.00000)   
##   
## std.tot 0.000   
## (0.000)   
##   
## teach.tot 0.000   
## (0.000)   
##   
## surv17 0.309\*\*\*   
## (0.043)   
##   
## povertyregion -0.300\*\*\*   
## (0.076)   
##   
## neverenrol\_uw2015\_pct -0.00004   
## (0.001)   
##   
## Constant 0.836\*\*\* 0.842\*\*\* 0.844\*\*\* 0.710\*\*\*   
## (0.007) (0.007) (0.007) (0.045)   
##   
## -----------------------------------------------------------------  
## Observations 366 366 366 306   
## Log Likelihood 3,034.301 2,964.712 2,305.021 1,902.407   
## Akaike Inf. Crit. -6,060.602 -5,919.425 -4,598.042 -3,780.814  
## Bayesian Inf. Crit. -6,044.991 -5,899.911 -4,574.626 -3,736.131  
## =================================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

# RBF contribution to learning outcomes (related to DLIs 6)

We now turn to exploring the association between RBF and learning outcomes, which is the rationale for the DLI 6.2. The analysis assesses the link between total amounts of RBF funding received for LGA DLIs, and learning outcomes as provided by PSLE and SFNA scores, for each LGA. It makes sense to consider the total RBF payments made to LGAs for this analysis for two reasons: 1) the DLI 6.2 is not an LGA DLI, so there were no payments associated with results on that DLI, going to LGAs; 2) the DLI 6 could be considered as a remote impact of RBF, with RBF contributing for only a small part in learning gain/loss, hence the more factors can be taken into account in the analysis (here, money from all LGA DLIs rather than one or another), the better.

The main findings are that there is little link between PTR and learning levels; and between RBF payments and learning levels.

## National exams

First we focus on national exams: Primary School Leaving Exams (PSLE) and Standard Four National Assessment (SFNA). Table 8 presents the number of schools for both exams across 2014-2019. SFNA 2014 is unavailable.

Table 8: Observations per year/exam

| year | PSLE | SFNA |
| --- | --- | --- |
| 2014 | 15,359 | N/A |
| 2015 | 15,585 | 16,657 |
| 2016 | 15,461 | 16,901 |
| 2017 | 15,679 | 17,163 |
| 2018 | 16,826 | 17,361 |
| 2019 | 17,047 | 17,560 |

### PSLE

We follow the same approach to detect outliers than before, excluding observations with more or less than 3 zscores from mean and differences within schools on time larger than 3SD. Figure 5 shows the relatively small proportion of excluded observations.

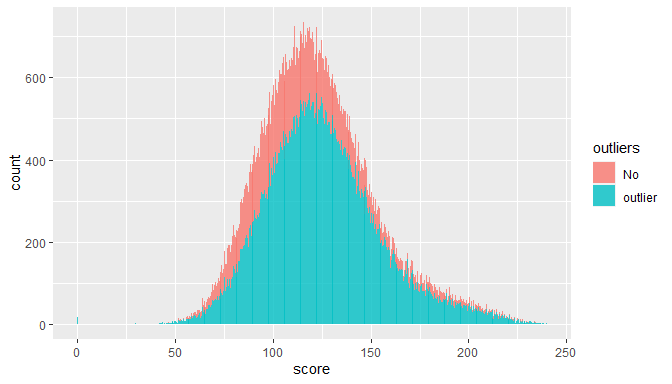


Figure 5: Histogram PSLE school-level scores with outliers identified

Figure 6 shows the association over time between LGA-level total RBF payments (in US$) and school-level PSLE scores for each year for which LGAs received funding (2016 to 2019). There is no clear slope which would suggest an association between both variables. This is also assessed through a mixed-effects regression model allowing for random intercepts for councils and random slopes for years, where we fail to reject the hypothesis of no association between them. After subsampling to years 2017-2019 we also did not find an association as shown across models (13)-(16).

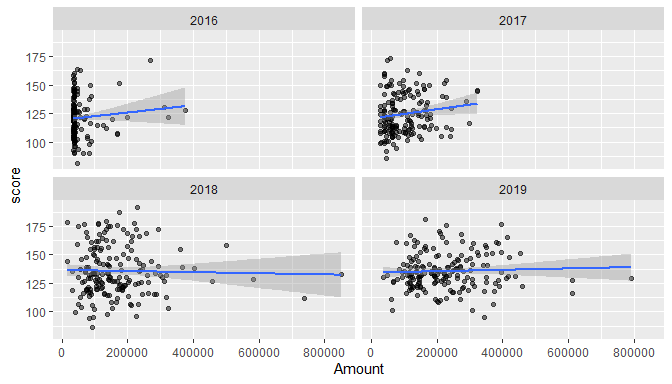


Figure 6: LGA-level average PSLE scores by LGA-level total RBF payments received

##   
## ===================================================================  
## Dependent variable:   
## -----------------------------------------------  
## score   
## (13) (14) (15) (16)   
## -------------------------------------------------------------------  
## Amount 0.00001 -0.00000 0.00001 -0.00000   
## (0.00000) (0.00001) (0.00000) (0.00001)   
##   
## year2017 3.370\*\*\* 3.527\*\*   
## (0.812) (1.481)   
##   
## year2018 13.738\*\*\* 12.484\*\*\* 10.297\*\*\* 9.498\*\*\*   
## (1.134) (1.788) (1.025) (2.073)   
##   
## year2019 14.047\*\*\* 13.006\*\*\* 10.729\*\*\* 9.624\*\*\*   
## (1.227) (1.923) (1.136) (2.184)   
##   
## Amount:year2017 0.00000   
## (0.00002)   
##   
## Amount:year2018 0.00001 0.00001   
## (0.00002) (0.00002)   
##   
## Amount:year2019 0.00001 0.00001   
## (0.00001) (0.00001)   
##   
## Constant 119.939\*\*\* 120.500\*\*\* 123.872\*\*\* 124.545\*\*\*   
## (1.221) (1.402) (1.322) (1.745)   
##   
## -------------------------------------------------------------------  
## Observations 9,179 9,179 6,960 6,960   
## Log Likelihood -43,104.880 -43,135.610 -32,701.550 -32,722.310  
## Akaike Inf. Crit. 86,241.760 86,309.210 65,425.110 65,470.630   
## Bayesian Inf. Crit. 86,355.750 86,444.580 65,500.430 65,559.650   
## ===================================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Model (17) shows a positive significant association between differences in PSLE school scores (difference.school) in the period 2014 and 2019 and the aggregated value for DLI payments for the period 2014 to 2018. After controlling for average scores, results become not significant in Model (18). The same analysis was performed at council-level, without finding a significant association - see Models (19)-(20).

##   
## ======================================================  
## Dependent variable:   
## -----------------------------------------  
## difference.school   
## linear OLS   
## mixed-effects   
## (17) (18) (19) (20)   
## ------------------------------------------------------  
## Amount.sum 0.00002\*\* 0.00000 0.00001 0.00000   
## (0.00001) (0.00001) (0.00001) (0.00001)   
##   
## score 0.700\*\*\* 0.637\*\*\*   
## (0.007) (0.033)   
##   
## Constant 11.780\*\*\* -74.734\*\*\* 6.419\*\*\* -69.012\*\*\*  
## (1.472) (1.580) (1.802) (4.069)   
##   
## ------------------------------------------------------  
## Observations 17,345 17,345 332 332   
## R2 0.004 0.537   
## Adjusted R2 0.001 0.534   
## ======================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

### SFNA

We follow the same approach to detect outliers than before, excluding observations with more or less than 3 zscores from mean and differences within schools on time larger than 3SD. Figure 7 shows the relatively small proportion of excluded observations.

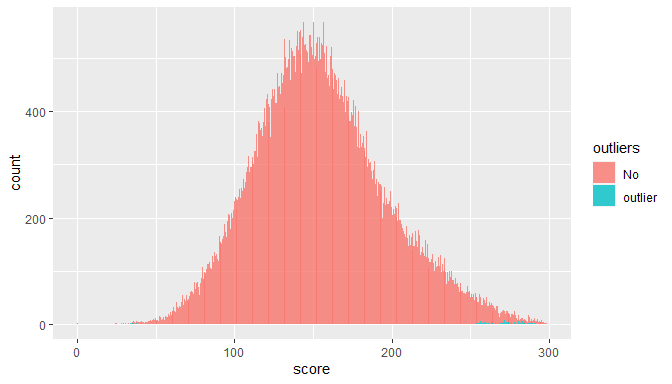


Figure 7: Histogram SFNA school-level scores with outliers identified

Figure 8 Figure 6 shows the association over time between LGA-level total RBF payments received and LGA-level SFNA scores, for each year for which LGAs received RBF payments. Like for the PSLE, there is no clear slope which would suggest an association between both variables. This is also assessed through a mixed-effects regression model allowing for random intercepts for councils and random slopes for years, where we fail to reject the hypothesis of no association between them. See models (21)-(23). Model (24) shows a significant coefficient due to the fact that the slope in year 2017 differs from the other years.

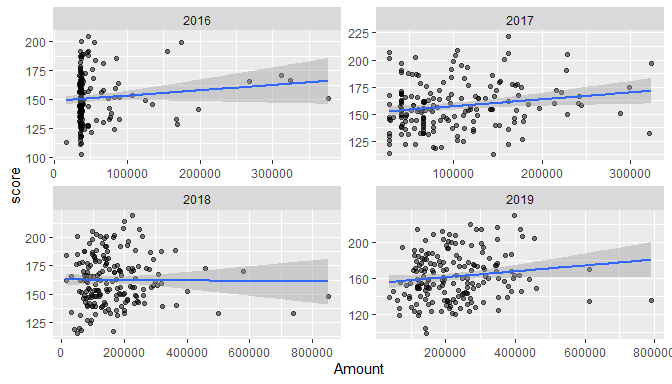


Figure 8: LGA-level average SFNA scores by LGA-level total RBF payments received

##   
## =======================================================================  
## Dependent variable:   
## ---------------------------------------------------  
## score   
## (21) (22) (23) (24)   
## -----------------------------------------------------------------------  
## Amount 0.00000 -0.00001 0.00000 0.00003\*   
## (0.00000) (0.00002) (0.00000) (0.00002)   
##   
## year2017 5.716\*\*\* 4.912\*\*\*   
## (0.903) (1.637)   
##   
## year2018 11.250\*\*\* 10.215\*\*\* 5.307\*\*\* 7.980\*\*\*   
## (1.369) (1.818) (1.127) (2.219)   
##   
## year2019 10.517\*\*\* 10.214\*\*\* 4.383\*\*\* 7.965\*\*\*   
## (1.854) (2.748) (1.692) (2.995)   
##   
## Amount:year2017 0.00001   
## (0.00002)   
##   
## Amount:year2018 0.00002 -0.00003   
## (0.00002) (0.00002)   
##   
## Amount:year2019 0.00001 -0.00003   
## (0.00002) (0.00002)   
##   
## Constant 151.080\*\*\* 151.796\*\*\* 156.494\*\*\* 154.178\*\*\*   
## (1.502) (1.727) (1.664) (2.237)   
##   
## -----------------------------------------------------------------------  
## Observations 68,704 68,704 51,891 51,891   
## Log Likelihood -340,466.700 -340,497.300 -257,664.400 -257,683.800  
## Akaike Inf. Crit. 680,965.400 681,032.600 515,350.800 515,393.600   
## Bayesian Inf. Crit. 681,111.700 681,206.300 515,448.200 515,508.800   
## =======================================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Models (25)-(28) show regression coefficients in models where we fail to reject the null hypothesis of an association between differences in SFNA school scores in the period 2015 and 2018 and DLI payments, both at school-levels and council-levels.

##   
## ======================================================  
## Dependent variable:   
## -----------------------------------------  
## difference.school   
## linear OLS   
## mixed-effects   
## (25) (26) (27) (28)   
## ------------------------------------------------------  
## Amount.sum 0.00001 -0.00001 0.00000 -0.00001   
## (0.00001) (0.00001) (0.00001) (0.00001)   
##   
## score 0.712\*\*\* 0.587\*\*\*   
## (0.008) (0.034)   
##   
## Constant 14.712\*\*\* -94.128\*\*\* 9.465\*\*\* -77.033\*\*\*  
## (2.010) (2.083) (2.324) (5.304)   
##   
## ------------------------------------------------------  
## Observations 18,530 18,530 339 339   
## R2 0.00000 0.469   
## Adjusted R2 -0.003 0.465   
## ======================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

# Conclusions

* Data quality:
  + data usually robust, with small number of unexpected outliers
* Inference:
* – Subsampling to avoid extreme measurement errors.
* – Association of reduction of PTR and increased amount received for the DLI 4.2 over time. No evidence of association of reduction of PTR with survival rates.
* – No evidence of association of LGA-level total payment amounts for combined local level DLIs and LGA-level learning scores.
  + Circumstantial evidence of associatiaton of scores differences and total amount DLI in PSLE.
* – Limitations: no counterfactual analysis, data quality problems (ghost teachers and students impossible to assess at regional level)

# Appendix

##   
## =====================================================================================  
## Statistic N Mean St. Dev. Min Pctl(25) Pctl(75) Max   
## -------------------------------------------------------------------------------------  
## PTR 47,583 49.815 21.337 2.379 34.900 62.600 130.400   
## Amount.t 47,584 91,800.770 102,899.700 0 13,677 128,032 564,774   
## Total Primary 47,583 542.163 387.525 6.000 298.000 672.000 7,365.000  
## Total Teachers 47,583 11.638 8.056 1.000 7.000 14.000 152.000   
## surv19 47,583 0.824 0.116 0.440 0.741 0.922 1.072   
## surv18 47,583 0.885 0.104 0.580 0.809 0.955 1.326   
## surv17 47,583 0.670 0.155 0.299 0.576 0.785 0.990   
## povertyregion 47,583 0.270 0.086 0.080 0.210 0.345 0.450   
## neverenrol\_uw2015\_pct 41,469 23.674 8.611 8.810 16.480 30.060 48.630   
## -------------------------------------------------------------------------------------  
##   
##   
## ===================================================================  
## Statistic N Mean St. Dev. Min Pctl(25) Pctl(75) Max   
## -------------------------------------------------------------------  
## score 169,594 121.598 28.478 29.867 101.644 138.111 236.730  
## -------------------------------------------------------------------  
##   
## ===================================================  
## Statistic N Mean St. Dev. Min Pctl(25) Pctl(75) Max  
## ===================================================

##   
## ===================================================================  
## Statistic N Mean St. Dev. Min Pctl(25) Pctl(75) Max   
## -------------------------------------------------------------------  
## score 103,192 151.836 41.595 24.250 122.963 177.175 298.500  
## -------------------------------------------------------------------

1. Univesidad Católica San Pablo [↑](#footnote-ref-1)
2. University of East Anglia [↑](#footnote-ref-2)