Analysis of factors affecting primary school students' performance

Abstract—This paper assesses the impact of gender, preschool and medium of the class in influencing academic results of primary students in government schools, with special attention given to performance in oral tests. The data used for the analysis was gathered by the Chennai chapter of Asha, an NGO working to improve the education of rural and urban schools in Tamil Nadu, India. They conducted oral and written assessments in about 90 primary government schools in Tamil Nadu in the subjects of Math, English, and Tamil. The data collected from these schools comprised of student's age, height, weight, parent's educational qualification and details about their preschool. Apart from these, details about the school were gathered from a website hosted and maintained by the National University of Educational Planning and Administration (NUEPA). This included school details such as the number of students, teachers and their qualifications, the facilities, etc. The results show that the medium of the class, the type of preschool attended and the gender of a student have a correlation with the performance while factors such as BMI, student-teacher ratio and school infrastructure quality have a small or negligible correlation with the performance.

Keywords—Preschool; gender; medium; performance

I. INTRODUCTION

In education, the decisions made by the teachers and educators may have the largest impact on a student's performance. Several times there are factors outside the control of a school that can affect how a student performs. Data analytics helps understand what these factors are and helps improve the quality of decision making of teachers and students. Data analytics creates new opportunities for educators to make these decisions early and provide maximum benefit to a larger number of people. It also helps the teachers to understand whether their intervention is effective or not. More importantly, it tells the decision-makers what changes will actually make a difference in a student's performance rather than something which seems good on paper but is not very effective.

There exist several papers on the topics regarding the effect of class size, school facilities, BMI and weight on a student's performance. In a review of that litera ture, Rivkin, Hanushek, and Kain (2005) analyse the relationship between class size, teacher quality, and student performance. [1] The results suggest that a low but statistically significant effect of class size on mathematics and reading scores exists and this effect declines as students' progress through school. The results also suggest that improvement in teacher quality has a positive effect than the reduction in class size. A paper written by Alswat, Al-Sheri, Aljuaid, Alzaidi, and Alasmari compares the BMI of children with their performance and the results suggest that there is no correlation between BMI and performance in the school.[2] In the existing works, reported the analysis based on a random distribution of students who were present in middle or high school. There is very little literature on the effect of various factors on primary education. Also, there is almost no work done on Indian students, including no work done comparing the performance of students in their native language versus performance in a foreign language.

The purpose of this paper is to investigate the following in primary students in both rural and urban Tamil Nadu in India-

- Effects of preschool education on a class 2 student's performance
- Effect of being in a native/foreign language medium class affects the performance of a student
- The role of gender on the performance of students

To answer the above research questions, we use the data collected by Asha Chennai and NUEPA. A very detailed data set is collected and maintained for nearly 90 schools over the last 3 years, which contains data at 3 levels: school, student and test paper and various related factors. We focused our study on students in the state Tamil Nadu. The schools are from all across Tamil Nadu including rural and urban schools giving us very varied results. We also used this data to generate other useful data, which helped us with the test. We performed several statistical tests on this data to try answering our research questions. The results suggest that Tamil section students tend to perform better in the oral tests as compared to English section students. It is also seen that attending more years of preschool has a positive effect on a student's performance. Finally, we also see that female students perform significantly better than male students in the oral tests.

II. LITERATURE REVIEW AND BACKGROUND

A. Related Works

In this section, we review the related works in analyzing the impact of certain factors on a learner's performance. In a paper written by Alswat, Al-Sheri, Aljuaid, Alzaidi and Alasmari, the relationship between body mass index and academic performance of students from the Kingdom of Saudi Arabia was studied. [2] Data were collected from 14 schools in the Taif city, Kingdom of Saudi Arabia. The BMI of each student was calculated and students were also categorized into two groups of high academic performance group (>90%) and low academic performance group (<90%) based on their GPA. It was seen that there was no correlation between academic performance and BMI.

Wenglinsky analyzed the relation between school spending and academic achievement. [3] The study was performed on a diverse range of students in class 8. It was seen that higher spending on resources such as teachers, which in turn reduced the student-teacher ratio and more qualified teachers, resulted in better academic achievement.

In a paper analyzing the impact of class size on student achievement, Borland starts by speaking about the factors which have previously hindered the papers. [4] He performs his tests, unhindered by these factors, and suggests that the

relationship between academic performance and class size is not only non-linear but also non-monotonous.

Though these papers are extremely informative, they fail to answer a few questions. All these papers work on a dataset which doesn't contain primary school students. There is almost no paper written about education in India and therefore fail to analyse the dynamic of teaching in native language vs foreign language. We try to answer these questions in our paper.

B. Asha Foundation

Asha was founded in the United States by a group of students interested in helping their home country, India. [5] The organization grew and spread to Canada, Singapore, and of course India, having nearly 60 chapters. Asha is also registered as a public charitable trust in India.

Several projects in Chennai and vicinity have been funded by Asha chapters. They are actively executing about 10 projects which together support more than 80 schools, and provide scholarships for over 100 students. They have been raising funds for these from individual donations, donations from corporations and organizations as well as from other Asha chapters. Asha collects relevant data such as height, weight, etc. which helped us with this study. Asha provides its own assessment several times a year to check the performance of the students. We used these test scores to perform our analyses for this paper. The Asha volunteers collect the data and update the database themselves.

C. DISE Data

To improve the educational statistics scenario in India, a pilot project was initiated at the National Institute of Educational Planning and Administration (NIEPA) in 1995. This project was provided with financial assistance from ICEF. The goal of the project was to examine issues related to the identification of data needs, processes and procedures for data collection, developing a framework for data flows and computerization, and facilitating the use of educational indicators in planning, management, monitoring, and evaluation. The first database, named as District Information System for Education (DISE), was released in 1995. The data collection process is expanded to include more schools over the years. A total of 604 districts across 35 States and UTs have been covered under DISE during 2005-06. [6]

The data provided by DISE includes key data on elementary education in terms of the number of schools, enrolment, and teachers; examination results for the previous academic session for the terminal classes at primary and upper primary levels of education; sex-wise enrolment of children with disabilities at primary and upper primary levels; enrolment by the medium of instruction and by school category, performance indicators in terms of school category; enrolment distribution: total, scheduled castes and scheduled tribes, percentage female enrolment; classrooms; dropout rate, quality indicators according to the category of schools, teacher-pupil ratio; blackboard and building; students classroom ratio, etc..

We used DISE data in our analysis. The data at the school level is extracted from a website maintained by the National University of Educational Planning and Administration (NUEPA). We used the data from District Information System for Education (DISE) database. The DISE data were preprocessed for our analysis.

III. DESIGN METHODOLOGY

The first few years of a child's life are supposed to be the most important years in the cognitive, motor, social and emotional development of a person's life cycle. To measure this, ASER Centre, which is a part of an NGO called Pratham, has been organizing tests using a common metric to ensure that students reach a basic level which is required. [7] This has proved to be an effective tool in seeing whether the several schemes implemented by the government are actually having an effect on a child's performance. According to ASER, a child's development can be measured by checking the competency in four domains: early language acquisition, early numeracy skills, cognitive abilities, and social and emotional learning. ASER designs its papers in such a way that it tests these skills and hence provides us with a good way to measure oral skills. Asha Chennai uses the test prepared by ASER to test the student's skills in both the oral tests as well as the written tests.

Along with these test scores, a very detailed data set is collected and maintained for nearly 90 schools over the last 3 years by Asha Chennai which has data at 3 levels: school, student and test paper. The data is collected by Asha volunteers in each school. We worked with Asha Chennai chapter which is present in the state Tamil Nadu to get data from all over the state, including both rural and urban schools. This gives us a very distributed dataset and hence we see varied results.

The dataset is divided into 3 main levels: school, student and test level. Firstly, at the school level, there is data such as scheduled caste percentage, scheduled tribe percentage, girl percentage, number of teachers, number of students and RTE score which is a marker out of 10 on how good the facilities such as Ramp, Playground, Boundary Wall, Drinking Water, Boys Toilet, Girls Toilet, Library, Pupil-teacher ratio, Student-Classroom ratio and Teacher-Classroom Ratio are. This data is collected directly from the DISE data and included in the database. It also contains the details of every section in the school along with the medium followed, the teacher in charge, and the percentage of girls and the average performance of the class. This data is calculated using the student's gender information. Then, at the student level, the dataset provides information on various topics such as individual performance of students along with details about their parents' education, preschool details, height, weight, and DOB. It also tells his section in class and hence the details of their class.

Finally, at the test level, there are details of every question answered by a student along with the response of the student, details of the question paper and the level of the student in the oral scale. The data contain test scores spanning grades 1 through 5 for the years 2016, 2017 and 2018.

A. Oral Analysis

The scores are based on the performance of the students in the 3 fields - Reading, Arithmetic, and English. Each section is graded out of 10 marks and split equally based on the number of levels. Table 1 shows how the grading is done based on the level of the student in each of the subjects.

A student's score is added up and then the percentage is calculated based on the number of tests they have written. For example, if a student is at one digit no in arithmetic, word in reading and small letter in English his score is 1.66+5+3.33 = 9.99. The percentage is 9.99/30 * 100 = 33.33%. If a person writes only 2 tests, his score is calculated out of 20. We used an absolute grading to calculate the oral test scores. This lets us compare the performance of students across schools but it doesn't let us compare the performance across classes because the oral ability of a student improves with age.

TABLE I.   ORAL SCORE DISTRIBUTION

Arithmetic	Reading	English
Pre number - 0	Pre letter - 0	Pre letter - 0
One digit number - 1.66	Letter - 2.5	Capital letter - 1.66
Two digit number - 3.33	Word - 5	Small letter - 3.33
Addition - 5	Sentence - 7.5	Read words - 5
Subtraction - 6.66	Paragraph - 10	Understand words - 6.66
Multiplication - 8.33		Read sentence - 8.33
Division - 10		Understand sentence - 10

IV. RESULTS

A. Effect of preschool on class 2 performance

To see the impact of preschool on a student's performance, we checked and performed tests on the performance of class 2 students. We had a dataset of 657 students in class 2 out of which there were 467 balwadi students and 132 KG students. There were also 58 students who didn't attend any form of preschool. We made 2 groups, students who went to balwadis and students who went to KG. Balwadis are smaller education centers present in almost every district in India. They focus on activities which help in early development and very little on math and English. KG, on the other hand, starts teaching kids basic math, English and Tamil, in this case.

Mann Whitney U test was used to verify the null hypothesis. We used Mann-Whitney U test because the distribution of the data was not normal and the number of values in each group was different. The null hypothesis was that the mean scores of Balwadi and KG students are equal. We took a two-tailed significance level at 5%. From the data, it can be concluded that the performance of the students who attended KG is significantly better than the students who attended balwadis (U = 35214.00, p = 0.0000181232, Z = -4.287). We see that the performance of the students who attended KG is significantly better than students who went to balwadis and the null hypothesis is not true. This suggests that students attending more years of preschool perform better and preschool plays a major part in the performance of students at the primary level. On checking, it is seen that students from both balwadis and KG perform significantly better than students who haven't attended either. This indicates that preschool education is extremely important in any form.

B. Effect of being in a Tamil/English section

The average performance of every section was calculated and it was divided into two parts- Tamil and English sections. It is very common to see schools in India teaching in the language prevalent in the region, in this case, Tamil. We have a total of 479 class sections which were used for this test, of which 302 are Tamil and 177 are English. The average oral test scores of each of the classes were calculated.

Mann Whitney U test was performed to check the null hypothesis that the performance in English and Tamil sections is equal. It was used because the distribution of the performances was not normal and the groups have an unequal number of sections. We took it to be a two-tailed significance level at 5%. It was seen that the performance of the Tamil sections was significantly better than the performance of the English sections in 2016 (U = 8370.50, p = 0.016, Z = 2.410) and 2017 (U = 18838.50, p = 0.0012, Z = 3.245). The disparity between Tamil and English sections reduced in 2018 and the result was considered to be equal in 2018 (U = 28019, p = 0.3769, Z = 0.884). Though written and oral scores show a positive correlation, it is seen that Tamil students perform comparatively better than the English students in the oral tests. Finally, it is also seen that the average score of the students is also increasing over the years.

C. Role of gender on a student's performance

The oral scores of the students were divided into two groups based on the gender of the students. There were a total of 4155 students of which 2167 were girls and 1988 were boys. The individual oral test scores were used for this test.

Mann Whitney U test was performed on the scores since the distribution of the scores was not normal and the number of students weren't equal. The null hypothesis was that the performance of the female students was equal to the performance of male students. We took a two-tailed significance level at 5%. From this data, it can be concluded that the performance of girls was significantly better than the performance of boys (U = 2356098.00, p = 1.39302e-7, Z = 5.266). It was seen that the performance of girls was significantly better than the performance of boys in all the years the test was taken. The null hypothesis was not true. The difference between the mean of the scores of the girls and the boys reduced from 2017 to 2018. Since the difference between the scores of both genders is decreasing, it is also seen that the male students are slowly catching up to the female students.

V. DISCUSSION

From our tests, we can conclude that going for preschool definitely increases the chance of a student performing better in the primary years, students in Tamil medium schools tend to perform better than their English peers, and female students tend to perform better at the primary level. Also,

more years of preschool also have a positive effect on the performance of the kids as the students in KG have a better performance than the students who only went to a balwadi. Kindergarten is an integral part of a child's early education as it acts as a transition from home to a school environment. It helps in the development of a child's social, emotional and literacy skills.

It is seen that children with richer parents tend to attend KG while poorer kids attend balwadis because the government provides it for all children. It was also observed from the data that the number of students who went to balwadi greatly outweighs the students who went to kindergarten. This could be attributed to the fact that until recently there were no kindergartens in the government schools of Tamil Nadu which led to a lesser number of students attending them in the rural areas probably due to lack of government-run kindergartens and high fees of those run privately. Moreover, the majority of the students attend balwadis because of proximity to homes in villages and economic conditions. Since the performance of balwadi students are poorer, steps should be taken to improve it and bring it to the same level as KG. Most of the students will benefit from this and it will lead to better results at the primary level. Parents must also be better informed about the benefits of preschool education for kids.

A majority of the students in rural Tamil Nadu are taught in school in Tamil as they're used to communicating in it from a young age. In many cases, English is taught as a second language in these schools. Tamil medium is the more common medium because it is taught in government schools and access to these schools is very easy as they are present in almost every district in Tamil Nadu.

From our results, we see that students studying in Tamil medium sections perform significantly better than students studying in English medium. This might be because of the fact that a third of the oral assessment is in Tamil. Therefore, it is much easier for Tamil medium students while it is the second language for the English medium students. It can be contributed to the fact that the students are very comfortable with their language at home. Students should be encouraged to take languages like English seriously as well. This will improve their overall learning experience, gain more knowledge and perform better in the real world.

At the primary level, it is also seen that girls perform better than boys. This is a very surprising fact because girls traditionally do more work at home in rural households. This suggests that male students don't spend as much time studying as female students do. Students must be explained the importance of academics at a young age because it sets up a foundation for future years.

VI. CONCLUSION

This study was performed to check the various factors affecting a student's performance at the primary level. Though various studies have been done before, this study was different because of the unique dataset used. Majority of the students were from primary schools in rural Tamil Nadu in India and hence the factors affecting such students were distinctive. The tests and the data used in this paper permit us to draw a number of sharp conclusions about primary education and to provide clear answers for the questions raised in the Introduction.

(i) *Preschool education provides an important foundation for students which helps them in their primary years.* Preschool education, especially kindergarten, is seen to have a positive impact on a student's performance. Students who went to kindergarten are also seen to perform better than students from balwadis. Preschool must be made more accessible for people in rural India to provide a better education for students.

(ii) *Students studying in their native language tend to perform better than students studying in a foreign language.* Students who studied in Tamil perform better than their English section counterparts because they are more comfortable in the language. English should be given more importance in school since it will help many of the students later in life as well.

(iii) *Girls perform better than boys in oral tests at the primary level.* It can be concluded that girls take academics more seriously and therefore perform better in class. This is a surprising results considering that girls also do a lot of household work and have to balance this with their academics as well. Boys must be explained the importance of education at the primary level.

In closing, we note that there are only a handful of studies relating external factors to a student's performance. Further work is clearly warranted given the importance of the issue. In addition, while we try to advance the literature, there are several limitations to our study that we acknowledge. We only used the oral test scores to perform tests and this may not be the best method. We were missing several factors which could have been used to achieve better results. Data like parents' occupation, guidance, self-study, teacher's quality, socio economic status, peer performance, family size and work at home could've been great indicators of a student's performance and it might be helpful to collect such data later. With the present data, we were unable to develop a model which could predict a student's score satisfactorily. It is of utmost importance that we strive to improve this work and bring about change which could improve the conditions of education in rural India.

REFERENCES

[1] S. G. Rivkin, E. A. Hanushek, and J. F. Kain, "Teachers, schools, and academic achievement," Econometrica, Vol. 73, No. 2, pp. 417–458, March, 2005.

[2] K. A. Alswat, A. D. Al-shehri, T. A. Aljuaid, B. A. Alzaidi, and H. D. Alasmari, "The association between body mass index and academic performance," Saudi Med J 2017, vol. 38 (2), pp. 186-191.

[3] H. Wenglinsky, "How Money Matters: The Effect of School District Spending on Academic Achievement," Sociology of Education, vol. 70, no. 3, p. 221, 1997.

[4] M. V. Borland, R. M. Howsen, and M. W. Trawick, "An investigation of the effect of class size on student academic achievement," Education Economics, vol. 13, no. 1, pp. 73–83, 2005.

[5] "Asha Chennai," The Chennai chapter of Asha for Education. [Online]. Available: https://chennai.ashanet.org/about/. [Accessed: 2019].

[6] "School Report Cards," School Report Cards. [Online]. Available: http://www.schoolreportcards.in/. [Accessed: 2019].

[7] "ASER Centre," ASER centre. [Online]. Available: http://www.asercentre.org/. [Accessed: 2019].