Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Evaluating the efficacy of basic level standard assessment in conflict-affected regions: Initial findings from 57,473 pupils in Northeastern Nigeria

Jamiu S Olumoh
American University of Nigeria
Audu Liman
American University of Nigeria
Anton Pak
University of Queensland
Presly R. Obukoadata
Delta State Univeristy
Oyelola A. Adegboye
oyelola.adegboye@menzies.edu.au

Charles Darwin University

Article

Keywords: Numeracy, Literacy, School effectiveness, Multivariate multilevel analysis, Structural equation modelling

Posted Date: April 29th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9088466/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Evaluating the efficacy of basic level standard assessment in conflict-affected regions: Initial findings from 57,473 pupils in Northeastern Nigeria

Abstract

Background: With the recent insecurities and violence against schoolchildren, the quality of the educational system in Nigeria has been declining, particularly in Northern Nigeria. This study aims to shed light on the situation by eliciting, describing, examining, and analyzing the foundational literacy and numeracy skills of primary school pupils. To achieve this, we have introduced the newly designed Basic Level Standard Assessment (BALSA) as part of the Strengthening Education in the Northeast States (SENSE) intervention project. The study also provided initial evidence of the effectiveness of the BALSA in addressing the educational challenges in the region.

Methods: A random effect-based structural equation modelling technique was used to analyze the aggregated primary surveyed data from 669 primary schools in Northern Nigeria. The data was collected using validated curriculum-based skill tests.

Results: The results showed significant inter-correlations between pupils' competence in literacy and numeracy subscales. Confirmatory factor analysis supported the factor structure of numeracy and literacy subscales with loadings ranging from 0.32 to 0.89. Across subscales and in both primary grades 2 and 4, the average *reading* score has been consistently high, and correlations between *reading* and other subscales were among the highest. Participants in grade 2 received the lowest scores in *fractions,* whereas participants in grade 4 received the lowest scores in *geometry*.

Conclusions: The study underscored the heterogeneity in educational attainment across pupil groups and highlighted the need to assess primary school students' learning outcomes when evaluating educational interventions. It is recommended that intervention projects should include research designs that account for students, school-level and teacher's characteristics.

Keywords: Numeracy, Literacy, School effectiveness; Multivariate multilevel analysis; Structural equation modelling

Introduction

Although Nigeria has seen some progress in reducing out-of-school rates among school-age children, from 37% in 2000 to 28% by 2021, it still has one of the highest numbers of out-of-school children worldwide (VOA, 2022, UNESCO 2022; Nwoke et al 2024). In 2021, they made up one in five of the global total of 244 million children aged 6-17. (VOA, 2022; UNESCO 2022; Nwoke et al 2024). The most recent estimates place the national figure at 18.3-20 million children and youth, which equates to 37% net attendance in primary and secondary levels, with stark regional disparities persisting: around 60% out-of-school rates in the North-East, ~50% in the North-West, and ~40% in North-Central regions, and to under 20% in the South (Nwoke, Oyiga, & Cochrane, 2024; Udefuna, Ezenwajiobi, & Udom, 2025; UNESCO, 2022; UNICEF, 2017). The insurgency and other violence-related activities mainly affecting Northern Nigeria have been cited as a major contributor to the disruption of school education (Bilyaminu, Iya, & Purokayo, 2017; Dauda, 2014; Muhammad & Mohammed, 2019; Nnaemeka, 2019). The violence has led to the destruction of many schools, which resulted in their closure and the displacement of families, teachers, and students. For about two decades, the insurgency has brought the educational system in Northeastern Nigeria to a near-total collapse, with more than one million school-aged children being forced to flee their homes and becoming disadvantaged in learning and labour opportunities at the height of the Boko haram insurgency (Ensign, 2016). About 802 schools remain closed in the region, and 1889 classrooms were reported as damaged or destroyed in 2014 (UNICEF, 2017). However, with the increasing stability in the region and improving support from international organizations in the past decade, most of the destroyed educational facilities have been repaired, and new schools have been built to cater for the educational needs of the region. These disruptions have had a long-lasting effect on the attainment of school knowledge in the region. The data on literacy and numeracy indexes revealed a

considerable gap in educational development between states in the Northeastern region and those in other regions (Haggai, 2009; Keck, 2016; Ogbonna, 2016).

Several factors contribute to the widespread poor academic performance in the examined context. These factors can be broadly categorized within the framework of teacher empowerment, exposure intensity, language skills (Oyetunde, 2016), and levels of school effectiveness, as defined within Scheerens' School Effectiveness School (SES) model. In this context, 'school' encompasses all the components, processes, and intricacies that define the educational experience and learning outcomes. The effectiveness of a school can be evaluated through the input-process-output model outlined by Scheerens (1990). Here, the teacher's role is considered essential input and significantly impacts students' academic achievements (Table 1). This impact is influenced by various factors, including teachers' qualifications, competencies, pre-service and in-service training, and the teacher-to-student ratio (Akinsolu, 2010; Akiri, 2013; Igbinedion, Newby, & Sparkes, 2017; Keck, 2016; ESSPIN, 2016; Owan, 2012).

The 'process' aspect of school effectiveness encompasses the various structures and methods that facilitate teaching and learning, such as curriculum design, school organization, and the overall school climate. In response to these challenges, a new educational initiative, "Strengthening Education in the Northeast States (SENSE)," was introduced in the Northeastern region in 2019 (Martens et al., 2020; Obukoadata, Hammler, Yusuf, Liman, & Olumoh, 2024; Rinnert et al., 2020). Within the context of the SES model, which identifies factors that limit school effectiveness, particularly in conflict-affected areas, this study aims to explore, describe, and assess the SENSE Basic Level Standard Assessment (BALSA) instrument.

Despite recent reconstruction efforts and international support, primary school pupils in Nigeria's Northeast states continue to lag far behind national and regional benchmarks in

foundational literacy and numeracy skills. The insurgency and its aftermath have created a "lost generation" of learners with severe learning gaps, exacerbated by teacher shortages, inadequate training, poor school infrastructure, and socio-economic barriers. This study addresses the critical problem of persistently low literacy and numeracy proficiency among primary school pupils in Adamawa and Gombe States, Nigeria, where the SENSE intervention aimed to improve educational outcomes; however, baseline data revealed significant deficits that require targeted evaluation.

This study aims to deepen understanding of foundational literacy and numeracy in conflict-affected Northeastern Nigeria and to appraise the BALSA assessment within the context of the SENSE project. The research is framed by the following questions:

RQ1: What are the baseline levels of literacy and numeracy proficiency among grade 2 and 4 pupils in the target LGAs of Adamawa and Gombe States as measured by BALSA?

RQ2: Which sociodemographic factors are significantly associated with higher or lower BALSA test scores?

RQ3: How well does BALSA perform as a measurement tool in this setting?

RQ4: What actionable insights can guide improvements in teacher training, school management, and curriculum delivery under Project SENSE?

Intervention Overview: Project SENSE and BALSA

Project SENSE was funded by the United States Agency for International Development (USAID) and was designed to address key challenges faced by the basic education system in the Northeastern region of Nigeria by improving the capacity of educational managers and using evidence-based methods to plan and deliver basic educational services in line with the states' education sector strategy. Further details of the SENSE project have been described elsewhere (Martens et al., 2020; Obukoadata et al., 2024; Rinnert et al., 2020). Briefly, the core premise of this initiative was that empowering trained

educational managers would, in turn, enhances teacher's instructional abilities, leading to improved educational outcomes for children in the states. The SENSE project adopted a collaborative approach, engaging with various education stakeholders, including national and international partners. Together, they co-designed activities, gained a comprehensive understanding of the existing educational landscape, and devised strategies to address a range of scenarios and interventions. The project's primary focus areas included English, mathematics, and technology in education (Martens , 2020; Rinnert , 2020).

As an integral part of this intervention program, the Basic Level Standard Assessment (BALSA) was administered at the outset among pupils in primary Grades 2 (P2) and 4 (P4) in select regions. Specifically, this assessment was conducted in five of the twenty-one local government areas in Adamawa State and three of the eleven in Gombe State. The BALSA test items were meticulously designed to assess students' literacy and numeracy skills in these two core subject areas, providing a baseline measure of their knowledge and abilities before the commencement of the targeted educational interventions in the selected schools.

Literature review

Despite sustained national and international interventions aimed at rebuilding education systems following the Boko Haram insurgency, schooling environments in the region continue to face structural and institutional challenges that limit effective teaching and learning. Studies examining information and communication technologies for education (ICT4E) in Adamawa and Gombe States consistently report damaged or inadequate infrastructure, overcrowded classrooms, unreliable electricity supply, weak internet connectivity, and shortages of essential teaching and learning materials (Awoyemi, 2023). Besides, educators and school leaders have limited digital competencies and low levels of data use for instructional improvement, reflecting a persistent disconnect between national ICT policy expectations and local implementation capacity (Fullan, 2013). National-level evidence similarly points to a persistent learning crisis despite increased school enrolment. Adeniran, Ishaku, and Akanni (2020) report that only a small proportion of early-grade pupils achieve minimum literacy and numeracy competencies, with children in rural areas, government schools, economically disadvantaged households, and northern regions disproportionately affected. Conflict-induced displacement further exacerbates educational inequality, contributing to marginalisation and disrupted learning trajectories among children living in IDP camps (Ensign, 2016; Rwang, 2023). These findings underscore that expanded access to schooling has not translated into improved learning outcomes, particularly in fragile contexts.

In response to these challenges, various policy and programme interventions have sought to address these challenges through both demand- and supply-side strategies. School feeding programmes have demonstrated positive effects on enrolment and academic participation (Barnabas, Bavorova, & Madaki, 2024), while initiatives aimed at improving school safety and physical infrastructure have attempted to restore conducive learning environments (Oladunjoye & Omemu, 2013). Additional interventions have focused on strengthening security

arrangements and providing learning resources in conflict-affected schools (Ataseriaba, 2016). However, empirical evidence suggests that implementation outcomes often fall short of policy expectations due to accessibility barriers, weak institutional coordination, and continued infrastructural deterioration. The divergence between policy commitments and lived educational realities remains a central concern across the literature.

In the northeast, targeted interventions have increasingly focused on context-sensitive and resource-appropriate approaches to improving foundational learning outcomes (Kyriakides, 2005; Kyriakides & Creemers, 2008). The USAID-funded SENSE project was designed to improve literacy and numeracy through teacher professional development, community engagement, and the delivery of low-technology instructional resources tailored to local conditions in Adamawa and Gombe States (Martens et al., 2020; Rinnert et al., 2020). Evidence from previous studies shows systemic barriers, which include: insufficient instructional materials, overcrowded classrooms, irregular teacher remuneration, and weak institutional support systems. Nevertheless, studies also reveal strong professional commitment among teachers and responsiveness to locally relevant innovation. Participatory co-design processes involving educators and school-based management committees have demonstrated that collaborative development of instructional materials improves cultural relevance, acceptability, and sustained use within communities (Martens et al., 2020; Rinnert et al., 2020).

Prolonged school closures during COVID-19 and insecurity further necessitated experimentation with alternative learning delivery models. The Transactional Radio Instruction (TRI) programme under the SENSE initiative employed interactive Hausa-language radio lessons supported by printed workbooks and community facilitation to sustain learning among Grade 3 pupils during disruptions. Quasi-experimental evaluation findings show that programme participants achieved significantly higher scores on early-grade reading assessments than non-participants after controlling for socio-demographic characteristics.

Learning gains were particularly strong among Hausa-speaking learners, highlighting the importance of linguistic alignment and contextual adaptation in instructional design. Qualitative evidence further indicates strong community acceptance of low-technology delivery mechanisms, reinforcing their relevance in fragile and resource-constrained educational contexts (Obukoadata, et al, 2024).

Scholars have also adopted diverse methodological approaches to understand learning dynamics in conflict settings. Geospatial analyses have examined how patterns of insurgency shape educational participation and community learning cultures (Sola, Chen, & Subrahmanian, 2025), while interdisciplinary studies have explored the interaction between health vulnerabilities, social instability, and educational access in violence-affected environments (Nggada et al., 2025). Qualitative research has moved beyond institutional analyses to capture everyday schooling experiences and social dimensions of educational access (Humphreys, Moses, Kaibo, & Dunne, 2015; Donnelly, 2026). Quantitative studies employing panel fixed-effects models have further demonstrated that variations in conflict intensity influence educational outcomes over time and across locations (Bertoni, Di Maio, Molini, & Nisticò, 2019). Emerging work on computer-based assessment systems also suggests potential pathways for strengthening learning measurement and accountability mechanisms (Alarape, 2024).

Nonetheless, these findings point to a direction that is relevant to deploying and testing the efficacy of the basic learning standard assessment for learners, as well as to adequately enhancing learning standards. Therefore, calling for a new method of learning and measuring outcomes. Our study looked beyond these findings, explored the interplay among them given the level of recent interventions as encapsulated in the SENSE policy thrust and determined what efforts are needed to optimise the gains achieved. Consequently, we looked beyond the provision of more teachers as envisaged by previous studies (Bissallah, 2024), insurgency

(Chatham House, 2016), adequate classrooms and funding (Oladunjoye & Omemu, 2013), engaged curriculum development (Chua & Don, 2013), religious inclination and conflicts (Ataseriaba, 2016) to providing advocacy for the process among the critical stakeholders - the beneficiaries.

Methodology and materials

Study design

This study employed a cross-sectional design, using baseline data from primary school pupils to describe foundational literacy and numeracy levels in conflict-affected Northeastern Nigeria and to assess the performance of the newly developed BALSA within the SENSE intervention.

Study Region and Sampling Method

This study encompassed 669 primary schools, strategically selected across eight randomly selected local government areas (LGAs) in the first phase of the SENSE intervention program in Adamawa and Gombe States, Nigeria (Figure 1). There are 21 LGAs in Adamawa State, out of which Mayo-Belwa, Numan, Yola South and Song were randomly selected. In Gombe, with 11 LGAs, Akko, Biliri, Gombe, and Nafada were the four randomly selected. In each LGA, we selected all primary schools for our study, except Akko LGA, where 100 schools were randomly selected from a pool of over 210. The study had a total enrolled capacity of 30,392 for P2 pupils and 27,091 for P4 pupils, which were available for the assessments. Thus, the sample size was the entire population.

Instruments and Data Collection

The SENSE BALSA instrument adopts a curriculum-based measurement approach to evaluate pupils' foundational numeracy and literacy skills. It was meticulously developed through multiple stages and tailored to the distinct needs of various classes and regions studied (Martin et al. 2020, Hernandez-de-Menendez et al., 2021). BALSA replaces multiple-

choice formats with immersive simulations that reflect real-world challenges and utilises blockchain to provide tamper-proof learning records (Conrad, Kablitz & Schumann, 2024). BALSA features a bias-mitigated algorithmic design for equitable outcomes across gender, ethnicity, and socioeconomic groups, offering individualised support in overcrowded classrooms (Zhang, Basham & Yang, 2020), tracks skill progression over decades to help policymakers identify educational gaps (Chavez-Rivera et al., 2024), and minimises human gender bias in evaluations (Figueras et al., 2025).

Question test banks were collaboratively developed by educators and scholars from each state under the supervision of the education resource center, responsible for curriculum development. To ensure impartiality, these questions underwent rigorous review and validation by the Quality Control Department of the Ministry of Education, and by subject teachers at Colleges of Education in both states.

The literacy test items were specifically designed to assess pupils' reading and phonics awareness. For Primary 2 (P2) pupils, there were ten reading items and five phonics awareness items, while Primary 4 (P4) pupils encountered ten reading items along with additional phonics items. This design aligns with the curriculum's emphasis on phonics in higher primary classes.

Conversely, the numeracy assessment was designed to assess pupils' functional mathematical proficiency. It comprised fifteen test items for P2 pupils and twenty for P4 pupils. In P2, these numeracy items were categorized into three sub-scales: basic operations (six items), mensuration (five items), and fractions (four items). Similarly, for P4, the categories consisted of six items on basic operations, four on mensuration, six on fractions, and four on geometry. These designations closely adhered to the basic education curriculum outlined in the Nigerian National Policy on Education (NERDC, 2014, p.9-12).

The BALSA test was administered in all 669 primary schools during phase one of the SENSE project, as detailed in the sampling section. However, as a novel instrument for assessing numeracy and literacy skills in primary school children, it may have inherent limitations. These limitations warrant further evaluation and refinement to enhance the national education policy and delivery mechanisms. The test findings serve as valuable recommendations for potential improvements in the educational system."

Primary Grades 2 and 4 were selected under the SENSE protocol to assess foundational literacy skills using the Early Grade Reading Assessment (EGRA) benchmarks. Grades 3 and 6 were excluded because they served as intervention endpoints in the SENSE programme. Eligible participants included pupils enrolled and present in SENSE-supported schools during the baseline assessment conducted in August 2019.

The BALSA literacy and numeracy tools demonstrated strong psychometric properties. All four scales exhibited high internal consistency, with Kuder-Richardson Formula 20 (KR-20) coefficients ranging from 0.858 to 0.874, surpassing the commonly accepted threshold of 0.80 for educational assessments (Ntumi et al, 2023). The item difficulty levels were within the acceptable range (0.20-0.80), confirming that the items were appropriately challenging. The corrected item-total correlations generally exceeded the recommended discrimination threshold of ≥0.30, indicating that most items effectively distinguished between higher- and lower-performing pupils. The only literacy item for Grade 4 was less discriminative, but removing it had an insignificant impact on the overall reliability. See Supplementary Material Tables S1-S5 for detailed reliability and validity analyses.

Statistical analysis

To assess the relationships between BALSA instruments, we employed a random-effects multivariate mixed model that combines elements of the linear mixed-effects model

(LMM) and the structural equations model (SEM). The SENSE baseline data used in this study exhibits a multilevel structure, with pupils nested within schools, and schools nested within the Local Government Area (LGA). This data structure implies that individual observations are not entirely independent, as they are correlated within levels. Consequently, traditional least-squares regression with individuals as the unit of analysis may not yield reliable results due to violations of the assumption of error term independence (Opdenakker & Van Damme, 2000). Multivariate multilevel models provide a more suitable and robust statistical approach for analyzing multiple correlated variables of interest (MacKinnon & Lockwood, 2003; Adegboye & Jawid, 2016).Outcome Variables

The outcome variables are measures of school effectiveness, defined by students' performance in the two components of the BALSA test: literacy (comprising two sub-scales) and numeracy (comprising five to six sub-scales).

Confirmatory Factor Analysis (CFA)

To validate the factor structures of the BALSA test components, we conducted confirmatory factor analysis (CFA) on the sub-scales. We assessed the goodness of fit of the CFA model using the following fit indices: Standardized Root Mean Square Residual (SRMR < 0.08, indicative of good fit), Root Mean Square Error of Approximation ($0.05 \leq$ RMSEA ≤ 0.08, indicative of good fit), Comparative Fit Index (CFI ≥ 0.95, indicative of good fit), and Tucker-Lewis Fit Index (TLI ≥ 0.95, indicative of good fit) (Boateng, Neilands, Frongillo, Melgar-Quiñonez, & Young, 2018; Hu & Bentler, 1999).

Multivariate Analysis

To explore the relationships between demographic variables (such as gender, state, and school location) and the components of the BALSA test, we employed multivariate analysis of variance (MANOVA). This statistical approach allowed us to assess how these demographic factors relate to performance across test’s components.

Furthermore. multilevel multivariate regression models (Cheung, 2013b; Rabe-Hesketh, Skrondal, & Pickles, 2004) was used to analyze the BALSA data. The details of the regression model specification are presented below.

Let Y_{ij} denote the ith (i=1-n_i) score at level 1 from jth (j=1-n_j) pupils at level 2. The model for the multivariate mixed-effects is:

$$Y_{ij} = X_{ij}\beta + Z_{ij}b_j + \epsilon_{ij} \quad (1)$$

where, X_{ij} and Z_{ij} are design matrices for the fixed effects and random effects, respectively; β represents the fixed effects of the independent variables; ϵ_{ij} are level 1 errors, with $\epsilon_{ij} \sim N(0, \Sigma)$. The $b_j's$ are level 2 random errors, with $b_j \sim N(0, D)$. Also, b_j and ϵ_{ij} are independent terms. The random-effects estimate $\hat{b}_j = E(b_j|y_{ij})$ is the empirical Bayes estimate for b_j, which reflects the stochastic variation between individuals.

The two-level SEM-based multivariate mixed models (SEM-MMMs) were formulated using the conditional mean vector and covariance matrix (Cheung, 2013a) as follows:

$$\mu_{ij}(\theta|x_{ij}) = x'_{ij}\beta \quad (2)$$

$$\sum_{ij}(\theta|x_{ij}) = Z_{ij}DZ'_{ij} + \Sigma \quad (3)$$

where θ is the vector of factor loadings (Cheung, 2015).

Both equations 2 and 3 above can be linked through the specification:

$$X_{ij} \,\&\, Z_{ij} \equiv \lambda_{ij},$$

$$\beta \equiv \alpha,$$

$$\Sigma_{ij} \equiv \theta_{ij},$$

$$\mathrm{D}_{ij} \equiv \psi_{ij},$$

where ψ represents the latent covariance matrix, λ is the factor loading representing the magnitude of change in the observed variable Y_{ij} for a unit change in the latent variable η.

The latent variable η has a mean matrix α (α_1 is the mean of the intercept and α_2 is the mean of the slope parameter or the average effect of X_{ij} on Y_{ij}, respectively) and θ is the residual variance-covariance matrix (Adegboye & Jawid, 2016).

All analyses were implemented in R version 3.6.3, SAS version 9.4 and Stata version 16.

Results

Exploratory data analysis

Table S6 presents an overview of the descriptive characteristics of the study sample, comprising 57,473 pupils across 669 schools. The sample is further delineated into 27,102 pupils from Adamawa State and 30,371 from Gombe State. Notably, most (88.34%) of the schools are located in rural areas, and the gender distribution reflects a slightly higher proportion of male pupils (54.15%) than female pupils (45.85%).

Descriptive statistics pertaining to the five subscale scores for Primary 2 (P2) and the six subscale scores for Primary 4 (P4) in both states are presented in Tables 2 and 3, respectively. To calculate these scores, we aggregated the results of all test items assessing each respective subscale. In the case of P2, a noteworthy observation is the significant correlation observed among the five subscale scores in both states ($p < 0.001$). The correlation coefficients, ranging from 0.42 to 0.69, underscore the interrelated nature of these components. Similarly, for P4, the six sub-scale scores exhibited significant correlations ($p < 0.001$), with correlation coefficients ranging from 0.21 to 0.66. These findings emphasize the interconnectedness of the different facets of assessment within each grade level and state.

Confirmatory Factor Analysis (CFA)

The results from the CFA model showing the relationship between each BALSA test item and the subscales are presented in Figures S1 and S2. In these figures, squares represent test items and ovals represent subscales. The numbers between the subscales represent the correlation between subscales, while the numbers by the one-directional arrows between the

sub-scales and the test items represent standardized subscale loading as previously argued by Martin et al. (2020), who maintained that the most observed learner characteristic of adaptive feedback and adaptive navigation were critical factors were critical for the learning processes, especially when deploying and adapting learning technology to assist individual learning and personalized growth.

Numeracy CFA model

In the CFA model for P2 numeracy, three subscales (*basic operations*, *fractions*, and *mensuration*) were significantly correlated ($p < 0.001$). The model fit indices (*CFI*=0.88, *TLI*=0.85, *RMSEA*=0.08, and *SRMR*= 0.06) fall within the acceptable range. Supplementary Figure S1A provides the standardized parameter estimates for the CFA model for P2. Notably, some items (fractional items in the BALSA tool for numeracy) exhibited lower loadings, with the mensuration item NQ10 and fraction test item NQ12 registering the lowest loadings at 0.48 and 0.38, respectively. For the P4 numeracy CFA model, 20 test items were allowed to load four subscales (Supplementary Figure S1B). While the model demonstrates reasonable fit (CFI=0.74, TLI=0.70, RMSEA=0.08, SRMR=0.07), it's noteworthy that the four subscales display high correlations. However, certain items exhibit lower factor loadings (< 0.5), specifically NQ1, NQ2, NQ5, NQ6, NQ11, and NQ16, as detailed in Supplementary Figure S2B.

Literacy CFA model

For P2 literacy, the CFA model encompassed 15 test items (Supplementary Figure S2A). The fit indices for this model, including CFI (0.879), TLI (0.857), RMSEA (0.08), and SRMR (0.061), indicate an appropriate model fit. Standardized subscale loadings generally exceed the 0.5 threshold, although there are some instances of lower loadings in specific test items, notably in LQ1, LQ2, LQ3, LQ5, LQ10, and LQ11. Similarly, the CFA model for P4 literacy, comprising 20 items (Supplementary Figure S2B), illustrates correlations among the

subscales and an appropriate model fit (CFI=0.829, TLI=0.808, RMSEA=0.077, SRMR=0.068). However, as with P2, several standardized subscale loadings fall below 0.5, primarily in test items related to reading ("LQ10") and phonemic awareness (LQ11, LQ12, LQ13).

Effects of demographics

The results from the Multivariate Analysis of Variance (MANOVA) are summarized in Table S7. At the baseline, notable differences were observed in P2 numeracy assessments across various dimensions. There were significant differences between genders. Univariate contrasts indicated that males outperformed females in basic operations, fractions, and mensuration as have been argued in previous studies under a normal educational ecosystem (Conrad, Kablitz & Schumann, 2024; Liu & Sun, 2024; Adeoti, 2017), but which we thought would be disaggregated under a disruptive edu-ecosystem as depicted by the insurgency ravaging the areas. Differences across states were also significant, with Adamawa State exhibiting higher scores than Gombe State. School location had a significant impact, with rural schools outperforming urban schools. For the P4 numeracy assessment, gender differences were observed, with males attaining higher scores in basic operations, mensuration, geometry, and fractions. Significant variations were evident across states, with Adamawa State surpassing Gombe State in scores. The findings showed that the location of schools influenced scores, with rural schools in Adamawa State outperforming urban schools in Gombe State.

Regardingliteracy for P2 pupils, gender differences were significant, with males scoring higher in reading. Variations across states were also observed, with Adamawa State pupils achieved higher scores. Similarly, school location affected performance, with rural schools in Adamawa State outperforming urban schools in Gombe State. No significant difference was observed in phonemic awareness scores. Similarly, for P4 literacy, significant

differences emerged across states' scores, with Adamawa State surpassing Gombe State, and the rural schools in Adamawa State outperforming urban schools in Gombe State. Gender performance did not show significant differences. Overall, while male pupils tended to score higher in reading and lower in phonemic awareness, these gender differences were not statistically significant. However, the influence of state and school location on scores in both numeracy and literacy assessments is evident.

The results from MANOVA are presented in Table S7. At baseline, there were significant differences in P2 numeracy assessments between genders (Wilk's λ= 1.00, $F_{(3)}$ = 4.18, p =0.01, η^2= 0.00), states (λ = 0.97, $F_{(3)}$ = 335.75, p < 0.001, η^2 =0 .03) and location of schools (λ = 0.98, $F_{(3)}$ = 171.31, p < 0.001, η^2 =0 .02). Additionally, univariate contrasts show that males obtained higher scores in *basic operations*, *fractions* and *mensuration* than females. Urban schools and Gombe state had significantly lower scores in *basic operations*, *fractions* and *mensuration* than rural schools and Adamawa State. For P4 numeracy assessment, our results showed significant differences between genders (λ = 0.999, $F_{(4)}$ = 4.337, p =0.002, η^2 = 0.01), states (λ = 0.94, $F_{(4)}$ = 408.02, p < 0.001, η^2 =0 .06) and location of schools (λ = 0.98, $F_{(4)}$ =122.22, p < .001, η^2 = 0.02). More precisely, the univariate contrasts show that on average, males obtained higher scores (p=0.002) in *basic operations*, *mensuration*, *geometry*, and *fractions* than females. Schools in urban locations and Gombe State had significantly lower scores (p < 0.001) in *basic operations*, *mensuration*, *geometry*, and *fractions* than schools in rural locations and Adamawa State.

With regards to literacy among P2, significant differences were found between genders (λ = 1.00, $F_{(2)}$ = 4.84, p =0.008, η^2 = 0.00), states (λ= 0.97, $F_{(2)}$ = 422.57, p < 0.001, η^2 =0 .03) and location of schools (λ = 0.98, $F_{(2)}$ =342.84, p < .001, η^2 = 0.02). Moreover, males obtained higher *reading* scores (p=0.002), but no significant difference was observed in *phonemic awareness* (p=0.069). Schools in urban locations and Gombe State had

significantly lower scores ($p < 0.001$) in *reading* and *phonemic awareness* than schools in rural locations and Adamawa State. Similarly for P4, significant differences were observed between states scores ($\lambda = 0.95$, $F_{(2)} = 646.77$, $p < 0.001$, $\eta^2 = 0.05$) and location of schools ($\lambda = 0.99$, $F_{(2)} = 186.15$, $p < .001$, $\eta^2 = 0.01$); however, gender performance was not significantly different ($\lambda = 1.00$, $F_{(2)} = 0.11$, $p = 0.895$, $\eta^2 = 0.00$). While male pupils had higher *reading* scores and lower *phonemic awareness* scores, this difference was not statistically significant. Schools in urban locations and Gombe state had significantly lower scores ($p < 0.001$) in *reading* and *phonemic awareness* than schools in rural locations and Adamawa State.

Pupil-specific SEM

In order to account for the pupils-specific effect, an SEM-MMM was formulated such that the diagonal elements of BALSA subscales variance-covariance and the factor loadings were freely estimated. The definition variables, X_i (independent variables gender, location of schools, and state), were allowed to load on each latent factor onto fix the model parameters (Figures 1-2). The results from the multilevel (pupils nested within schools) SEM-MMM with correlated errors across literacy and numeracy measurements are presented in Tables 2-3 and depicted in Figures 2-3.

Significant associations were found for gender, location, and state, as well as for *reading* and *phonemic awareness* scores among P2 pupils (Table 4 and Figure 2A). P2 males performed significantly better than females in *reading*, 1.123 (95% *CI*: 0.042; 2.203), p=0.042 and *phonemic awareness,* but not statistically significant (0.973, 95% *CI*: -0.143; 2.090, p=0.087). In addition, large and significant differences were observed in school locations, with pupils in urban areas performing worse in *reading* (-10.965, 95% CI: -15.446; -6.484, p<0.001) and *phonemic awareness* (-7.777, 95% CI: -12.674; -2.880, p=0.002) compared to pupils in rural areas. Adamawa pupils also outperformed Gombe pupils in

reading and *phonemic awareness* by 4.399 (95% CI: 0.565; 8.234), p=0.025 and 16.597 (95% CI: 11.148; 22.046), p<0.001, respectively. For the P4 students, no significant association was detected between literacy subscales (i.e. *reading* and *phonemic* scores) and gender. However, for the location variable, urban pupils reported lower *reading* (-7.170, 95% CI: -12.366; -1.973, p=0.007) and *phonemic awareness* (-5.035, 95% CI: -9.105; -0.965, p=0.015) scores on average than to pupils residing in rural areas. Large and significant effects on literacy subscales by state are also observed, with students in Adamawa State consistently outperforming students in Gombe (Table 4 and Figure 2B).

The numeracy results for P2 and P4 are presented in Table 5 and Figure 3. Similar to Table 4, the effect of gender on numeracy performance is limited. The most significant difference is observed for P4 *mensuration and statistics* (1.441, 95% CI: 0.175; 2.707, p=0.026) and, in some cases, weakly or not statistically significant. Across all numeracy subscales for P2 and P4, urban locations had a substantial adverse effect on performance. For example, the most significant decreases in performance associated with urban areas in *mensuration and statistics* were -9.644 (95% CI: -14.150; -5.138), p<0.001 for P2 and -7.576 (95% CI: -11.814; -3.338), p=0003 for P4, respectively. Except for the *fractions* dimension for P2, the effect of the Adamawa State school system compared to the system in Gombe on pupils' numeracy performance has been significant and large. For example, the effect was 7.573 (95% CI: 3.537; 11.608), p<0.001 for P2 basic operations and 14.019 (95% CI: 10.118; 17.920), p<0.001 for P4 *fractions* subscales. These magnitudes approximately correspond to an additional correct answer to the test questions.

Discussion

The study posited that a greater insight into the differentials in pupils' performance in literacy and numeracy is critical in planning a targeted intervention. The issue of performance

in learning has been identified by various theoretical frames to help define how effective a model performs in ensuring school effectiveness. Performance is seen as a veritable part of the SES model ecosystem. It is instructive to note that the BALSA intervention was first carried out during the post-peak-insurgency era, but with the challenges provided by COVID-19, the intervals of gathering data were also expanded to ensure that the incidences of performances were not a one-off occurrence but an aggregate of the performances of the pupils over the periods. Within this testability period, the BALSA is observed to improve traditional assessments, adapting based on performance. It is very useful in capturing biometric data, and uses immersive simulations in restive areas. However, the defined inhibitions are still very evident regarding accessibility in low-resource areas that require continuous audits of quantum computing issues, as evidenced in conflict-prone areas.

The declining quality of education in Nigerian schools is alarming and raises concerns about the pupils' academic performance, especially at the primary school level. This decline is not unconnected with the lack of attention to perceived factorials identified in Scheerens' model, and amplified by Muhammad & Mohammed (2019) and Nnaemeka et al. (2019) which correlate variedly with the various stakeholders, the subject assessed, the quality of stakeholders' involvement, the environment and the interventions needed in dealing with education in conflict-prone areas. Although there have been several intervention studies in Nigeria, they mainly focused on a particular demographic group, based on a small population, specific location of schools (e.g. rural) or examined the quality of teachers (Adeoti et al., 2017; Moja, 2000; Oyetunde, 2016; Pellens et al., 2016). There has been little focus on the aspects of the subjects that pupils find difficult and the associated factors impacting their performance. The SENSE intervention team, together with stakeholders, co-designed a novel test, BALSA, to assess the performance of pupils in literacy and numeracy skills. Below, we discuss the associated factors underlying the performance of pupils in literacy and numeracy.

Assessment of the instrument

In light of our first research question to assess the reliability and validity of the BALSA test instrument in enhancing school effectiveness, we found significant inter-correlations between literacy and numeracy subscales in this study. While the correlation matrix ranged from moderate to high, it showed empirical predictive validity evidence that the performance of pupils in these subscales was all positively associated. Among the subscales, the average score in *reading* has been consistently high, and the correlation between *reading* and other subscales is among the highest. This result is consistent with previous studies that found a positive relationship between *reading* and *arithmetic* (Akbasli, Sahin, & Yaykiran, 2016; Rutherford-Becker & Vanderwood, 2009), and literacy and numeracy skills (Jordan, Hanich, & Kaplan, 2003; Rutherford-Becker & Vanderwood, 2009). Similarly, we found a correlation between *basic operation* and *fractions* to remain consistently the highest in the two states and both for P2 and P4. The BALSA tool signifies a transformative approach to evaluating essential literacy and numeracy skills in unstable settings. BALSA was able to overcome the shortcomings of traditional assessments that fail to reflect real-time learning conditions during conflict and displacement due to its incorporation of adaptive testing (which modifies question difficulty based on responses) and biometric measures (like eye-tracking). Its use of technology for secure records guarantees data integrity, an essential aspect in regions where educational metrics can be compromised by institutional instability and insurgency, as these areas studied. Nonetheless, the study points out the technological challenges in low-resource settings, highlighting the need for localized adaptations

Regarding our second question on the factor structure of the BALSA instruments in helping educational stakeholders make the school system more effective and strengthened for development, the psychometric properties of the BALSA instrument showed a moderate to a

good fit for all loadings. The significance of this is that within an ecosystem of conflicts and insurgency, deploying such tools guarantees educational progress and deepens learning and engagement, especially when technologies are deployed. Although CFA supported the factor structure of numeracy and literacy subscales for the two primary levels, P2 and P4, respectively, some factor loadings were low (<0.5). For example, in numeracy P4 subscales, the highest and lowest loadings were in *fractions* item NQ9 (0.61) and item NQ2 (0.32), respectively. It is important to reiterate that the 'low' levels and not indicative of poor values for such subscales as in fractions, but very indicative of a subconscious lack of attention to those subscales because they seem to have no socio-cultural and applicable values for the respondents. For instance, the notion of having factional values in discussing monetary or family issues within the study ecosystem seems misplaced as 'one is one, and more is more'. However, factor loadings in P2 are relatively higher, ranging from 0.38 in NQ12 (*fractions*) to 0.72 in NQ4 (*fractions*). Similar lower loadings were recorded in literacy, *reading* item LQ10, *phonemic awareness* items LQ11, LQ12 and LQ13 in Primary 4, suggesting that future BALSA instruments would benefit from modification specifically designed to target those subject areas with low loadings to ensure that only good question items are retained. Consequently, we can argue that the study finds moderate-to-strong correlations between literacy and numeracy subscales (e.g., reading and arithmetic skills), underscoring the interrelation of foundational abilities. This finding supports global data (e.g., Jordan et al., 2003) and contests isolated pedagogical strategies. In areas affected by conflict, where broken education systems often emphasize rote memorization, BALSA's findings support integrated curricula that connect literacy and numeracy. The strong correlation between basic operations and fractions in P4 further emphasizes the need for enhanced conceptual understanding in mathematics, a gap worsened by socio-cultural disconnects (e.g., fractional concepts are often irrelevant in agrarian societies).

Differences in performance

To simultaneously model several dependent variables and allow for correlation within each individual using a set of covariates, a structural equation modelling (SEM)-based on a random-effects model was considered; this strategy is based on mean and variance/ covariance structures. We observed gender differences in the performance of pupils in P2. For example, males performed better in *reading* and *phonemic skills* than females. Even in numeracy, males scored higher than females in all the subscales (except in *fractions*). The reason for this marked difference between genders could be because of the lower enrollment of females in schools due to the effects of insurgency activities (Dauda, 2014; Muhammad & Mohammed, 2019). In Primary 4, there is no significant difference between the genders in literacy skills, although males scored higher than their female counterparts. On the other hand, males performed significantly better than females in numeracy subscales, except in geometry. This further compounds the debates over gender based performances among school children (Obukoadata et al., 2024)

This study observed disparities between urban and rural locations. In terms of site, pupils from rural backgrounds performed better than the pupils in urban areas in both subjects. This is significant for us in that by the general norms and from studies, pupils in urban schools were expected to score higher than pupils in rural schools due to the notion that pupils in rural locations are educationally disadvantaged (Howley & Gunn, 2003; Johnson, Kuhfeld, & Soland, 2021; Nworgu & Nworgu, 2013). However, in some rural areas considered in this study, it was observed that some parents are educated, and they play roles at home in helping their children's performance in terms of giving them moral support and motivation to learn. Additionally, some of these communities have partnered with religious organizations to support teaching children at school. Nonetheless, we are mindful of the near lack of conceptual aggregation of what constitutes urban and rural settlement within the Nigerian sociological

framework. In some instances, lack of certain infrastructure is deemed as the defining point, while in others, population metrics define urban and rural areas despite the presence or absence of such basic infrastructure.

In addition to insurgency activities, kidnappings and farmer-herder violence, other reasons for not being in school include the family's socioeconomic status and preference for traditional Islamic schools (Adeoti et al., 2017; Haggai, 2009). The high number of out-of-school children also has far-reaching consequences, in that if those in school who are the test subjects for this study barely fared fairly well, how well will those who are out of school? Also, those out-of-school children have been argued to provide the firepower and willing tools for an insurgent with which to ravage the communities. Other reasons for being out of school included disruption of education for many students and an inability to attain higher education, especially for girls, because of cultural and religious issues. Therefore, the findings in this study are enormously helpful in designing the subsequent academic interventions for at-risk children with difficulty in literacy and numeracy and making them much more effective for learning. In a recent study, Obukoadata et al (2024) examined the effectiveness of the SENSE-TRI program in improving literacy skills among grade 3 learners in conflict-affected regions of Nigeria during the COVID-19 lockdown. They found that those who participated in the program significantly outperformed those who did not, across various literacy tasks, demonstrating the program's success despite challenging conditions Obukoadata et al. (2024).

Significantly, we assume from the findings that the study uncovers significant disparities in educational outcomes. It shows a snag in the gender gaps during learning under any circumstances. Male students excelled over females in numeracy (e.g., P4 mensuration scores), likely reflecting systemic issues such as lower female enrollment due to risks associated with insurgency and cultural biases. There is also the dysfunctionality and paradox

595 inherent in the urban and rural learning patterns and outcomes (Marzano, 2003). Contrary to
596 worldwide trends, and in support of Rutter (1979) findings, rural students outperformed urban
597 counterparts in literacy and numeracy, due to community support (e.g., educated parents,
598 religious collaboration) and the overcrowded conditions and teacher shortages faced by urban
599 schools following insurgency. These results highlight the necessity for interventions that are
600 sensitive to local contexts, such as gender-sensitive initiatives and infrastructure
601 improvements in urban educational facilities in line with the Scheerens model (Scheerens,
602 1990).

603 This study is consistent with a previous study based on the baseline data on Girl's
604 Education Project (GEP) Phase 3, in which the education data, research and evaluation in
605 Nigeria (EDOREN) highlighted several key challenges faced by the early learning
606 intervention in Northern Nigeria (Pellens et al., 2016). They identified the location of
607 schools, poor infrastructure, quality of teachers (low levels of knowledge and skills amongst
608 teachers), inadequate availability of teaching and learning materials and low level of pupils'
609 knowledge and skills expected at pupils' level (e.g. pre-literacy skills) as major challenges.

610 **Limitations and strengths of this study**

611 We acknowledge the following limitations in this study. Firstly, school-level
612 information (such as student-teacher composition and dropout rates) and student-level (family
613 background) were not collected for this study. Secondly, from the exploratory and confirmatory
614 analyses, the performance of students sampled is poor in all the topics tested in literacy and
615 numeracy. Future study designs should explore student and school-level variables, including
616 the teacher's effect.

617 To the best of our knowledge, this intervention is the first large intervention
618 undertaken in the region that explored the aspects of students' performance in literacy and
619 numeracy. One of the major strengths of this study is the large sample size and the broad

geographical reach of the data. Similarly, the advantages of using the SEM-based random effects models are multi-fold. Firstly, it allows us to get the best of both worlds (Cheung, 2013a; Rovine & Molenaar, 2000); the random effects or latent variables in mixed-models are often interpreted as unobserved heterogeneity at the different levels inducing dependence among all lower-level units in the same higher-level unit. On the other hand, when variables of interest cannot be measured directly, in SEM, a set of items reflecting a hypothetical construct is formed (Rabe-Hesketh et al., 2004). Secondly, the SEM-based random-effects model has the ability to impose and test equality and inequality constraints on the parameters easily (Cheung, 2013b). Thirdly, it facilitates the testing of any error covariance structure (Rovine & Molenaar, 2000), and lastly, it allows the flexibility of model specification.

Conclusion

This study finds relevance within the STEM discourse in several respects. It helped promote efficient learning by underscoring prior limitations to primary education learning within the Pre K-16 circle by suggesting innovative pedagogies and evaluation mechanisms that not only amplify the integrated-inclusive approaches accommodated in the School Effective School (SES) model by Scheerens, but also highlight the imperative of the external touch of effectiveness-mediator geared at activating the intermediate-quality-assurance-mechanism. Secondly, in supporting the postulations of the SES model, it arguably indexed the overwhelming roles of different layers of teachers and teachers' education in bettering the learning experience of the primary school child in Nigeria and, by extension, in most other similar environments.

These findings align with global evidence linking literacy and numeracy outcomes, reinforcing BALSA’s utility as a diagnostic tool in disrupted educational ecosystems. However, the assessment’s innovative features, such as adaptive question difficulty and biometric data integration, face practical hurdles in low-resource settings. While BALSA’s

adaptive engine theoretically personalizes learning, rural schools (88.34% of the sample) often lack the infrastructure for high-definition adaptiveness, risking exclusion of the most vulnerable pupils. Additionally, lower factor loadings in geometry (0.32) and fractions (0.38) suggest cultural misalignment; concepts like fractional arithmetic may hold limited relevance in agrarian communities, skewing performance metrics. This calls into question whether BALSA's design sufficiently adapts to local contextual realities beyond technical adaptability.

The observed performance disparities further complicate its efficacy. Rural pupils outperformed urban counterparts, a counterintuitive finding given urban areas' typically superior resources, partly attributed to community-led educational support and parental involvement in rural settings. Yet, this advantage may be illusory: urban schools in Gombe and Adamawa, still reeling from insurgency-driven displacement, face overcrowding and teacher shortages unaccounted for in the study's pupil-level focus. Similarly, gender gaps favouring males in numeracy likely reflect systemic barriers rather than innate ability, as female enrollment lagged, a legacy of insurgent threats and cultural biases limiting girls' access to education.

BALSA's implications for policy are twofold. Positively, its real-time data on skill gaps, such as P4 geometry scores 22% below national benchmarks, enables targeted interventions, exemplified by the SENSE-TRI program's 34% literacy improvement post-lockdown. Conversely, overreliance on technocentric solutions risks neglecting socio-structural drivers of educational inequity, such as teacher training deficits and familial poverty. Future iterations must balance algorithmic precision with grassroots engagement, investing in rural tech infrastructure while addressing curricular relevance and gender-specific barriers. Ultimately, BALSA's success in northeastern Nigeria hinges on its

669 evolution from a diagnostic tool into a catalyst for holistic, context-sensitive educational
670 recovery.

671 From the foregoing, we recommend that, from the BALSA's detection of weak
672 subscales (e.g., P4 geometry, factor loading = 0.32), reforms in curriculum should
673 emphasizes the contextualization of abstract concepts to local situations. It is also important
674 that, since the study aligns with Scheerens' SES model, promoting interventions that
675 encompass teacher training, school environment, and parent involvement must be
676 consistently driven to improve the outcomes of education. For instance, the SENSE-TRI
677 program's effectiveness in enhancing literacy during COVID-19 lockdowns (Obukoadata et
678 al., 2024) exemplifies the impact of community-inclusive approaches. And, that, the urban-
679 rural performance paradox highlights the pressing need for fair distribution of resources,
680 including technology and teacher training in under-resourced areas. Finally, we reiterate that
681 further intervention targeted at these schools should explore research design that examines
682 students' school-level and teachers' level information and what is needed to make the system
683 more effective. The current study presents valuable information at baseline that will assist in
684 evaluating the impact of more programmes, such as SENSE intervention implementation, on
685 learning outcomes that are geared at making the school system more effective in troubled
686 areas, such as in Northeastern Nigeria.

687 **Abbreviations**

688 *SENSE:* Strengthening Education in the Northeast States; *BALSA:* Basic Level Standard
689 Assessment; *LMM:* Linear Mixed-effects Model; *SEM:* Structural equation modelling
690 *CFA:* Confirmatory factor analysis; *CFI:* Comparative fit index; *RMSEA:* Root mean square
691 error of approximation; *SRMR:* Standardized root mean square residual; *TLI:* Tucker–Lewis
692 index; *MANOVA:* multivariate analysis of variance; *NQ:* Numeracy question item; *LQ:*
693 Literacy question item; *P2*: Primary two (2) pupils; *P4*: Primary four (4) pupils.

Ethical approval

Ethical approval for Strengthening Education in Northeast Nigeria (SENSE) project in Adamawa and Gombe States (PRO-19-09-03) was granted by the Institutional Review Board of the American University of Nigeria with approval code AUN-19-09-03. The respective states' (Adamawa and Gombe State) Ministries of Education provided approval for the study. School principals provided institutional assent. Parents/guardians provided were informed of the purpose of the study. This study was conducted in accordance with the Declaration of Helsinki

Informed consent

Permission was obtained from the school authority since the study involved minors. The IRB waived the requirement for parental permission as the research posed no more than minimal risk; however, parents were informed about the study.

Conflict of interest

Authors declare no competing interests.

Authors' contributions

JSO conceived the study and completed the majority of the writing and data analysis. AL conceived the study and assisted with writing the manuscript; the principal investigator on the SENSE project was responsible for the study's design and execution. AP and PRO assisted with data analysis and writing the manuscript. OAA conceived the study and completed the majority of the data analysis and writing. All authors read and approved the final manuscript.

Availability of data

The datasets generated and/or analysed during this study are not publicly available due to confidentiality, but are available from the SENSE team upon reasonable request.

Acknowledgements

We would like to thank all the teachers and pupils who participated in the SENSE project

References

Adegboye, O. A., & Jawid, A. (2016). Multivariate multilevel models for attitudes toward statistics: multi-disciplinary settings in Afghanistan. Journal of Applied Statistics, 43(1), 244-261.

Adeniran, A., Ishaku, J., & Akanni, L. (2020). Is Nigeria experiencing a learning crisis? Evidence from curriculum-matched learning assessment. Centre for the Study of the Economies of Africa. Working Paper.

Adeoti, J., Onwuemele, A., Obayelu, E., Gitter, S., Rochow, G., Poulsen, L., & Ghubril, A. (2017). Impact Evaluation of UNICEF Nigeria Girls' Education Project Phase 3 (GEP3) Cash Transfer Programme (CTP) in Niger and Sokoto States. Retrieved from Abuja, Nigeria:

Akbasli, S., Sahin, M., & Yaykiran, Z. (2016). The Effect of Reading Comprehension on the Performance in Science and Mathematics. Journal of education and practice, 7(16), 108-121.

Alarape, A. A. (2024). Perceived Effectiveness of Computer-Based Test in Curbing Examination Malpractices and Revitalizing Quality Control by Undergraduates of Federal Universities in Southwestern Nigeria. International Journal of Research and Innovation in Social Science, VIII(V), 2408–2417. https://doi.org/10.47772/IJRISS.2024.805174

Ataseriaba, O. E. H. (2016). Education as a tool for resolving conflicts and enhancing national security in Nigeria. International Journal of Educational Research and Reviews, 4(3), 666–677.

Awoyemi, O. A. (2023). Evaluation of the implementation of basic education minimum standard in IDP camp schools in North Eastern Nigeria. Unilorin Journal of Lifelong Education, 7(2), 124–135.

Barnabas, B., Bavorova, M., & Madaki, M. Y. (2024). The Effect of School Feeding Program on Pupils' Enrolment, Attendance and Performance in Northeastern Region of Nigeria. Sage Open, 14(4), 21582440241255809. https://doi.org/10.1177/21582440241255809

Bertoni, E., Di Maio, M., Molini, V., & Nisticò, R. (2019). Education is forbidden: The effect of the Boko Haram conflict on education in North-East Nigeria. Journal of Development Economics, 141(C). https://doi.org/10.1016/j.jdeveco.2018.06.007

Bilyaminu, M., Iya, I. B., & Purokayo, S. G. (2017). ANALYSIS OF THE IMPACT OF BOKO HARAM INSURGENCY ON EDUCATION IN ADAMAWA STATE, NIGERIA. International Journal of Academic Research and Reflectio, 5(6).

Bissallah, C. E. (2024). Environmental factors as correlates of academic performance among internally displaced pupils in North Central Nigeria. Sokedu Review, 23(1), 14–25.

Chua, Y. P., & Don, Z. M. (2013). Effects of computer-based educational achievement test on test performance and test takers' motivation. Computers in Human Behavior, 29(5), 1889–1895. https://doi.org/10.1016/j.chb.2013.03.008

Cheung, M. W.-L. (2013a). Implementing restricted maximum likelihood estimation in structural equation models. Structural equation modeling: a multidisciplinary journal, 20(1), 157-167.

Cheung, M. W.-L. (2013b). Multivariate meta-analysis as structural equation models. Structural equation modeling: a multidisciplinary journal, 20(3), 429-454.

Cheung, M. W.-L. (2015). metaSEM: An R package for meta-analysis using structural equation modeling. Frontiers in psychology, 5, 1521.

Dauda, M. (2014). The effect of Boko haram crisis on socioeconomic activities in Yobe state. The International Journal of Social Sciences and Humanities Invention, 1(4).

Donnelly, E. (2026). Doctoral Candidate and Graduate Teaching Assistant, Department of Politics and International Studies, SOAS.

Ensign, M. (2016). Local action to protect communities in Nigeria. Forced Migration Review.

Fullan, M. (2013). Stratosphere: Integrating technology, pedagogy, and change knowledge. Alberta Journal of Educational Research, 62(4), 429-432.

Haggai, M. P. (2009). Basic Education and Education for All in Nigeria: The Role of Multinational Corporations.

Howley, C. B., & Gunn, E. (2003). Research about mathematics achievement in the rural circumstance. Journal of Research in Rural Education, 18(2), 86-95.

Humphreys, S., Moses, D., Kaibo, J., & Dunne, M. (2015). Counted in and being out: Fluctuations in primary school and classroom attendance in northern Nigeria. International Journal of Educational Development, 44, 134–143. https://doi.org/10.1016/j.ijedudev.2015.08.004

Johnson, A., Kuhfeld, M., & Soland, J. (2021). The Forgotten 20%: Achievement and Growth in Rural Schools Across the Nation. AERA Open, 7, 23328584211052046.

Jordan, N. C., Hanich, L. B., & Kaplan, D. (2003). A longitudinal study of mathematical competencies in children with specific mathematics difficulties versus children with comorbid mathematics and reading difficulties. Child development, 74(3), 834-850.

Keck, K. (2016). ESSPIN Composite Survey 3: Kano State Report. Retrieved from

Kyriakides, L. (2005). Extending the comprehensive model of educational effectiveness by an empirical investigation. School Effectiveness and School Improvement, 16(2), 103-152.

Kyriakides, L., & Creemers, B. P. (2008). A longitudinal study on the stability over time of school and teacher effects on student outcomes. Oxford review of education, 34(5), 521-545.

Martens, M., Hajibayova, L., Campana, K., Rinnert, G. C., Caniglia, J., Bakori, I. G., . . . Oh, O. J. (2020). Being on the wrong side of the digital divide": seeking technological interventions for education in Northeast Nigeria. Aslib Journal of Information Management.

Marzano, R. J. (2003). What works in schools: Translating research into action: Ascd.

Moja, T. (2000). Nigeria education sector analysis: An analytical synthesis of performance and main issues. World Bank Report, 3, 46-56.

Muhammad, F. D., & Mohammed, A. (2019). Analysis Of The Effects Of Insurgency On Education In North Eastern Nigeria; A Case Study Of Yobe State Of Nigeria.

Nggada, M. D., Istifanus, B., Ebirere, O. R., Suleiman, M., Abubakar, M., & Faruq, H. A. (2025). Assessment of the needs of Internally Displaced Persons (IDPS) on Health Care and Educational Needs: A Case Study of Borno State, Nigeria. International Journal of Education Effectiveness Research. https://doi.org/10.70382/hijeer.v08i8.029

Nnaemeka, A. N., Waba Simon,, Aliyu Adamu. (2019). Analysis of the impact of boko-haram insurgency on public institutions in Bama local government area of Borno state, Nigeria. North Asian International Research Journal of Social Science & Humanities, 5(6).

Nwoke, C., Oyiga, S., & Cochrane, L. (2024). Assessing the phenomenon of out-of-school children in Nigeria: Issues, gaps and recommendations. Review of Education, 12(3), e70011.

Nworgu, B. G., & Nworgu, L. N. (2013). Urban-rural disparities in achievement at the basic education level: The plight of the rural child in a developing country. Developing Country Studies, 3 (14), 128–140. Downloaded on 30th August, 2017 from.

Obukoadata, P. R., Hammler, K., Yusuf, H., Liman, A., & Olumoh, J. S. (2024). Transacting knowledge when there are no schools during the COVID-19 lockdown in Nigeria: the SENSE-transactional radio instruction experience. Humanities and Social Sciences Communications, 11(1), 1-12.

Ogbonna, S. (2016). Education Sector Support Programme in Nigeria (ESSPIN) Assignment Report.

Oladunjoye, P., & Omemu, F. (2013). Effect of Boko Haram on school attendance in Northern Nigeria. British Journal of Education, 1(2), 1–9.

Oyetunde, T. O. (2016). Improving literacy instructional practices in primary schools in Nigeria: Strategies that work. Literacy Information and Computer Education Journal, 6(2), 2323-2328.

Pellens, T., Outhred, R., Majeed, Z., Kveder, A., Binci, M., Wallin, J., . . . Rai, S. (2016). Evaluation of UNICEF Girls' Education Project Phase 3 (GEP3). Baseline Synthesis Report', prepared by EDOREN on behalf of UNICEF GEP3.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). Generalized multilevel structural equation modeling. Psychometrika, 69(2), 167-190.

Rinnert, G. C., Campana, K., Martens, M., Mupinga, D., Caniglia, J., Malgwi, G., . . . Filippi, A. (2020). Synergy through Making: Co-designing with Educational Stakeholders in Northeastern Nigeria.

Rovine, M. J., & Molenaar, P. C. (2000). A structural modeling approach to a multilevel random coefficients model. Multivariate Behavioral Research, 35(1), 51-88.

Rutherford-Becker, K. J., & Vanderwood, M. L. (2009). Evaluation of the relationship between literacy and mathematics skills as assessed by curriculum-based measures. The California School Psychologist, 14(1), 23-34.

Rutter, M. (1979). Fifteen thousand hours: Secondary schools and their effects on children: Harvard University Press.

Rwang, V. K. (2023). Quality basic education provision for internally displaced children in internally displaced persons' camps. A case study of some camps in Northcentral and Northeast of Nigeria. [Phd, University Library, University of Southampton]. https://eprints.soton.ac.uk/477009/

Scheerens, J. (1990). School effectiveness research and the development of process indicators of school functioning. School Effectiveness and School Improvement, 1(1), 61-80.

Sola, L., Chen, Y., & Subrahmanian, V. S. (2025). A quantitative geospatial analysis of the risk that Boko Haram will target a school. PLOS One, 20(6), e0320939. https://doi.org/10.1371/journal.pone.0320939

Udefuna, P. N., Ezenwajiobi, C. C., & Udom, M. S. (2025). Addressing the crisis of out-of-school children in Nigeria and analysing government interventions so far: Areas for legislative intervention.

UNESCO. (2022). New estimation confirms out-of-school population is growing in sub-Saharan Africa. Global education monitoring report: policy paper, 48 and fact sheet 62.

UNICEF. (2017). Education. Retrieved from https://www.unicef.org/nigeria/education

Zhang, L., Basham, J. D., & Yang, S. (2020). Understanding the implementation of personalized learning: A research synthesis. Educational Research Review, 31, 100339. https://doi.org/10.1016/j.edurev.2020.100339

Table 1. Scheerens' SES model application to BALSA/SENSE study

Level	Key factors	Relevance to this study
Input	Teacher training Resources Leadership	Northeast teacher Shortages post-imsurgency
Process	Curriculum delivery School climate Time-on-task	BALSA measures Skills delivery (RQ1/RQ2)
Output	Pupil achievement (literacy/numeracy)	Factor associations with BALSA scores (RQ3/RQ4)

Table 2: Descriptive statistics of the BALSA test instrument subscale scores and their intercorrelation (Primary 2)

Subscales (Items)	Number of items	1	2	3	4	5
Adamawa P2						
1.Reading (LQ1-LQ10)	10					
2.Phonemic Awareness (LQ 11-LQ15)	5	0.60				
3.Basic Operations (NQ2, NQ5-NQ9)	6	0.63	0.56			
4.Fractions (NQ, NQ3 ,NQ4, NQ12)	4	0.60	0.53	0.69		
5.Mensuration and Statistics (NQ10-NQ15)	5	0.52	0.55	0.61	0.54	
Mean ± SE		5.712±0.03	2.63±0.02	3.25±0.01	2.13±0.01	2.27±0.02
Gombe P2						
1.Reading (LQ1-LQ10)	10					
2.Phonemic Awareness (LQ 11-LQ15)	5	0.49				
3.Basic Operations (NQ2, NQ5-NQ9)	6	0.62	0.47			
4.Fractions (NQ, NQ3 ,NQ4, NQ12)	4	0.58	0.42	0.67		
5.Mensuration and Statistics (NQ10-NQ15)	5	0.45	0.55	0.55	0.47	
Mean ± SE		5.12±0.02	1.96±0.01	2.84±0.02	2.05±0.01	1.65±0.01

Table 3: Descriptive statistics of the BALSA test instrument subscale scores and their intercorrelation (Primary 4)

Subscales	Number of items	1	2	3	4	5	6
Adamawa Grade 4							
1.Reading (LQ 1-LQ10)	10						
2.Phonemic Awareness (LQ 11-LQ20)	10	0.57					
3.Basic Operations (NQ3-NQ5, NQ15, NQ17, NQ18)	6	0.57	0.56				
4.Fractions (NQ1, NQ2, NQ8-NQ10, NQ12)	6	0.55	0.53	0.66			
5.Mensuration and Statistics (NQ7, NQ6, NQ19, NQ20)	4	0.5	0.51	0.64	0.57		
6.Geometry (NQ11, NQ13, NQ14, NQ16)	4	0.36	0.37	0.5	0.48	0.44	
Mean ± SE		5.80±0.03	5.46±0.03	3.06±0.02	3.32±0.02	2.06±0.01	1.45±0.01
Gombe Grade 4							
1.Reading (LQ 1-LQ10)	10						
2.Phonemic Awareness (LQ 11-LQ20)	10	0.34					
3.Basic Operations (NQ3-NQ5, NQ15, NQ17, NQ18)	6	0.48	0.41				
4.Fractions (NQ1, NQ2, NQ8-NQ10, NQ12)	6	0.49	0.39	0.59			

5.Mensuration and Statistics (NQ7, NQ6, NQ19, NQ20)	4	0.35	0.44	0.57	0.52		
6.Geometry (NQ11, NQ13, NQ14, NQ16)	4	0.21	0.39	0.43	0.45	0.42	
Mean ± SE		5.24±0.03	4.11±0.03	2.30±0.01	2.56±0.01	1.48±0.01	1.06±0.01

Table 4. SEMs with correlated errors across literacy measurements

	Estimate	95% CI		P-value
		Lower	Upper	
	Primary 2			
Reading				
Male (vs Female)	1.123	0.042	2.203	0.042
Urban (vs Rural)	-10.965	-15.446	-6.484	<0.001
Adamawa (vs Gombe)	4.399	0.565	8.234	0.025
Phonemic awareness				
Male (vs Female)	0.973	-0.143	2.09	0.087
Urban (vs Rural)	-7.777	-12.674	-2.88	0.002
Adamawa (vs Gombe)	16.597	11.148	22.046	<0.001
	Primary 4			
Reading				
Male (vs Female)	-0.092	-1.221	1.036	0.873
Urban (vs Rural)	-7.17	-12.366	-1.973	0.007
Adamawa (vs Gombe)	4.928	1.064	8.792	0.012
Phonemic awareness				
Male (vs Female)	0.57	-0.327	1.467	0.213
Urban (vs Rural)	-5.035	-9.105	-0.965	0.015
Adamawa (vs Gombe)	12.51	8.71	16.309	<0.001

Table 5. SEMs with Correlated Errors Across Numeracy Measurement

	Estimate	95% CI		P-value
		Lower	Upper	
	Mensuration (P2)			
Mensuration (P2)				
Male (vs Female)	1.221	-0.06	2.501	0.062
Urban (vs Rural)	-9.644	-14.15	-5.138	<0.001
Adamawa (vs Gombe)	11.165	7.339	14.991	<0.001
Fractions				
Male (vs Female)	0.753	-0.173	1.68	0.111
Urban (vs Rural)	-4.626	-9.471	0.219	0.061
Adamawa (vs Gombe)	3.054	-0.749	6.857	0.116
M1[schoolname]	-3E+06	-3E+06	-3E+06	<0.001
Basic Operation				
Male (vs Female)	1.211	0.246	2.177	0.014
Urban (vs Rural)	-7.527	-12.466	-2.589	0.003
Adamawa (vs Gombe)	7.573	3.537	11.608	<0.001
	P4			
Male (vs Female)	1.441	0.175	2.707	0.026
Urban (vs Rural)	-7.576	-11.814	-3.338	<0.001
Adamawa (vs Gombe)	13.849	10.233	17.465	<0.001
Fractions				
Male (vs Female)	0.835	0.058	1.612	0.035
Urban (vs Rural)	-7.011	-10.803	-3.219	<0.001

Adamawa (vs Gombe)	14.019	10.118	17.92	<0.001
M1[schoolname]	-3E+06	-3E+06	-3E+06	<0.001
Basic Operation				
Male (vs Female)	1.355	0.314	2.396	0.011
Urban (vs Rural)	-5.839	-9.972	-1.706	0.006
Adamawa (vs Gombe)	13.97	10.005	17.935	<0.001
M1[schoolname]	-3E+06	-3E+06	-3E+06	<0.001
Geometry				
Male (vs Female)	0.892	-0.163	1.946	0.097
Urban (vs Rural)	-4.786	-8.436	-1.137	0.01
Adamawa (vs Gombe)	10.027	6.824	13.231	<0.001

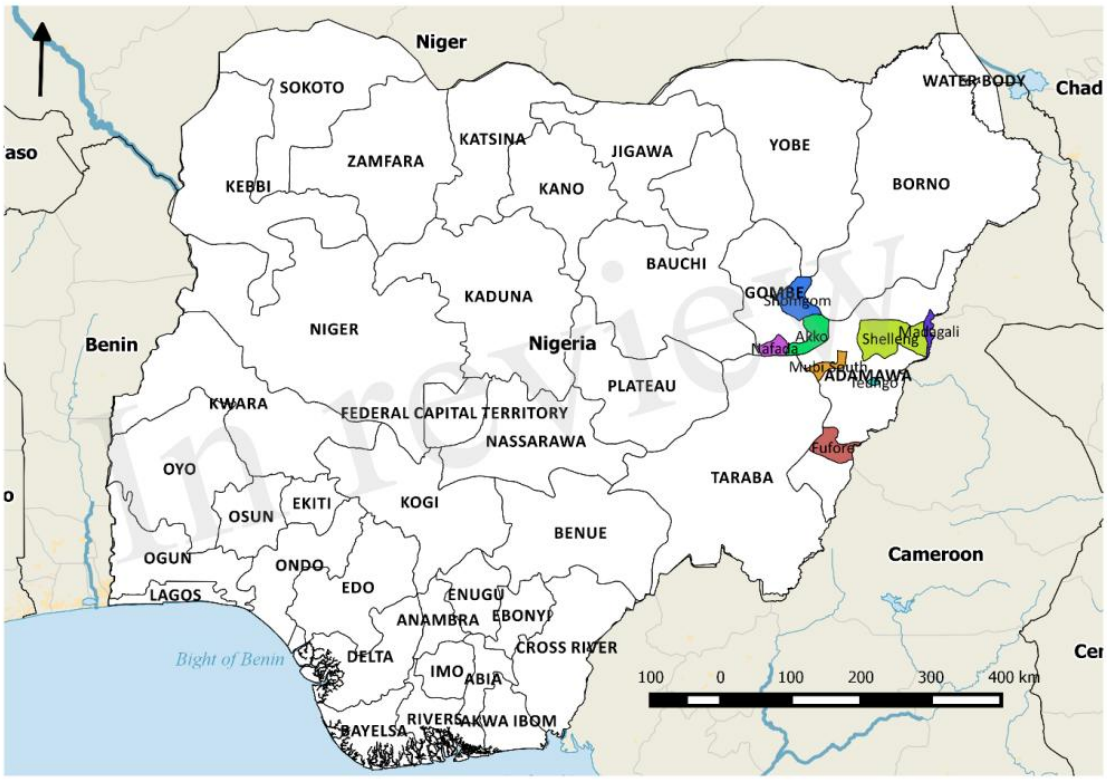
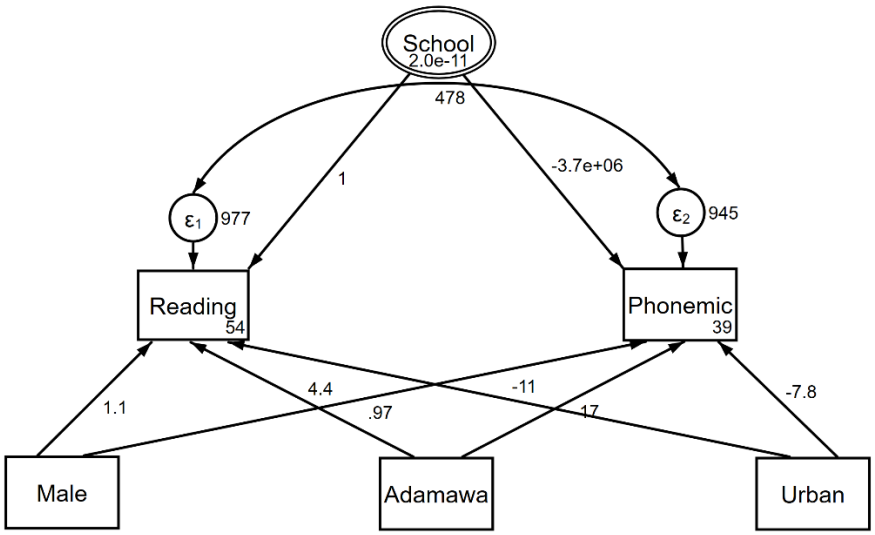


Figure 1. Map of Nigeria showing the SENSE phase 1 study region

(A)



(B)

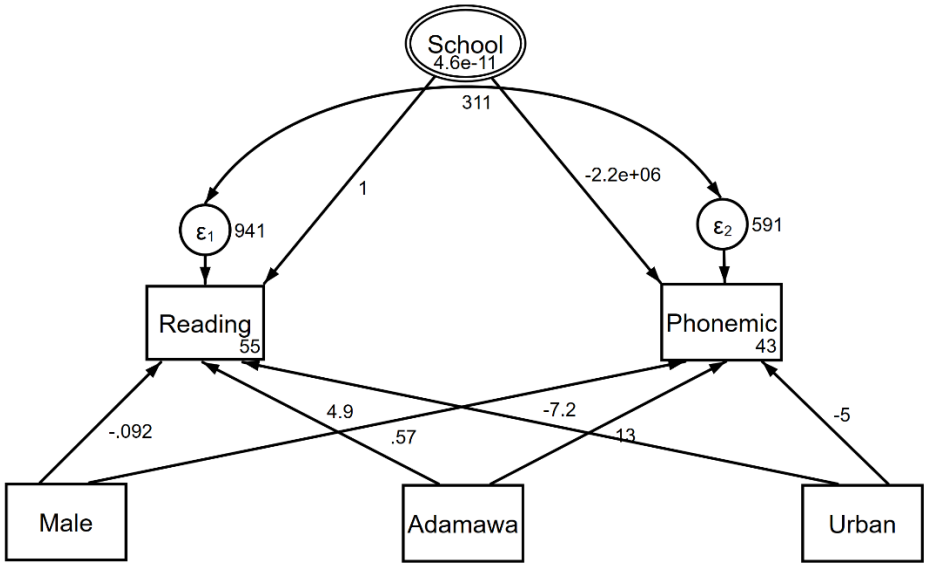
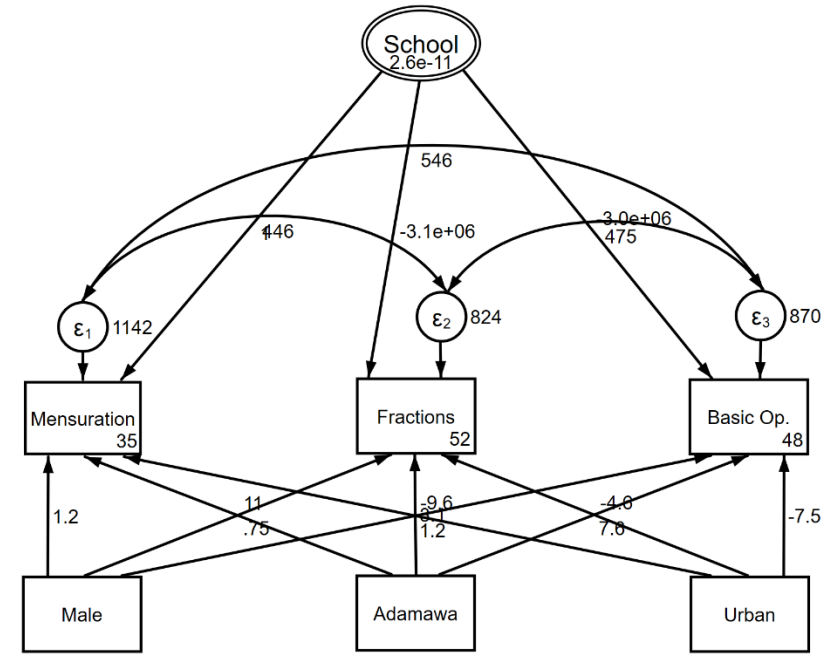


Figure 2. Multilevel Structural Equation Model for performance in literacy, (A) Primary 2, (B) Primary 4.

(A)



(B)

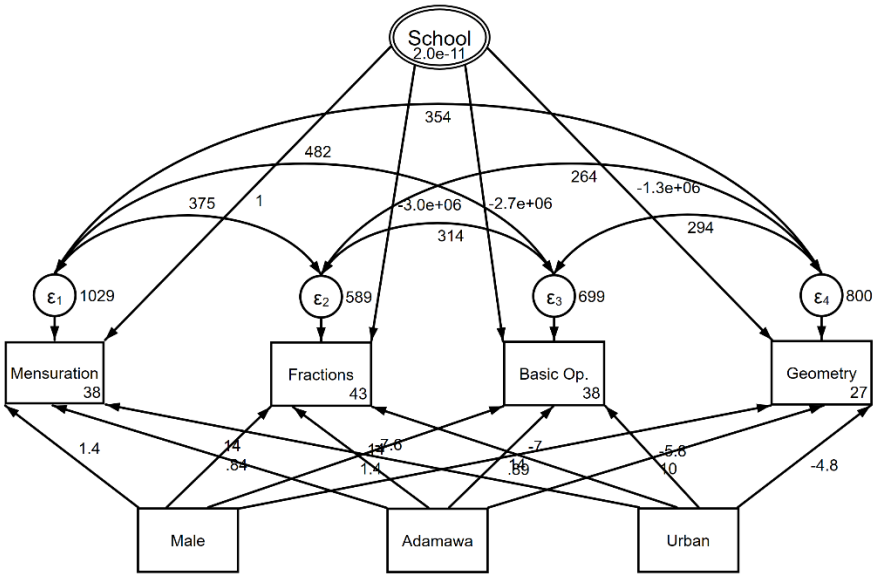


Figure 3. Multilevel Structural Equation Model for performance in Numeracy, (A) Primary 2, (B) Primary 4.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SUPPLEMENTARYMATERIALS.docx