**Literacy Analysis in Primary School Pupils – A Survey of Schools in Kuje Local Council in the Federal Capital Territory, Nigeria.**

A Project Report to Africa Data School Nairobi, Kenya

Submitted by Modupe Owolabi – Abuja, Nigeria

**Introduction**

The backdrop to this project was the result of a number of years working in an elementary school library in a rural community. In this school most of the pupils do not speak English outside of school hours, and even when they do, it is usually a broken form of English. The majority of the pupils were from a less privileged background, and were so disadvantaged economically that many did not even have regular story books at home. As such, several concepts were strange to them even on reading some of the classic stories like: “Cinderella”, “Thomas the Tank Engine”, “Cat in the Hat”, “Clifford the Big Red Dog”, and “The Gruffalo”. In spite of the fact that the stories held their attention – the vocabulary, the settings, the characters were all quite strange.

When Africa Data School asked the students to do some exploratory work and come up with a project that was both financially viable and socially relevant, the idea of collating data from several schools in the area, and running a series of analysis was born. Whether or not there would be some economic advantage remained to be seen. The Proof of Concept (please find attached, Appendix A) was submitted, and approved and work began in earnest in December 2022. A series of schools were outlined for visits in the area around the community where the initial school is located. A target group of 1000 pupils was proposed. A written letter of introduction was obtained from Africa Data School, and in time over a period of three months twenty-three schools were visited. Permission was sought from the Proprietors (School owners) in most cases, and pupils from the classes where it was expected they could read on their own were selected. Sadly, not all the pupils could read fluently even at the fourth and fifth years of basic education (primary school, grades 4 and 5 – ages 9-12.)

The study was conducted primarily to carry out a descriptive analysis of possible language background effects on English comprehension in elementary education among pupils from a rural background. The insight gained was invaluable. A survey was carried out among over one thousand respondents. The forms distributed required basic data collection and selecting the correct option from multiple choice answers to questions on a simple comprehension passage in English. Although the majority of the pupils understood fairly basic English as shown by the test scores, the sheer number of languages spoken was overwhelming and indicated a wide series of combinations.

The process took longer than anticipated. Over one thousand forms were handed personally to pupils and/or teachers in the various schools. A sample of the Form is shown in Appendix B. The pupils were closely monitored, and in some cases, the questions had to be read to them and the options explained to them. For the most part the pupils were eager and willing to participate. (Personal Data was discouraged in a few schools. The personal data was not used at all in the data analysis. It was really to be able to ensure that the forms were individually filled.)

**Main Points:**  The data entry was carried out over the period of time the data was being collected. However, it was not very well regulated. As a result the pre-processing stage took a very long time. Although much of the cleaning was done on Microsoft Excel before it was converted to a pandas data frame, it really made it clear that the process was more tedious than it should have been. Given this understanding, a future collation could be done from the forms into a spreadsheet via Google Forms. It would make the format cleaner and more standardized.

The general plan was to run this analysis in such a way so as to answer the following questions:

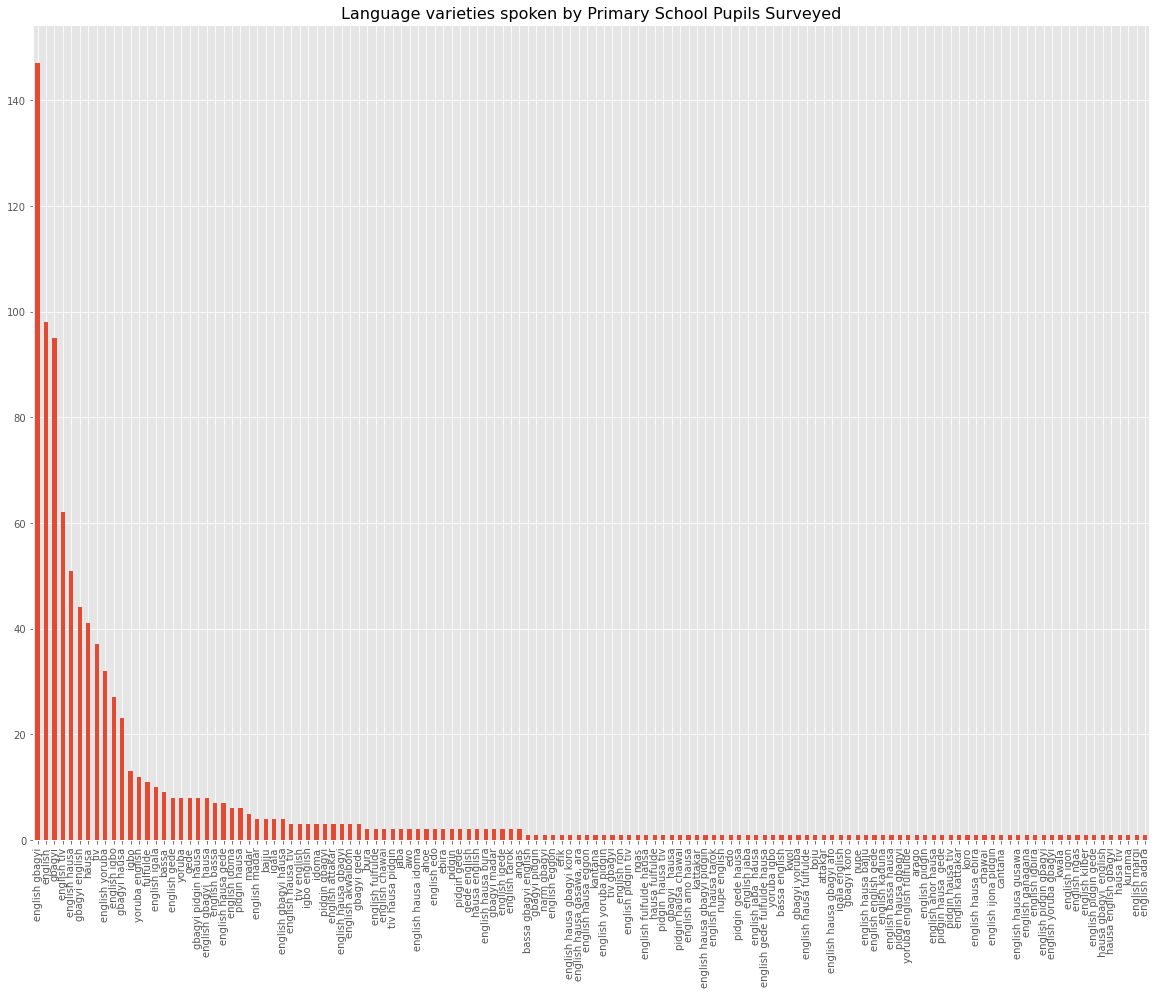
1. What number of pupils speak more than one language?
2. What language combination occurs most frequently?
3. What's the next highest frequency combination.?
4. Does the number of languages spoken have a bearing on pupil's understanding of English?
5. Does the language background affect English comprehension levels? (Question # iv. being more specific than Question #v)
6. Is it possible to map the scores of the comprehension with the languages spoken?

**Exploratory Focus**

Knowing that the machine learning aspect was a big challenge, it seemed easier to present data visualizations as the key component of displaying insights to the data collected. Generally, the statistics shown were just to highlight some of the significant observations. Data visualization as a means of exploratory data analysis is an aspect that really serves to explain insights gleaned from the data in a colorful and interesting way, in a way that explains the values and information being portrayed. Here are the charts to answer the questions listed above:

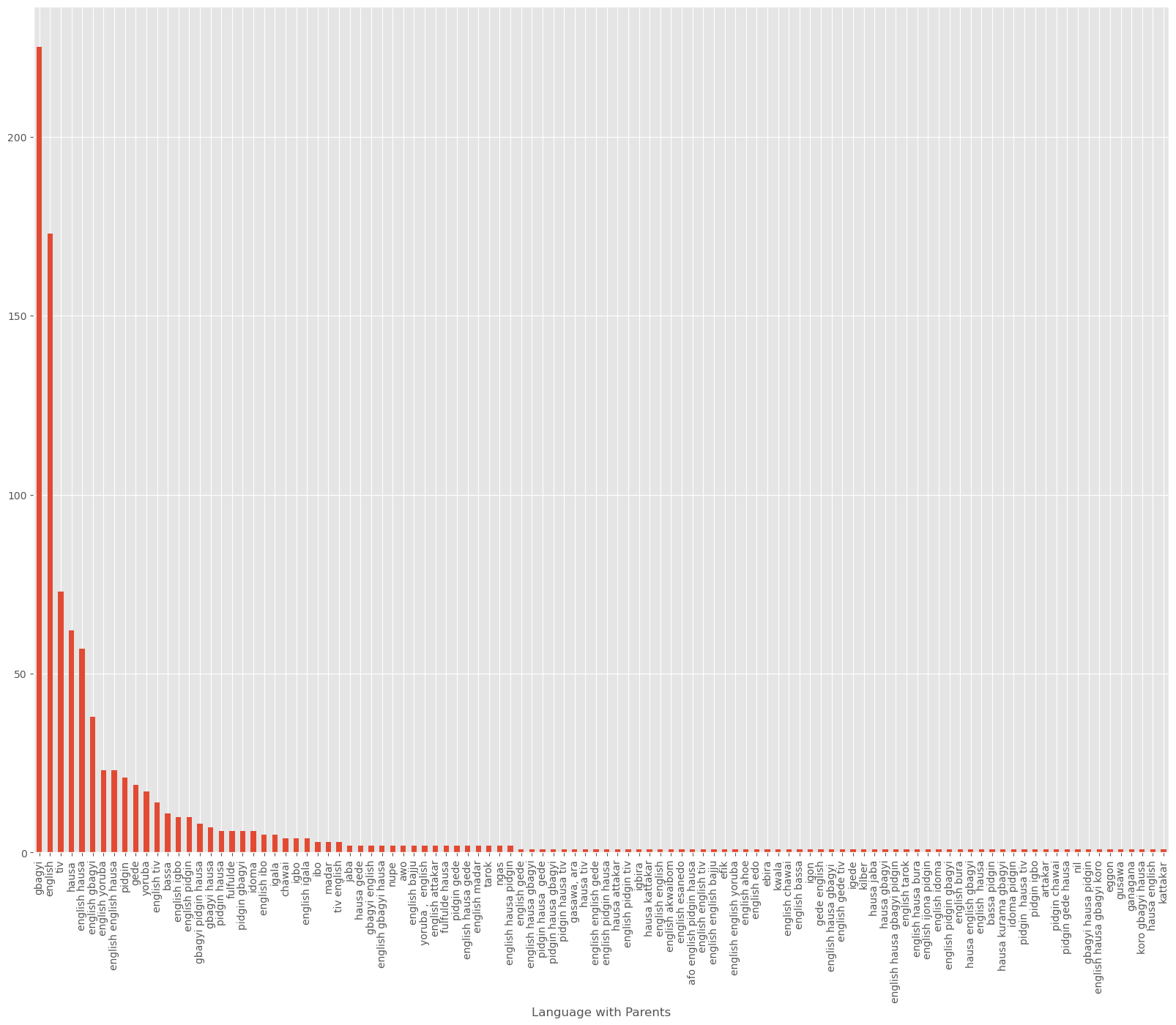
**Language charts…**

Fig i. Language Combinations spoken by Pupils



1. Pupils who spoke only English from this chart are about 95. This implies that 805 pupils spoke more than one language. Some as many as four different languages.
2. From this chart, the most prevalent language combination was English and Gbagyi (about 195 out of 900), which is not surprising given that the indigenous language of the FCT is Gbagyi. ( The first and the sixth bars are the same language combination)
3. The combination with the next highest frequency is English and Tiv. (The fourth bar)

Fig ii. Languages Spoken with Parents



In the course of the findings, it became clear that the number of languages spoken was quite large (over 70 different langauges). Also the combination of languages was quite diverse.

**Fig iii. PairPlot Analysis**



The pairplot shows a definite correlation, but there is not enough to indicate a causal relationship so as to be able to answer Questions iv, v, and vi. The results of this analysis indicates that there is some correlation between the languages spoken and the level of comprehension.

**Recommendations and Conclusion**

In the course of carrying out the analysis, it was discovered that there is no strong causal argument for **if** and **how** a non-English language background would affect the comprehension. The pair plots done show some correlation, which in some cases is explicable. But there was no strong evidence for a **causal** relationship. This could not be readily shown by the data collected, nor by the analysis carried out. Therefore, it would be necessary to work out a more detailed test to show English comprehension as well as a more thorough data collection format. The whole project was a series of lessons in learning *on-the-go*. Much of the work was done over and over again in order to present findings and insights from the data analysed.

**Appreciation.**

The author would like to thank the following for their support, encouragement, and technical help (especially in the analysis): Professor O. Owolabi, Mr. Albert O. Onamusi, Dr. Charles I. Saidu.

Much appreciation goes to Africa Data School instructors – John Muchiri and Jacques Khisa for the many hours of teaching, and training, and patient instruction.

February 7, 2024

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Appendix A**

**Proof of Concept for Africa Data School Project.**

* *Submitted Jan 2022*

**THEME: Literacy Analysis in Primary School Pupils - Survey of 6(?) schools in Kuje Local Council in the Federal Capital Territory, Nigeria.**

**PURPOSE:** Literacy levels are generally seen to be failing. Several factors have been pinpointed as responsible. The ability of primary school pupils to read fluently and also understand what they are reading (and be able to express themselves) is a growing concern. There is a need to be able to address the various factors. However, these factors need to be discovered, itemized, displayed in such a way that the school owners and teachers, and other members of the educational workforce can more readily point out the needed solutions.

**AIM:** To conduct a survey of approximately one thousand pupils regarding literacy levels, particularly to come up with a standard of what is seen to work in the Nigerian educational setting considering second language speakers, and fluency in reading and understanding English at grade school level.

**TIME FRAME:** The survey should be carried out over a period of one month and using several schools in the FCT. As a precursor, a sample survey will be carried out in one community school using a few pupils from different classes, chosen specifically to explore feasibility. The data collating and cleaning process will require an additional time span. All in all, the project should not exceed 6 months, giving allowance for delays in communication and other situations.

**Pros and Cons:** To get up to 1000 respondents may require covering more schools. The data capture will include actual visits to schools. This may cause some delays as well as require some written permission. (Maybe even from parents)

**INSTRUMENTS:** A simple data sheet with personal details and a sample comprehension article will be handed out to students to fill and submit presumably same day. Alternatively, an e-form could be sent to parent’s email addresses, but that would require parents having time to sit the child down and fill the form and mail back, which precludes internet access, data and a host of other options.

**Pros and Cons:** Permission would have been sought beforehand. – Do I need a letter of introduction from Africa Data School?

**OBJECTIVE:** The data received will be collated and stored in an Access(?) Database. The data will then be retrieved at various forms, and exploratory data analysis used to clean up and display findings. The role of various descriptives in determining the fluency level will be explored with the aim of promoting ideals that build up fluency in reading. A secondary aim for a future project is to promote distribution of materials that will increase fluency presumably at different levels of basic education. This will be based on the model developed.

**Pros and Cons:** How will the machine learning model expose fluency level based on just one comprehension write-up?

*Further Objectives: To determine reading comprehension ability among children based on their language skills or number of languages spoken…*

*AIM1: Testing Comprehension using a time-based test*

*AIM 2: Number of languages spoken at school entry with ability to comprehend*

*AIM3: Determine the impact of English as a second language has on reading fluency (how children coming to school with no English background are affected in their learning and comprehension)- Do we know how teaching in English affects learning.*

**ACTIVITY:** Using the backdrop of classification algorithm in machine learning much of the activity to be covered will use Explanatory Data Analysis and Feature Engineering to explore various views. And then develop a machine learning model that can serve the educational establishments to project literacy levels and fluency across their schools. This will help to establish how various features are directly and indirectly responsible for literacy levels in primary school pupils. Hopefully also it will present a case study for future developments and expectations with digitized learning.

**Pros and Cons:** Will this analysis be given or sold to schools? How does this become a viable commercial project?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Appendix B**

**ADS Project Data Survey… 2023 NO.:\_\_\_\_\_\_\_\_**

**Section A.**

1. Write the name of your school: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
2. Surname: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ First Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Other Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. Write today’s date: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 4. Boy? \_\_\_\_\_\_\_\_\_\_\_ or Girl? \_\_\_\_\_\_\_\_\_\_\_\_
2. Write your date of birth.\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Section B. Answer the following questions by ticking the box or circle that is correct for you.**

**6. What language/s do you speak?**

* English □ Gbagyi □ Tiv □ Fulfulde
* Hausa □ Ibo □ Yoruba
* Other language/s (write the other language/s)\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**7. What class are you now (January-March2023)**

O Basic 3 O Basic 4 O Basic 5 O Basic 6

**8. What year did you start primary school?**

O 2020 O 2019 O 2018 O 2017 O 2016

**9. What language do you speak with your parents at home? Tick all that apply.**

* English
* Pidgin English
* Hausa
* Other (please write which other language or languages you speak with your parents?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**10. What language do you speak with your siblings (brothers and sisters) at home?**

* English
* Pidgin English
* Hausa
* Other (please write which other language or languages you speak with your siblings?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**11. What language do you speak with your friends at home?**

* English
* Pidgin English
* Hausa
* Other (please write which other language or languages you speak with your friends?)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**12. What language do you speak in school?**

* English
* Pidgin English
* Hausa

**13. When did you start speaking English?**

O. Primary School O. Pre-primary

O. Play group O. At home (that is before school)

**Section C Read the short passage below and tick the circle with the correct answer:**

Yusuf was a happy nine-year old boy who liked to sing and dance. He was often acting drama in school. He played football during break with his friends. But he did not like to read. His mother and father and all his teachers were always scolding him “Yusuf, you must read your books and study hard to become great in life.”

Yusuf came from a family of six; his mother, his father and two brothers and a sister. He lived in the village of Zarmaganda and went to school near his village. The name of his school was Sabon Rai Community School. Yusuf was in Basic 3.

**14. What was the name of Yusuf’s' school?**

O. Sabon Layi Community School O. Sabon Rijiya Community School

O. Sabon Rai Community School O. None of the above

**15. How old was Yusuf?**

O. Six O. Seven O. Eight O. Nine

**16. How many brothers and sisters did Yusuf have?**

O. Four O. Two O. Three O. Five

**17. What was Yusuf happy doing?**

O. playing football O. dancing only

O. drama only O. singing and dancing

**18. What was the name of Yusuf’s village?**

O. Zarmagi O. Zarmaga

O. Zarmaganda O. Zarmabada