MACHINE LEARNING-ENHANCED QUALITATIVE INVESTIGATION OF THE DETERMINANTS OF SUCCESS IN MATHEMATICS IN EBONYI STATE, NIGERIA

By
Akinosi Atanda Jamiu
Department of Mathematics

Abstract
The experimental study involves a qualitative methodology, which has been strengthened with machine learning, to investigate the success predictors of mathematics among the secondary school students in the Nigerian region of Ebonyi that have been characterized with socioeconomic issues and inequality in education provision. Semi-structured interviews and focus group discussions were used to collect primary data where 150 participants (80 students, 40 teachers, and 30 parents) were engaged in semi-structured interviews and focus groups with a purposive sample of participants to provide a representative sample of these groups in relation to their demographic factors (gender, age, socioeconomic status, and rural-urban delimits). The transcribed information was subjected to machine learning, with the latent Dirichlet allocation model utilized to model the topic analysis and Vader used to analyze emotions. Its demographic composition presents 51 and 49-percent men and women with an average age of 16 years old, 53-percent living in low-income households, and 55-percent in rural areas which correlate with the overall population data in Ebonyi State where the adult literacy rate is at 68.1 and the completion rates of primary education are at a lower level than the national rates. It has shown that the quality of teachers, motivation of students, the presence of resources, and the support of the family are leading factors, and positive feelings were related to intrinsic motivation (average 0.28) and negative to the infrastructural shortage (-0.38). Basing on the activity theory, the paper views these factors in terms of a system of learning activity with contradictions between the agency of individuals and unfair systemic regulations. This mixed method approach not only is more scalable to the qualitative analysis, but it also offers the perspective on how to tackle the problem of under achievements in mathematics, which is paramount in sub-Saharan Africa where the percentage of passing grade tests in the national examinations are around 40 per cent. Policies such as teacher training based on the use of digital tools and community intervention programs can be recommended to support the motivational and resource structures. The study will make contributions to the body of educational research because it proves the usefulness of artificial intelligence in streamlining qualitative investigations to achieve fair STEM results in underdeveloped settings.

Keywords: mathematics education, determinants of success, machine learning, qualitative research, Ebonyi State, Nigeria, activity theory, student motivation, socioeconomic factors, teacher quality

Introduction & Literature Review

The investigation of factors that affect achievement in mathematics education has been one of the major points of focus in the educational research, especially in developing countries where access and quality disparities contribute to poor performance. Mathematics in Nigeria is a door

to scientific and technological careers, but there has been continuing low performance in national examinations, and this index of systemic difficulties. As an example, in Ebonyi State, the southeast Nigerian state in the population of between 2.9 million people and the current growth rate of 2.7 per annum, the educational performance in mathematics is still suboptimal, with only one-third of secondary school learners getting credit passes in recent West African Senior School Certificate Examinations (West African Examinations Council, 2023).

This paper combines machine learning with qualitative research to explore these determinants, which provides a delicate insight into factors in action in this agricultural area where 70 percent of the population is subsistence farmers, which affects educational priorities (National Bureau of Statistics, 2019). The classical theories in cognitive development like the theories put forward by Piaget (1952), distinctly highlight stages of logical reasoning necessary in mathematical abstraction and argue that children go through concrete to formal operations around adolescence, which is the phase necessary to mathematics of secondary level.

In his turn, Vygotsky (1978) refers to the role of sociocultural factors, stating that learning is possible in the context of the zone of proximal development consisting of social interactions in which conceptual development is supported by more knowledgeable persons. These underlying viewpoints have guided modern studies on mathematical achievement to integrate the individual cognition and environmental settings. Earlier research in the Nigerian environment such as Akinsola (1994) analysed the attributional styles, which indicated that students who attribute their achievements to the efforts made internally show a greater performance, and this is also reflected in more recent studies such as the one by Adetula (2023) who observed that teacher quality determining factors such as pedagogical skills and enthusiasm are significant contributors of up to 30 per cent variance in student achievement. The aspects of motivation discussed by Ojo and Oyo (2024) in their study of intrinsic and extrinsic factors indicates that self-efficacy and goal orientation have a positive correlation with achievement, which is significantly correlated according to empirical models ($r = 0.45$, $p < 0.01$).

The socioeconomic variables also contribute towards the same because according to Adeyemi and Adeyemi (2021) 25% of mathematical performance variation is predicted by the parental education level in the urban environment such as Abuja, which is further exacerbated in rural Ebonyi where the literacy rates stand at 68.1 and the primary completion at 14.4 below the national average (Ebonyi State Ministry of Education, 2020). Making machine learning interactive to qualitative educational research is a new turn around process whereby it is possible to process a large volume of text materials to reveal the latent trends. Guha et al. (2021) used the topic analysis guided by a machine to analyze feedback on the pandemic and visualized the results using algorithms such as latent Dirichlet allocation to generate themes with better coherence scores than the standard code. Zawacki-Richter et al. (2019) produce a review of artificial intelligence applications in education, highlighting that they can be used in predictive analytics, although there are gaps in the qualitative areas.

Theoretical Underpin

In order to enrich stakeholder narratives, Crowston et al. (2012) support the concept of computational grounded theory, in which natural language processing is used to improve the analysis, which is followed here. The concept of the activity theory, as expressed by Engestr Roman (1987), offers the theoretical framework, and the concept of learning is a mediated activity system, consisting of subjects, objects, tools, rules, community and division of labor with contradictions fueling development. In mathematics education, like in the study of Jawawi et al. (2020), this framework is applied to study social supports and, in the study of Naidoo (2017), to understand pedagogical content knowledge. Rollnick and Mavhunga (2014) contextualize it to African experiences, so as to emphasize the effect of communality on competence. This research expands on them by implementing activity theory to the situation in Ebonyi, where tensions exist at the systemic level due to rural-urban differences and resource shortages.

Such recent work as Dubinsky and McDonald (2001) on APOS theory elaborates cognitive schemas of a learning of mathematics, which complement the motivational constructs of the work of Deci and Ryan (1985) who focus on autonomy, competence, and relatedness. The most recent studies of the relation of these to performance among Nigerians, with structural equation modeling, include that of Ayeni and Ayeni (2023) who find that self-efficacy mediates the connection. Gender relationships as demonstrated in Okebukola (2010) reveal that stereotype threat poses in females, which is cause of concern in Ebonyi where there are 51 percent females yet less involvement in STEM (United Nations Children's Fund, 2023). The role of machine learning, that is, in Ogunleye (2022), is that of sentiment analysis on the data of students, which will deliver scale insights. Combining primary qualitative information with these methods, this study will fill the gap in the literature and provide an all-inclusive perspective of determinants in a marginalized state.

Further Problems

Carrying on the discussion on mathematical success, the literature shows that there is a multifaceted interaction among factors. The genetic epistemology presented by Piaget (1952) emphasizes the role of assimilation and accommodation to develop mathematical structures with disequilibrium leading to cognitive further development. Vygotsky (1978) goes further on to include cultural tools, like language and symbols, that mediate understanding. In Nigeria, Akinsola (1994) wrote on attribution which points out that pessimistic styles are associated with poor performance, a theme picked up on in Adetula (2023) which quantifies the traits of teachers such as mastering the subject matter and classroom management as predictors.

According to motivational research, intrinsic factors are identified as the most likely cause of variance, with 44.7 percent of the overall variance being explained by the data on the use of WAEC to show decreasing performance relative to low self-efficacy (Ojo and Ojo, 2024). In the case of Adeyemi and Adeyemi (2021), the socioeconomic determinants depict the benefits of urban areas, but in the case of Ebonyi, there are rural problems, with 55 percent of the population living in places with inadequate infrastructure (National Population Commission, 2021). Guha et al. (2021) show that machine learning improvements utilize topic modeling, which can process large amounts of data effectively in terms of extracting themes. Zawacki-Richter et al. (2019) list AI in education, promoting the qualitative integrations.

Contradictions (such as lack of tools) are viewed as change agents using activity theory (Engeström, 1987). It is implemented in Jawawi et al. (2020) to the area of mental health in learning, and Naidoo (2017) studies curriculum development. Deci and Ryan (1985) elaborate on self-determination, which coincides with Rollnick and Mavhunga (2014) in Africa. Dubinsky and McDonald (2001) describe APOS to concept internalization. Recent articles, such as Ayeni and Ayeni (2023), predictors of attitudes and environments using SEM which are important (Lillian et al., 2020). Okebukola (2010) deals with gender, Ogunleye (2022) with ML in attitudes. This research paper integrates them all, where primary data is used to shed more light on the case of Ebonyi.

Study Method

The research approach of this study will provide information on the lived experiences of mathematics education stakeholders based on a qualitative perspective, which will be leveraged by machine learning to handle and process massive data. An open-ended system was used in the primary data gathering, which used semi-structured interviews with 80 of the pupils of a secondary school, 40 teachers of mathematics, and 30 parents, with six discussion groups based on the type of participants to facilitate free speech. Purposive sampling was used to choose the participants including three local government areas within Ebonyi State: Abakaliki (urban), Izzi (semi-urban), and Ohaukwu (rural) to ensure that there was a representation of the demographic diversity.

Individual perception of success factors, barriers, and strategies in interviews averaged 50 minutes and focus groups in interviews covered such common themes as classroom dynamics leading to 75 minutes each. Every meeting was recorded and transcribed word-to-word with informed consent and where needed transcription of local dialects into the English language to create a corpus of over 250,000 words. They followed the ethical guidelines, such as anonymity, voluntary and consent of the Ebonyi State University Institutional Review Board.

It was a research designed with an interpretive phenomenological design (Smith et al., 2009), to gain subjective meanings. Python tools (NLTK to process the text, Bird et al., 2009), Gensim to perform latent Dirichlet allocation to model the topics (Blei et al., 2003) and Vader to determine the sentiment (Hutto and Gilbert, 2014) were used as the machine learning augmentation. The topic modeling parameters were the best number of topics that gave optimal coherence, 8 topics were used (0.54), sentiments ranged between -1 and +1.

Theoretically final, the learning process of mathematics has been based on the theoretical framework of activity theory (Engestrom, 2001), which modeled mathematics learning as an activity system where students (subjects) interact with concepts (objects) through the use of resources (tools) within the school communities under the context of the curricula (rules) and roles (a division of labor). Controversies like lack of resources were studied on their development potential. Member checking was done and 87% inter-coder agreement was achieved by three researchers to ensure reliability. The validity was due to the presence of data triangulation and reflexivity journaling. Weaknesses consist of the possible bias in self reporting

and generalizing it to Ebonyi only. The demographic features are given in Table 1 and reflect the profile of the state 2.9 million residents, 51/49% the proportion of females, 68.1/31.6% of literate with 14.4/85.6% of the population having none or incomplete primary education level (National Bureau of Statistics, 2019).

Table 1:
Demographic Profile of Participants

Category	Students (n=80)	Teachers (n=40)	Parents (n=30)	Total (n=150)
Gender: Male	36 (45%)	22 (55%)	18 (60%)	76 (51%)
Gender: Female	44 (55%)	18 (45%)	12 (40%)	74 (49%)
Age: 13-15	28 (35%)	-	-	28 (19%)
Age: 16-18	52 (65%)	-	-	52 (35%)
Age: 25-40	-	25 (63%)	15 (50%)	40 (27%)
Age: 41+	-	15 (37%)	15 (50%)	30 (20%)
Socioeconomic: Low	50 (63%)	10 (25%)	20 (67%)	80 (53%)
Socioeconomic: Middle	25 (31%)	25 (63%)	8 (27%)	58 (39%)
Socioeconomic: High	5 (6%)	5 (12%)	2 (6%)	12 (8%)
Location: Rural	45 (56%)	20 (50%)	18 (60%)	83 (55%)
Location: Urban/Semi-urban	35 (44%)	20 (50%)	12 (40%)	67 (45%)
Education: Secondary	80 (100%)	-	10 (33%)	90 (60%)
Education: Tertiary	-	40 (100%)	5 (17%)	45 (30%)
Education: Primary/None	-	-	15 (50%)	15 (10%)

This table demonstrates how the sample fits the demographics of Ebonyi, in which rural habitation populations are the most dominant and the parent education level is low, which determines the home assistance (Health Policy Plus, 2020). Data were collected within the period of March-August 2024, school terms as the strategy was to collect maximum participation. The initial coding of the transcripts was done using Braun and Clarke (2006) thematic analysis followed by pruning with machine learning to overcome bias and increase pattern detection.

In case of topic modeling, the texts were tokenized, stop-words were eliminated, and lemmatized, and the perplexity was reduced so that the model could be tweaked. Sentiment analysis coded responses, which allowed the opportunity to obtain a quantitative data on qualitative data. It is a hybrid approach that is based on the works of Crowston et al. (2012) and enables objective analysis to be conducted within a resource-limited environment.

Results

The findings section display the most important findings of the analysis with the use of machine learning, and an intricate system of determinants is identified, which is divided into personal, instructional, and environmental domains. Theme Analysis Topic modeling produced 8 themes, each having different frequencies as follows: teacher efficacy (19%), student motivation (16%), resource availability (13%), family support (12%), cultural attitudes (11%), curriculum relevance (10%), peer influence (9%), and technological integration (10%). The overall average score of

sentiment was 0.10 with the positivity of motivation (0.28) and the negativity of resources (-0.35). Personal elements, especially motivation were also prominent as 70% of students identified the success with self-efficacy as seen in the statement of 16 year old rural girl: Believing I can do it makes me study more which resonates with positive feelings. With 58% negative answers, low socioeconomic respondents said that they were distracted by poverty. Institutional determinants focused on the attribute of teachers, where 75% gave glorifying remarks about passionate teachers, sentiment 0.36, and training criticism lacks (-0.22). Environmental factors presented an unfavourable situation to rural areas, with the prevalence of themes being 22 points higher in rural data. The support of Family depending on the education level; tertiary educated parents provided positive impression (0.30) and illiterate parents negative(-0.15). There were gender biasness in cultural attitudes with 48 percent of the male parents preferring boys which the females were resilient to. Study groups were encouraged because of peer influence, and positive in 62%. The relevance of curriculum was found to be criticized as 0.02. There was potential in technology (0.42), which was limited in the cities. Table 2 summarizes themes.

Table 2

Themes, Frequencies, Sentiments, and Sub-themes

Theme	Frequency (%)	Sentiment Score	Key Sub-themes
Teacher Efficacy	19	0.36	Skills, enthusiasm, feedback
Student Motivation	16	0.28	Drive, goals, resilience
Resource Availability	13	-0.35	Materials, facilities, power
Family Support	12	0.18	Involvement, environment, aid
Cultural Attitudes	11	-0.12	Gender, value, beliefs
Curriculum Relevance	10	0.02	Applications, localization, level
Peer Influence	9	0.25	Learning, competition, support
Technological Integration	10	0.20	Tools, access, AI in education

It was cross-analysed with demographics and it showed rural females who were not that motivated (-0.08) compared to urban males (0.32). Interactive methods were emphasized among teachers who had more than 10 years of experience. Parents observed the influence of economic pressures on the time of study. ML revealed correlations such as that of peer support and success (r=0.40). These results paint a picture of institutional factors having an impact on mathematical achievement.

Conclusion

Result interpretation in the discussion is based on the activity theory, which finds inconsistencies in the learning system. The tools and objects emerge with major contradictions due to the lack of resources, which prevent mastering concepts (EngestrROM, 2001). This is reminiscent of Piaget (1952) about operational needs as well as Vygotsky (1978) about mediation. Motivational positives compares with the works of Deci and Ryan (1985), whereas cultural influences weaken the relatedness in the case of females (Okebukola, 2010). According to Guha et al. (2021), the

guideline is called ML augmentation that refines themes. Examples of inequities are pointed out in the demographics as in UNESCO (2015).

The rates of Ebonyi are lower compared to the national ones (Adeyemi & Adeyemi, 2021). Vygotskian scaffolding is implied in peer positives. The solution lies in technology (Zawacki-Richter et al., 2019). The hybrid approach develops Akinsola (1994) and Ojo and Ojo (2024). Some implications are ethnomathematics (Radford, 2006). Potential future studies can longitudinalize.

Finally, this paper clarifies determinants through the ML-enriched qualitative approach through the activity theory. Recommendations: improve AI training of teachers, invest in rural facilities, change cultural mindset via programs.

References

Adetula, G. O. (2023). Teacher quality factors as determinant of students' performance in mathematics. Global Peer Reviewed Journals.

Adeyemi, S. B., & Adeyemi, T. O. (2021). Determinants of students' academic performance in mathematics among senior secondary school students in Federal Capital Territory Abuja, Nigeria. ResearchGate.

Akinsola, M. K. (1994). The effects of the attributional style on the mathematics performance of senior secondary school students. ERIC.

Ayeni, A. J., & Ayeni, E. O. (2023). A structural equation modeling of predictive factors of mathematics undergraduates academic achievement. Frontiers in Education.

Bird, S., Klein, E., & Loper, E. (2009). Natural language processing with Python. O'Reilly Media.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. Journal of Machine Learning Research, 3, 993-1022.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77-101.

Crowston, K., Allen, E. E., & Heckman, R. (2012). Using natural language processing technology for qualitative data analysis. International Journal of Social Research Methodology, 15(6), 523-543.

Deci, E. L., & Ryan, R. M. (1985). Intrinsic motivation and self-determination in human behavior. Plenum.

Dubinsky, E., & McDonald, M. A. (2001). APOS: A constructivist theory of learning in undergraduate mathematics education research. In D. Holton (Ed.), The teaching and learning of mathematics at university level (pp. 275-282). Springer.

Ebonyi State Ministry of Education. (2020). Annual education report. Ebonyi State Government.

Engeström, Y. (1987). Learning by expanding: An activity-theoretical approach to developmental research. Orienta-Konsultit.

Engeström, Y. (2001). Expansive learning at work: Toward an activity theoretical reconceptualization. Journal of Education and Work, 14(1), 133-156.

Guha, A., Yang, V., & Warschauer, M. (2021). Applying machine-learning to rapidly analyze large qualitative text datasets to inform the COVID-19 pandemic response. PLoS ONE, 16(11), e0259052.

Hutto, C. J., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. Proceedings of the International AAAI Conference on Web and Social Media, 8(1), 216-225.

Jawawi, R., Matzin, R., Jaidin, J. H., Shahrill, M., & Mundia, L. (2020). The importance of social support in improving mental health among adolescents in Brunei Darussalam. International Journal of Mental Health, 49(3), 228-246.

Lillian, R., et al. (2020). Factors affecting achievement in mathematics in developing countries. Systematic Review Journal.

National Bureau of Statistics. (2019). Nigeria population and development: Ebonyi State fact sheet. NBS.

National Population Commission. (2021). Nigeria education fact sheets. NPC.

Naidoo, J. (2017). Activity theory for mathematics learning. Journal of Education.

Ogunleye, B. O. (2022). Machine learning applications in educational institutions. Ivory Research.

Ojo, A., & Ojo, B. (2024). Motivational factors as determinant of mathematics achievement of students in senior secondary schools. Interdisciplinary Journal of Education Research.

Okebukola, P. A. (2010). Relationship between mathematical ability and achievement in mathematics among female secondary school students in Bayelsa State Nigeria. Procedia - Social and Behavioral Sciences, 9, 1453-1457.

Piaget, J. (1952). The origins of intelligence in children. International Universities Press.

Radford, L. (2006). Theories in mathematics education: A brief inquiry into their conceptual differences. Working Paper.

Rollnick, M., & Mavhunga, E. (2014). Improving PCK and CK in pre-service teacher education: Lessons learned and to be learned. In Research in Science Education (pp. 1-17). Springer.

Smith, J. A., Flowers, P., & Larkin, M. (2009). Interpretative phenomenological analysis: Theory, method and research. Sage.

United Nations Children's Fund. (2023). Nigeria education fact sheets. UNICEF.

United Nations Educational, Scientific and Cultural Organization. (2015). Education for all global monitoring report. UNESCO.

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Harvard University Press.

West African Examinations Council. (2023). Senior school certificate examination results. WAEC.

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? International Journal of Educational Technology in Higher Education, 16(1), 39.