Econometric Analysis of Gender Disparities in STEM Enrollment and Performance in Sub-Saharan African Secondary Schools

Saina Philiph Kipkosgei
Department of Applied Mathematics, University of Eldoret, P.O. Box 1125-30100, Eldoret, Kenya
Correspondence: sainamailbox@gmail.com; ORCID: 0000-0001-8543-9976

Abstract

Gender disparities in Science, Technology, Engineering, and Mathematics (STEM) remain among the most entrenched structural inequalities in Sub-Saharan African (SSA) secondary education, with far-reaching consequences for human capital development and inclusive economic growth. Despite incremental progress in achieving gender parity at the primary level, significant participation and performance gaps re-emerge and intensify within STEM tracks at the secondary level, constraining the region's long-term development trajectory. This study presents a rigorous, multi-method econometric analysis of gender gaps in STEM enrollment and examination performance across five SSA countries—Kenya, Tanzania, Uganda, Ghana, and Nigeria—using a pooled dataset of 1,423,807 upper secondary students spanning 2018 to 2023. The analytical framework integrates binary logistic regression and probit models to estimate the probability of STEM enrollment as a function of gender and a comprehensive set of socioeconomic and institutional covariates, alongside the Oaxaca-Blinder decomposition technique to partition the gender performance gap in national examinations into explained (endowment) and unexplained (structural) components. Two-way panel fixed-effects models with country and year controls provide additional robustness. Descriptive results confirm that female students are 23.4 percentage points less likely to enroll in STEM tracks than their male counterparts. Logistic regression analysis yields a fully controlled log-odds estimate for the female indicator of −0.812 (Average Marginal Effect = −18.4 percentage points; $p < 0.001$). Oaxaca-Blinder decomposition reveals that 38.6% of the observed performance gap of 0.41 standard deviations is attributable to measurable endowment differences, while the remaining 61.4% is structurally unexplained, implicating stereotype threat, implicit teacher biases, gendered curricula, and cultural norms. The share of female STEM teachers emerges as a significant moderator of female performance. The study concludes with evidence-based policy recommendations directed at teacher recruitment and gender-responsive training, curriculum reform, conditional cash transfer programmes targeting STEM-enrolled girls, and equitable STEM infrastructure investment across rural and urban schools.

Keywords: *gender disparities; STEM education; Sub-Saharan Africa; logistic regression; probit model; Oaxaca-Blinder decomposition; secondary education; gender gap; human capital; educational inequality*

1. Introduction

The underrepresentation of girls and women in Science, Technology, Engineering, and Mathematics (STEM) disciplines remains one of the most structurally persistent educational challenges in Sub-Saharan Africa (SSA). Although the region has made measurable strides towards gender parity in primary school enrolment over the past two decades, these gains erode sharply as students transition to secondary education and subject specialisation becomes compulsory. Gender-differentiated tracking into STEM streams at this critical juncture has consequences that extend far beyond the educational system: they reproduce occupational segregation, constrain household income diversification, and limit the region's capacity to generate the technical human capital necessary for sustainable industrialisation and digital transformation [1], [2], [3].

Sub-Saharan Africa confronts a dual developmental imperative that makes this issue especially acute. On one hand, the region's demographic structure—characterised by a rapidly expanding working-age population—presents an unprecedented economic opportunity, provided that human capital accumulation is broad-based and inclusive [4]. On the other hand, the accelerating diffusion of automation, artificial intelligence, and digital production technologies is reshaping global labour markets in ways that systematically reward STEM competencies and penalise their absence [5]. The systematic exclusion of half the potential labour force from these pathways therefore constitutes not merely an equity failure but a macroeconomic inefficiency with compounding long-run costs.

The existing body of literature has documented gender gaps in STEM participation across SSA through qualitative case studies, small-scale surveys, and cross-country descriptive reports [6], [7]. However, rigorous econometric analyses that exploit nationally representative examination microdata, apply decomposition techniques to disentangle the structural sources of these gaps, and leverage panel variation to account for country-level heterogeneity remain conspicuously scarce [8]. This analytical deficit is consequential: policy prescriptions formulated without a granular, causally informed understanding of the mechanisms driving gender disparities are unlikely to be well-targeted or cost-effective.

This study addresses this gap by making four substantive contributions to the quantitative literature on gender and STEM education in SSA. First, it applies binary logistic regression and probit models to a large, multi-country dataset to estimate the determinants of STEM enrollment, reporting Average Marginal Effects with standard errors clustered at the school level. Second, it employs the Oaxaca-Blinder decomposition methodology—adapted from its origins in labour economics wage gap analysis—to partition the observed gender performance gap in national STEM examinations into explained endowment differences and unexplained structural components. Third, it exploits a six-year panel structure covering five countries (2018-2023) to conduct two-way fixed-effects estimation that absorbs time-invariant country heterogeneity. Fourth, it translates these empirical findings into actionable, context-calibrated policy recommendations anchored in the institutional realities of SSA education systems [9], [10].

The remainder of this paper is organised as follows. Section 2 reviews the theoretical frameworks and empirical evidence on gender disparities in STEM education in SSA. Section 3 presents the data sources, variable definitions, and estimation strategies. Section 4 reports the empirical results. Section 5 discusses the implications of findings within the broader literature. Section 6 synthesises the policy implications. Section 7 concludes.

2. Literature Review

2.1 Theoretical Foundations

Several distinct but complementary theoretical frameworks inform the analysis of gender disparities in STEM education. Human capital theory, in the tradition of Schultz [11] and Becker [9], conceptualises education as an investment decision in which households allocate resources across children based on expected future returns. When structural labour market discrimination reduces the realised wage premium to STEM qualifications for women relative to men, rational investment calculus will produce systematically lower household commitment to female STEM education—even in the absence of explicit gender preferences. This demand-suppression mechanism is compounded by credit constraints faced by poor households, which force prioritisation of male children's education as a lower-risk investment [21].

Gender socialisation theory complements this economic framing by foregrounding the role of cultural norms, gendered social scripts, and institutional practices in shaping adolescent girls' occupational aspirations and subject choices [12]. From early childhood, girls in many SSA contexts are exposed to domestic role scripts that devalue analytical and technical competencies as inappropriate or unnecessary for women. These socialised expectations are internalised and reinforced through interactions with parents, teachers, and peers, creating a powerful normative barrier to STEM participation that operates independently of household resources.

Steele and Aronson's [13] stereotype threat theory provides a psychological mechanism through which internalised gender stereotypes about mathematical ability translate directly into measurable performance deficits. The key insight is that when female students are made aware—implicitly or explicitly—of prevailing cultural stereotypes about women and mathematics, cognitive resources are redirected from task engagement towards managing evaluation anxiety, producing systematic underperformance relative to actual ability. Experimental evidence suggests this mechanism is particularly potent in high-stakes examination contexts, precisely the settings used to measure STEM achievement in this study [13].

Eccles et al.'s expectancy-value model [14] provides a further psychological micro-foundation for subject choice. The model posits that educational decisions are jointly determined by students' expectancies of success in a given domain and their subjective valuation of that domain. Gendered expectancies—shaped by stereotype threat, teacher feedback, and parental attitudes—systematically depress female students' confidence in STEM competence, while gendered cultural values reduce the perceived utility of STEM engagement for their projected life trajectories. Together, these forces produce the enrollment and performance patterns observed in this study.

2.2 Empirical Evidence from Sub-Saharan Africa

UNESCO Institute for Statistics data indicate that the Gender Parity Index (GPI) for upper secondary enrolment across SSA stood at 0.84 in 2022, reflecting a persistent aggregate deficit in female participation [15]. However, this headline figure masks substantial heterogeneity: the GPI ranges from 0.52 in Niger to 0.97 in Rwanda, and the gap is systematically larger in STEM-specific tracks than in overall secondary enrolment [16], [17]. Across the region, girls constitute approximately 34% of upper secondary students enrolled in science and technical streams [16], a proportion that has remained stubbornly stable despite improvements in overall enrolment parity.

Country-level examination data corroborate these enrolment patterns. Analyses of Kenya's Kenya Certificate of Secondary Education (KCSE) results consistently document male advantages in mathematics and physical sciences, with performance gaps widening substantially at the upper tail of the score distribution [18]. This upper-tail divergence is particularly consequential because entry into tertiary STEM programmes is typically determined by scores in precisely this range. Tanzania's National Examinations Council data reveal analogous patterns, with girls' relative performance deficits most pronounced in physics and chemistry [19]. In Ghana, West Africa Senior School Certificate Examination data document a 14.3 percentage point gender gap in credit-level passes in science subjects [20].

The literature identifies demand-side and supply-side drivers operating through distinct channels. On the demand side, household poverty constrains investment in female secondary education through direct cost and opportunity cost effects [21]; geographic distance to secondary schools with qualified science teachers creates a spatial barrier that falls disproportionately on girls due to safety concerns about long travel distances [22]; early marriage and adolescent pregnancy interrupt schooling trajectories at rates significantly higher for girls than boys [23]; and the scarcity of visible female STEM role models limits the formation of aspirational identities among female students [24]. On the supply side, inadequate laboratory infrastructure, male-normative science textbook content, and gender-insensitive pedagogical practices collectively reproduce an unwelcoming environment for female STEM learners [25], [26].

2.3 Econometric Literature and Identified Research Gaps

The application of discrete choice models to educational participation decisions in SSA has expanded considerably in recent years [27]. Binary logistic regression and probit specifications have been used to model school enrolment, subject selection, and examination entry, with findings broadly confirming the negative effect of female gender on STEM-related choices. However, many existing studies rely on school-level or district-level aggregates rather than individual-level microdata, limiting precision and introducing ecological inference problems [28]. Studies with individual-level data frequently lack the rich household covariate information needed to separate gender effects from correlated socioeconomic confounders.

The Oaxaca-Blinder decomposition technique, developed originally to analyse racial and gender wage gaps in labour economics [29], [30], has been adapted to educational contexts to decompose gaps in test scores, attainment, and enrolment rates. Applications in SSA contexts remain sparse, and existing decomposition studies typically focus on overall literacy or numeracy gaps rather than STEM-specific performance in nationally standardised examinations [8]. This study fills that lacuna by applying the decomposition to national STEM examination scores, identifying the relative contributions of endowment differences and structural disadvantages to the observed gender performance gap.

Panel econometric approaches that control for country and time fixed effects have been underutilised in the SSA educational gender gap literature relative to their prevalence in wage gap research [31]. The cross-country heterogeneity documented in the preceding review highlights the potential for omitted country-level variables—including colonial institutional legacies, cultural attitudes towards female education, and national education financing systems—to confound cross-sectional estimates. The two-way fixed-effects specification employed in this study addresses this concern while preserving the efficiency advantages of pooling data across countries and years.

3. Research Methodology

3.1 Research Design

This study adopts a quantitative, explanatory cross-national panel design. The core methodological logic is to exploit variation in gender composition across students, schools, countries, and years to estimate the causal mechanisms through which gender shapes STEM educational outcomes. The research design is structured around three analytically coherent and sequentially related modules: (i) descriptive statistical characterisation of gender-disaggregated STEM enrolment and performance indicators, including trend analysis; (ii) binary discrete choice modelling through logistic regression and probit specifications to estimate the probability of STEM enrolment as a function of gender and a progressive sequence of covariate sets; and (iii) Oaxaca-Blinder decomposition of the gender performance gap in standardised national STEM examination scores. Two-way fixed-effects panel estimation serves as a robustness check across all modules [28], [31].

3.2 Data Sources and Sample Construction

Three primary data sources are integrated in this study.

National Secondary School Examination Microdata. Individual-level examination records were obtained from five national examination bodies: the Kenya National Examinations Council (KNEC) for Kenya Certificate of Secondary Education (KCSE) results; the Tanzania National Examinations Council (NECTA) for Certificate of Secondary Education Examination (CSEE) results; the Uganda National Examinations Board (UNEB) for Uganda Certificate of Education (UCE) results; and the West African Examinations Council (WAEC) for West Africa Senior School Certificate Examination (WASSCE) results in both Ghana and Nigeria. Each dataset contains individual candidate records including: candidate gender; school identifier and school-level administrative data; subject enrolment; raw scores and grade classifications in each STEM subject; and school type (government/private) and district. Following data cleaning, deduplication, and exclusion of records with missing gender information, the merged dataset comprises 2,847,614 individual examination records across five countries and six years (2018-2023).

UNESCO Institute for Statistics Data

Country-year level data on gender-disaggregated gross enrolment ratios, Gender Parity Indices, trained teacher rates by gender, pupil-to-teacher ratios, and public expenditure on education as a share of GDP were obtained from the UIS Data Centre [15]. These country-year aggregates are merged to the examination microdata by country and year to provide contextual covariates for the panel fixed-effects specifications and to enable macro-level cross-validation of enrollment patterns.

Demographic and Health Survey (DHS) Data

Household-level covariates—including the parental education composite (highest level attained by either parent), household wealth index quintile, and rural-urban residential classification—are derived from the most recent DHS rounds for each country [32]. These data are merged to the examination microdata using district (or equivalent administrative unit) identifiers, assigning district-level median household characteristics to individual students. This district-level linkage introduces measurement error in individual-level household characteristics, a limitation explicitly acknowledged in Section 5.3.

The analytical sample for the logistic and probit models comprises 1,423,807 upper secondary students (Forms 3-4 or their national equivalents) with non-missing values on all covariates. The sample excludes Forms 1-2 students, as subject tracking and STEM-specific choice are most relevant in the upper secondary years when subject selection has been finalised. The STEM enrolment indicator is operationalised as enrolment in a minimum of three of the following core subjects: mathematics, physics, chemistry, biology, and technical or applied mathematics.

3.3 Variable Definitions and Operationalization

Table 1. *Variable Definitions, Operationalisation, and Data Sources*

Variable Name	Type	Operationalisation	Source
STEM_Enroll	Dependent (Binary)	1 = enrolled in ≥3 STEM subjects; 0 = otherwise	KNEC / NECTA / UNEB / WAEC
STEM_Score	Dependent (Continuous)	Standardised aggregate STEM examination score (μ=0, σ=1)	KNEC / NECTA / UNEB / WAEC
Female	Key Independent	1 = female candidate; 0 = male candidate	Examination records
ParentEdu	Control	Highest parental education: 0=none, 1=primary, 2=lower secondary, 3=upper secondary, 4=tertiary	DHS
WealthQuint	Control	Household wealth index quintile: 1 = poorest to 5 = wealthiest	DHS
Urban	Control	1 = school located in urban area; 0 = rural	Ministry of Education records
PublicSchool	Control	1 = government-funded school; 0 = private or faith-based	Ministry of Education records
FemaleTchrRatio	Control	Proportion of school STEM teachers who are female (0–1 continuous)	UIS / National MoE data
PupilTchrRatio	Control	Number of enrolled students per qualified STEM teacher	UIS / National MoE data
CountryFE	Fixed Effect	Country binary indicators; Kenya = reference category	Derived
YearFE	Fixed Effect	Examination year binary indicators;	Derived

Variable Name	Type	Operationalisation	Source
		2018 = reference year	

Note. Variable definitions are consistent across all five countries. Continuous variables are standardised before entry into regression models. District-level DHS medians are used for household covariates. Source: Author's compilation.

3.4 Binary Logistic Regression and Probit Specifications

Both binary logistic regression and probit models share a latent variable conceptualisation in which the observed binary enrolment decision reflects an underlying continuous propensity that crosses a threshold [27]. Let Y^* denote this latent propensity, with $Y = 1$ observed if $Y^* > 0$. The systematic component is:

$$Y_i^* = X_i\beta + \varepsilon_i$$

Under the logistic specification, ε follows a standard logistic distribution, yielding:

$$P(Y = 1 \mid X) = \Lambda(X\beta) = \exp(X\beta)/[1 + \exp(X\beta))]$$

Under the probit specification, ε follows a standard normal distribution, yielding:

$$P(Y = 1 \mid X) = \Phi(X\beta) = \int_{\infty}^{X\beta} \varphi(t)\ dt$$

where Λ(·) is the logistic cumulative distribution function, Φ(·) is the standard normal cumulative distribution function, and φ(·) is the standard normal probability density function. Both models are estimated by maximum likelihood. Since the log-odds coefficients from the logistic model and the index function coefficients from the probit model are not directly interpretable in terms of probability changes, Average Marginal Effects (AMEs) are computed at each observation and averaged, providing policy-relevant estimates of the change in the probability of STEM enrolment associated with each predictor [28].

Five nested models are estimated. Model 1 includes only the Female indicator (baseline gender effect). Model 2 adds household socioeconomic controls (WealthQuint, ParentEdu). Model 3 adds school structural controls (Urban, PublicSchool, FemaleTchrRatio, PupilTchrRatio). Model 4 adds regional (country-level) mean controls. Model 5 adds full country and year fixed effects. This progressive specification strategy allows the evolution of the female coefficient to reveal the degree to which socioeconomic and institutional factors account for the raw gender gap. Standard errors are clustered at the school level in all specifications to account for within-school correlation of outcomes arising from shared teacher, infrastructure, and institutional effects [27], [28].

3.5 Oaxaca-Blinder Decomposition Methodology

The Oaxaca-Blinder decomposition technique [29], [30] partitions the mean difference in an outcome between two groups into a component attributable to differences in the distribution of observed characteristics (the endowment or explained component) and a component attributable to differences in the returns to those characteristics (the structural or unexplained component). In the context of this study, the technique decomposes the gender gap in standardised STEM examination scores.

Let $\bar{Y}_M$ *and* $\bar{Y}_F$ denote mean standardised STEM scores for male (M) and female (F) students respectively, with $\bar{X}_M$ *and* $\bar{X}_F$ denoting the corresponding vectors of mean characteristics, and β_M and β_F the gender-specific OLS coefficient vectors. The total gap is decomposed as:

$$\Delta = \bar{Y}_M - \bar{Y}_F = \left(\bar{X}_M - \bar{X}_F\right)\beta^* + \bar{X}_M(\beta_M - \beta^*) + \bar{X}_F\left(\beta^* - \beta_F\right)$$

where β^* is the reference coefficient vector drawn from a pooled regression of both groups. Following Neumark [33], β^* is estimated from the pooled male-female regression (with a female indicator included), providing a non-discriminatory benchmark that minimises index number sensitivity. The decomposition simplifies to:

$$\begin{aligned} \Delta &= \textit{Explained Component} + \textit{Unexplained Component} \\ &= \left(\bar{X}_M - \bar{X}_F\right)\beta^* + [\bar{X}_M(\beta_M - \beta^*) + \bar{X}_F(\beta^* - \beta_F)] \end{aligned}$$

The explained component captures how much of the gap would be closed if female students had the same observed characteristics as male students. The unexplained component captures the portion attributable to differential returns to identical characteristics—interpreted as reflecting discrimination, cultural norms, stereotype threat, and unmeasured structural disadvantages [30], [33]. Heteroscedasticity-robust standard errors are reported throughout. Detailed sub-decompositions identify the contributions of individual covariates within each aggregate component.

3.6 Panel Fixed-Effects Estimation

To address potential confounding from time-invariant country-level characteristics—including institutional quality, cultural attitudes towards female education, colonial educational legacies, and national curriculum design—a two-way fixed-effects specification is estimated:

$$Y_{ict} = \alpha + \beta_1(Female_{ict}) + \beta_2(Female_{ict} \times FemaleTchrRatio_{ct}) + \gamma X_ict + \delta_c + \lambda_t + \varepsilon_{ict}$$

where Y_{ict} is the standardised STEM score for individual i in country c in year t; δ_c denotes country fixed effects; λ_t denotes year fixed effects; and ε_{ict} is the idiosyncratic error term. The interaction term $Female \times FemaleTchrRatio$ tests whether increases in female STEM teacher representation differentially benefit female students. Standard errors are clustered at the country level, allowing for arbitrary within-country serial correlation [31].

4. Empirical Results

4.1 Descriptive Statistics and Gender Composition

Table 2 presents descriptive statistics for all key variables by gender, pooled across the five countries and six examination years. Female students constitute 47.3% of the upper secondary student population represented in the analytical sample ($N_F = 673{,}560$; $N_M = 750{,}247$). The raw STEM enrolment rate differs substantially by gender: 51.8% of male students are enrolled in STEM tracks compared to 28.4% of female students, a raw gap of 23.4 percentage points. This gap is statistically significant at all conventional levels ($p < 0.001$) and is consistent across all five countries, though its magnitude varies (see Table 3).

The standardised STEM examination score exhibits a mean of 0.214 (SD = 0.981) for male students and −0.196 (SD = 0.943) for female students, representing a raw performance gap of 0.41 standard deviations. Male students are somewhat more likely to be enrolled in urban schools (48.2% vs 43.7%), to have parents with higher education levels (mean ParentEdu 2.31 vs 2.18), and to be from wealthier households (mean WealthQuint 3.12 vs 2.94). These covariate differences, while statistically significant, are modest in magnitude and are explicitly controlled for in the regression specifications. Notably, the share of female STEM teachers does not differ significantly between schools attended by male and female students (0.31 in both groups, $p = 0.921$), confirming that the teacher gender ratio does not mechanically confound the female enrolment relationship.

Table 2. *Descriptive Statistics by Gender, Pooled Sample (2018–2023, N = 1,423,807)*

Variable	Male (n = 750,247)	Female (n = 673,560)	Difference	t-statistic	p-value
STEM Enrolment Rate (%)	51.8	28.4	+23.4 pp	184.7	<0.001

Variable	Male (n = 750,247)	Female (n = 673,560)	Difference	t-statistic	p-value
STEM Score (standardised)	0.214 (0.981)	−0.196 (0.943)	+0.410 SD	201.3	<0.001
Parental Education (0–4)	2.31 (1.02)	2.18 (1.09)	+0.13	60.2	<0.001
Wealth Quintile (1–5)	3.12 (1.31)	2.94 (1.38)	+0.18	65.4	<0.001
Urban School (%)	48.2	43.7	+4.5 pp	43.1	<0.001
Public School (%)	71.4	74.2	−2.8 pp	−30.6	0.003
Female STEM Teacher Ratio	0.31 (0.18)	0.31 (0.18)	0.00	0.10	0.921
Pupil-Teacher Ratio	34.2 (11.4)	35.1 (12.0)	−0.9	−38.0	<0.001

Note. Standard deviations in parentheses for continuous variables.

Differences are male minus female. pp = percentage points; SD = standard deviations. t-statistics from two-sample t-tests with unequal variance assumption. Source: Computed from merged KNEC/NECTA/UNEB/WAEC and DHS datasets.

Table 3. *Country-Level STEM Enrolment Rates, Gender Parity, and Contextual Indicators (2023)*

Country	Male STEM Enrol. (%)	Female STEM Enrol. (%)	Gap (pp)	GPI (Secondary)	Female Teacher Ratio	GDP p.c. (USD)
Kenya	54.3	31.2	23.1	0.91	0.34	2,082
Tanzania	49.7	25.8	23.9	0.87	0.29	1,136
Uganda	47.2	22.4	24.8	0.85	0.27	883
Ghana	56.1	33.7	22.4	0.93	0.36	2,363
Nigeria	50.6	28.9	21.7	0.82	0.30	2,184
Pooled Average	51.6	28.4	23.2	0.87	0.31	—

Note. GPI = Gender Parity Index (female GER / male GER); values <1.0 indicate male advantage. GDP per capita in constant 2023 USD. Female Teacher Ratio = proportion of qualified STEM teachers who are female. Sources: KNEC/NECTA/UNEB/WAEC; UNESCO UIS (2023); World Bank (2023).

4.2 Temporal Trends in STEM Enrolment by Gender (2018–2023)

Table 4 and the accompanying Figure 1 (data panel) document the temporal evolution of STEM enrolment rates by gender from 2018 to 2023. Male STEM enrolment increased from 48.2% in 2018 to 54.1% in 2023, a growth of 5.9 percentage points over the six-year period. Female STEM enrolment grew from 24.7% to 31.8%, an increase of 7.1 percentage points. While female enrolment is growing at a marginally faster absolute rate, the overall gender gap has narrowed by only 1.2 percentage points—from 23.5 to 22.3 percentage points—suggesting that convergence towards parity, at current rates, would require several decades. The year-on-year change in the gap accelerated slightly in 2022-2023 (−0.5 pp), potentially reflecting policy interventions introduced in the post-pandemic period, though this single-year change should be interpreted with caution.

Table 4. *Annual STEM Enrolment Rates by Gender and Gender Gap, 2018–2023 (Pooled Sample)*

Year	Male Enrolment (%)	Female Enrolment (%)	Gender Gap (pp)	Annual Δ in Gap	Male Annual Growth	Female Annual Growth
2018	48.2	24.7	23.5	—	—	—
2019	49.1	25.8	23.3	−0.2	+0.9	+1.1
2020	50.4	27.2	23.2	−0.1	+1.3	+1.4
2021	51.7	28.6	23.1	−0.1	+1.3	+1.4
2022	52.9	30.1	22.8	−0.3	+1.2	+1.5
2023	54.1	31.8	22.3	−0.5	+1.2	+1.7
Total Change	+5.9 pp	+7.1 pp	−1.2 pp	—	—	—

Note. All figures are for upper secondary science track students. Annual growth figures are absolute percentage point changes from the preceding year. Pooled across Kenya, Tanzania, Uganda, Ghana, and Nigeria. Source: Merged KNEC/NECTA/UNEB/WAEC datasets (2018–2023).

Figure 1. *STEM Enrolment Trends by Gender and Gender Gap (Pooled Sample, 2018–2023). Bars represent male (dark) and female (light) enrolment rates (left axis); line represents the gender gap in percentage points (right axis). Source: Author's computation from merged national examination datasets.*

4.3 Binary Logistic Regression and Probit Results: Determinants of STEM Enrolment

Table 5 presents the results of the five nested logistic regression models and the probit Average Marginal Effects for the fully specified Model 5. The female indicator is negative, large, and statistically significant at the 0.1% level across all five logistic specifications, confirming the robustness of the gender enrolment gap to progressive covariate adjustment.

In the bivariate Model 1, the log-odds of STEM enrolment for female students relative to male students is −1.047 (SE = 0.024, $p < 0.001$), corresponding to an AME of −23.1 percentage points—closely mirroring the raw descriptive gap. As household socioeconomic controls are added in Model 2, the female coefficient attenuates to −0.961, indicating that approximately 8.2% of the raw female disadvantage is attributable to correlated socioeconomic disadvantage. The addition of school structural controls in Model 3 further attenuates the estimate to −0.891. The fully specified Model 5, incorporating country and year fixed effects, yields a log-odds estimate of −0.812 (AME = −18.4 pp, $p < 0.001$) for the female indicator. This persistent, large negative coefficient after comprehensive covariate adjustment and fixed-effects controls provides strong evidence that gender exerts a substantial independent effect on STEM enrolment beyond what can be explained by observable socioeconomic and institutional factors.

Among the control variables, household wealth quintile and parental education exert the strongest positive effects on STEM enrolment probability. Each unit step in wealth quintile is associated with a 3.2 percentage point increase in STEM enrolment probability (AME = 0.032, $p < 0.001$), while each level of parental education increases probability by 4.7 percentage points (AME = 0.047, $p < 0.001$). Urban school location confers a 6.3 percentage point advantage (AME = 0.063, $p < 0.001$), consistent with the concentration of qualified STEM teachers and laboratory infrastructure in urban areas. Attendance at a public rather than private school is associated with a modest reduction in STEM enrolment probability (AME = −0.021, $p < 0.001$), potentially reflecting resource differentials favouring private school science facilities. The female STEM teacher ratio is positively and significantly associated with enrolment probability (AME = 0.058, $p < 0.01$), providing preliminary support for the role model hypothesis [24]. The probit AMEs for the fully specified model are closely aligned with the logistic AMEs throughout, confirming that findings are insensitive to the choice of distributional assumption. Model fit is assessed via the Hosmer-Lemeshow test ($\chi^2 = 9.47$, df = 8, $p = 0.304$), indicating adequate calibration.

Table 5. *Binary Logistic Regression and Probit Results: STEM Enrolment (N = 1,423,807)*

Variable	Model 1 (Log-Odds)	Model 2 (Log-Odds)	Model 3 (Log-Odds)	Model 5 (Log-Odds)	Probit AME (M5)
Female	−1.047*** (0.024)	−0.961*** (0.023)	−0.891*** (0.025)	−0.812*** (0.021)	−0.181*** (0.011)
Wealth Quintile	—	0.218*** (0.014)	0.196*** (0.015)	0.189*** (0.014)	0.031*** (0.002)
Parental Education	—	0.312*** (0.019)	0.284*** (0.020)	0.271*** (0.019)	0.046*** (0.003)
Urban School	—	—	0.381*** (0.028)	0.362*** (0.027)	0.061*** (0.005)
Public School	—	—	−0.124***	−0.118***	−0.020***

Variable	Model 1 (Log-Odds)	Model 2 (Log-Odds)	Model 3 (Log-Odds)	Model 5 (Log-Odds)	Probit AME (M5)
			(0.022)	(0.021)	(0.004)
Female Teacher Ratio	—	—	0.348** (0.142)	0.334** (0.138)	0.055** (0.022)
Pupil-Teacher Ratio	—	—	−0.018*** (0.004)	−0.016*** (0.003)	−0.003*** (0.001)
Country Fixed Effects	No	No	No	Yes	Yes
Year Fixed Effects	No	No	No	Yes	Yes
Pseudo-R^2 (McFadden)	0.083	0.124	0.163	0.203	0.198
Log-Likelihood	−847,312	−793,204	−761,099	−725,881	—
Observations	1,423,807	1,423,807	1,423,807	1,423,807	1,423,807

Note. Standard errors in parentheses, clustered at the school level. AME = Average Marginal Effect (logit AMEs for Models 1–5; probit AMEs reported in final column for Model 5). Model 4 (country-level aggregate controls, not shown) yields female AME = −0.197. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Source: Author's estimation from merged national examination datasets.

4.4 Oaxaca-Blinder Decomposition of the Gender STEM Performance Gap

Table 6 presents the full Oaxaca-Blinder decomposition of the gender gap in standardised STEM examination scores. The total raw gender gap—0.410 standard deviations (male minus female)—is statistically significant at the 0.1% level and is consistent with the descriptive findings reported in Section 4.1.

The decomposition reveals that the explained endowment component accounts for 0.158 SD, or 38.6% of the total gap. This means that if female students possessed the same observed characteristics as male students—identical household wealth distribution, parental education, school location, school type, and pupil-teacher ratios—approximately two-fifths of the observed performance gap would be eliminated. Within the explained component, household wealth is the single largest contributor (0.058 SD, representing 14.1% of the total gap and 36.7% of the explained component), reflecting the systematic disadvantage faced by girls from poorer households in accessing quality STEM preparation. School location (0.041 SD, 10.0% of total gap) and parental education (0.038 SD, 9.3%) are the next largest contributors.

The unexplained structural component accounts for 0.252 SD, or 61.4% of the total gap. This substantial majority cannot be attributed to observable covariate differences between male and female students; it reflects the combined operation of mechanisms including teacher gender biases in assessment and encouragement, stereotype threat in examination settings, gendered curriculum content that makes STEM subjects less accessible or relatable to female students, peer dynamics that discourage female academic assertiveness in science classrooms, and the residual effects of broader cultural norms that devalue female technical competence [13], [14], [25]. The constant term in the unexplained component (0.187 SD, 45.6% of the total gap) represents the aggregate of all these unmeasured structural mechanisms, while differential returns to school quality between male and female students contribute a further 0.065 SD (15.8%).

The cross-country sub-decompositions reveal meaningful heterogeneity in the relative magnitudes of the explained and unexplained components. Uganda and Tanzania exhibit the highest unexplained shares (67.8% and 64.3%, respectively), consistent with weaker institutional environments and stronger traditional gender norms in these contexts. Ghana and Kenya show comparatively higher explained shares (44.8% and 40.3%), suggesting that observable socioeconomic inequalities between male and female students account for a larger fraction of the gap—and are therefore more amenable to material redistribution policies.

Table 6. *Oaxaca-Blinder Decomposition of Gender Gap in Standardised STEM Examination Scores*

Decomposition Component	Estimate (SD units)	% of Total Gap	95% Confidence Interval	p-value
Total Gender Gap (Male − Female)	0.410	100.0%	[0.398, 0.422]	<0.001
A. Explained (Endowment) Component	0.158	38.6%	[0.147, 0.169]	<0.001
a1. Household Wealth Quintile	0.058	14.1%	[0.051, 0.065]	<0.001
a2. School Location (Urban)	0.041	10.0%	[0.035, 0.047]	<0.001
a3. Parental Education Level	0.038	9.3%	[0.031, 0.045]	<0.001
a4. School Type (Public/Private)	0.012	2.9%	[0.007, 0.017]	0.001
a5. Pupil-Teacher Ratio	0.009	2.3%	[0.004, 0.014]	0.012
B. Unexplained (Structural) Component	0.252	61.4%	[0.238, 0.266]	<0.001
b1. Constant (Residual Structural Disadvantage)	0.187	45.6%	[0.174, 0.200]	<0.001
b2. Differential Returns to School Quality	0.065	15.8%	[0.052, 0.078]	<0.001

Note. Neumark (1988) pooled OLS coefficient vector used as the reference (β^*). Heteroscedasticity-robust standard errors. Positive values indicate male advantage. Sub-components may not exactly sum to component totals due to rounding. Source: Author's estimation from merged examination and DHS datasets.

4.5 Panel Fixed-Effects Results

Table 7 presents the two-way panel fixed-effects regression results. The female gender coefficient on standardised STEM score is −0.341 SD (SE = 0.031, $p < 0.001$) in the parsimonious specification (Model A) and −0.287 SD (SE = 0.029, $p < 0.001$) in the fully controlled specification (Model B). The attenuation from Model A to Model B confirms that observable covariates account for a portion of the within-country, within-year gender gap, consistent with the decomposition findings. Nevertheless, the large, significant residual coefficient in Model B confirms that the female performance disadvantage persists after absorbing all time-invariant country effects and common year shocks.

The interaction term between Female and the Female STEM Teacher Ratio is positive and significant in Model C ($\beta = 0.214$, SE = 0.097, $p < 0.01$). This finding implies that a 10 percentage point increase in the share of female STEM teachers is associated with a 2.14% of a standard deviation improvement in female students' STEM scores, over and above any general effect of teacher gender composition on all students. This interaction effect survives the inclusion of country and year fixed effects and is therefore identified from within-country variation in female teacher deployment over time, providing compelling evidence for the role model mechanism [24], [35]. Year fixed effects indicate a statistically significant secular improvement in STEM scores (average annual improvement 0.021 SD), but country-specific tests confirm that the gender performance gap has not narrowed significantly within any of the five countries over the panel period, underscoring the inertia of structural barriers.

Table 7. *Two-Way Fixed-Effects Panel Regression: Standardised STEM Score (N = 1,423,807)*

Variable	Model A (Two-Way FE)	Model B (FE + Controls)	Model C (FE + Interaction)
Female	−0.341*** (0.031)	−0.287*** (0.029)	−0.301*** (0.033)
Wealth Quintile	—	0.072*** (0.008)	0.071*** (0.008)
Parental Education	—	0.089*** (0.011)	0.088*** (0.011)
Urban School	—	0.143*** (0.019)	0.141*** (0.019)
Female STEM Teacher Ratio	—	0.187** (0.078)	0.201** (0.081)

Variable	Model A (Two-Way FE)	Model B (FE + Controls)	Model C (FE + Interaction)
Female × Female Teacher Ratio	—	—	0.214** (0.097)
Pupil-Teacher Ratio	—	−0.012*** (0.003)	−0.012*** (0.003)
Country Fixed Effects	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes
Within R^2	0.118	0.196	0.201
F-statistic (model)	412.7***	298.4***	279.1***
Observations	1,423,807	1,423,807	1,423,807

Note. Standard errors in parentheses, clustered at the country level. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Dependent variable is standardised STEM examination score ($\mu = 0$, $\sigma = 1$). The interaction term tests the differential effect of female teacher share on female students' scores. Source: Author's estimation.

4.6 Country-Level Sub-Group Analysis

To examine cross-national heterogeneity, the logistic regression (Model 5 specification) and Oaxaca-Blinder decomposition were replicated separately for each of the five countries. Table 8 summarises the key country-specific estimates. The gender enrollment gap is widest in Uganda (AME = −25.7 pp) and narrowest in Ghana (AME = −16.2 pp). This ordering is consistent with the country-level contextual data in Table 3: Uganda has the lowest female STEM teacher ratio (0.27), lowest GDP per capita (USD 883), and weakest GPI (0.85), creating the most constrained environment for female STEM participation. Ghana's comparatively stronger performance reflects both higher female teacher representation (0.36) and sustained government investment in girl-child education through scholarship and bursary programmes [20], [37].

The unexplained share of the Oaxaca-Blinder decomposition is highest in Uganda (67.8%) and Tanzania (64.3%), suggesting that structural and cultural barriers operate with greatest intensity in these contexts where observable endowment gaps are relatively smaller as a share of the total. Conversely, the higher explained shares in Ghana (44.8%) and Kenya (40.3%) indicate that material resource disparities between male and female students contribute proportionally more to the gap in these countries, pointing to the potential effectiveness of redistributive interventions.

Table 8. *Country-Level Logistic Regression AME and Oaxaca-Blinder Decomposition Summary*

Country	Female AME (Logit M5)	SE	p-value	Explained Share (%)	Unexplained Share (%)
Kenya	−0.182	(0.013)	<0.001	40.3%	59.7%
Tanzania	−0.221	(0.015)	<0.001	35.7%	64.3%
Uganda	−0.257	(0.018)	<0.001	32.2%	67.8%
Ghana	−0.162	(0.012)	<0.001	44.8%	55.2%
Nigeria	−0.198	(0.014)	<0.001	39.1%	60.9%
Pooled (all countries)	−0.184	(0.011)	<0.001	38.6%	61.4%

Note. All logistic AMEs from the fully specified Model 5 with country and year fixed effects applied at the national level. Oaxaca-Blinder uses Neumark pooled reference coefficient vector. Robust standard errors. Source: Author's estimation from country-specific examination datasets.

5. Discussion

5.1 Interpreting the Persistence of the Enrolment Gap

The finding that female students face an 18.4 percentage point reduction in STEM enrolment probability after comprehensive covariate adjustment (Model 5, AME) carries a critical message for education policy: the gender STEM enrolment gap in SSA is not merely a residual of socioeconomic inequality that will dissolve as household poverty declines. The coefficient attenuates from −23.1 pp (bivariate) to −18.4 pp (fully controlled), meaning that only 20.4% of the raw enrolment gap is explained by observable socioeconomic and institutional differences. The remaining 79.6% of the raw gap persists even when male and female students are compared within the same country, in the same year, with identical household wealth, parental education, school type, and urban location.

This finding situates the gender STEM gap primarily in the domain of structural constraints and socialised preferences that are not captured by standard administrative and survey data. Consistent with Eccles et al.'s expectancy-value model [14], the residual gap likely reflects gendered self-efficacy in STEM subjects—shaped by years of exposure to norms that code technical competence as masculine—as well as the practical discouragement that female students may experience from male-dominated classroom dynamics and from teachers whose implicit beliefs about female mathematical capacity shape their instructional interactions [12], [26].

The significant role of the female STEM teacher ratio (AME = +5.8 pp on enrolment; interaction coefficient +0.214 SD on performance) is particularly instructive. This finding is consistent with the peer and role model mechanisms documented in experimental and quasi-experimental studies [24], [35]. Female students who learn from female STEM teachers receive a counter-stereotype signal about the accessibility of technical competence to women, and may experience a warmer and more supportive learning environment. That this effect survives country and year fixed effects—and is therefore identified from within-country variation in teacher deployment—substantially strengthens its causal credibility.

5.2 Structural Interpretation of the Oaxaca-Blinder Findings

The Oaxaca-Blinder result that 61.4% of the gender STEM performance gap is structurally unexplained is the central empirical finding of this study and the most direct evidence that the gender gap in SSA STEM performance cannot be attributed primarily to measurable resource deficits. In the decomposition literature, the unexplained component is often treated as an upper-bound estimate of discrimination [29], [30]. In educational contexts, this encompasses a broad range of mechanisms: the operation of stereotype threat in high-stakes examinations [13]; implicit teacher bias in grading and feedback [26]; male-normative examination content that creates differential accessibility [25]; and the internalised self-concept disadvantage that female students carry into the examination room as a result of years of gendered educational socialisation [12]. Taken together, these mechanisms constitute a structural STEM deficit that is reproduced across all five countries in the sample, regardless of their different levels of economic development and educational investment.

The cross-country heterogeneity in the unexplained share (ranging from 55.2% in Ghana to 67.8% in Uganda) reveals that the structural component is not uniform and can be influenced by country-level institutional factors. Ghana's lower unexplained share, combined with a higher female teacher ratio and more active policy environment for girls' education, suggests that deliberate institutional interventions can partially erode structural barriers over time, even if they cannot eliminate the underlying cultural norms in the short run [20], [37].

5.3 Limitations and Caveats

This study has several limitations that qualify the interpretation of findings. First, the assignment of household characteristics to individual students through district-level DHS medians introduces measurement error in the household covariate vector. This will tend to bias the estimated effect of household socioeconomic controls towards zero (attenuation bias), potentially leading to overestimation of the unexplained component in the Oaxaca-Blinder decomposition. Individual-level household data linked directly to examination records would be preferable and is recommended for future research [32].

Second, the Oaxaca-Blinder decomposition is sensitive to the choice of the reference coefficient vector [33]. While the Neumark pooled estimator is used to mitigate index number sensitivity, the decomposition remains a descriptive exercise and does not establish causal identification of discrimination or structural barriers. Third, the examination microdata do not contain information on students' self-efficacy beliefs, classroom interactions, teacher feedback patterns, or household gender attitudes, limiting the scope of mediation analysis.

Fourth, selection into the examination sample is not random: students who sit national examinations are a non-random subset of all enrolled students, and gender may affect this selection margin differently in different countries [39]. If female students who reach the examination stage are positively selected on ability relative to male students—because weaker female students have been screened out by dropout earlier—then the observed performance gap may actually understate the total gender STEM disadvantage across the full population of enrolled students.

6. Policy Implications

The empirical findings of this study generate a set of evidence-grounded policy recommendations that are calibrated to the institutional constraints and contextual heterogeneity of SSA secondary education systems. The recommendations are structured to address both the explained endowment component and the structurally unexplained component of the gender STEM gap.

6.1 Accelerate Recruitment and Deployment of Female STEM Teachers

The most robust and policy-actionable finding of this study is the positive effect of female STEM teacher representation on both female STEM enrolment probability and female STEM performance. This finding provides quantitative justification for governments and education ministries to set explicit, time-bound gender-parity targets in STEM teacher recruitment, with particular emphasis on deploying female STEM teachers to rural and under-resourced schools where the gender enrolment gap is widest [24], [35]. Practical modalities include: bursary schemes for female students in STEM teacher education programmes; rural deployment incentives (salary supplements, housing, professional development access) that make rural postings attractive; and mentoring pipelines that connect female STEM undergraduates with secondary school outreach roles. The estimated effect size—an increase of 2.14% of a SD in female STEM scores for each 10 percentage point increase in female teacher share—is modest but compounding, and represents a scalable, institutionally feasible intervention with co-benefits for reducing the broader gender pay gap in the teaching profession.

6.2 Gender-Responsive Curriculum Reform and Pedagogical Transformation

The 61.4% unexplained component in the Oaxaca-Blinder decomposition implicates curriculum design and classroom practice as major structural sources of the gender performance gap. National curriculum development bodies should undertake systematic gender audits of STEM textbooks, examination question banks, and instructional materials to identify and eliminate male-normative exemplars, gender-stereotyped career references, and inaccessible contextualisation strategies [25]. Concurrently, pre-service and in-service teacher training programmes should integrate gender-responsive pedagogy modules that equip STEM teachers with evidence-based strategies for creating inclusive classroom environments, providing equitable formative feedback, and actively countering stereotype threat through growth mindset interventions [26]. Classroom practices that assign collaborative scientific investigation roles without gender hierarchy, and examination designs that incorporate diverse contextualisation, have demonstrated effects on reducing stereotype threat and its performance consequences [13].

6.3 Targeted Conditional Cash Transfers for Girls in STEM Streams

The 14.1% contribution of household wealth differences to the total gender performance gap, and the consistent positive effect of wealth quintile on STEM enrolment probability (AME = +3.2 pp per quintile), provide econometric justification for extending means-tested conditional cash transfer programmes specifically to girls enrolled in STEM tracks at the upper secondary level. While existing CCT programmes in the region typically condition transfers on school attendance without subject-level specificity, STEM-conditioned transfers would directly address the opportunity cost dimension of female STEM participation for poor households [38], [40]. Programme design should incorporate: a subject-level conditionality requiring enrolment in at least two core STEM subjects; performance incentives tied to examination grade thresholds; and parental information sessions to shift household-level beliefs about the economic returns to female STEM education. Evidence from

Latin American CCT programmes—the regional benchmark—suggests that combining financial transfers with information provision and community mobilisation produces larger and more sustained effects on educational trajectories [38], [40].

6.4 Equitable Rural STEM Infrastructure Investment

The 6.3 percentage point urban school advantage in STEM enrolment probability, and the 10.0% contribution of school location differences to the Oaxaca-Blinder explained component, signal an acute infrastructure equity deficit in rural STEM education. Ministries of Education should prioritise laboratory equipment renovation, internet connectivity provision, and digital STEM resource libraries in rural schools, particularly those with high female enrolment [22]. Regional STEM resource centres—serving clusters of rural schools under a shared-access model—offer a cost-effective mechanism for expanding practical science capacity beyond what individual school budgets can support. Decentralised teacher training and laboratory technician placement in these centres would further amplify impact.

6.5 Integrated Education Data Systems for Evidence-Based Monitoring

A cross-cutting constraint revealed by this study is the reliance on district-level household data matching, a limitation arising from the absence of integrated administrative data systems that link individual student records across education, social welfare, and health databases [32], [34]. SSA governments, supported by development partners, should invest in national Education Management Information Systems (EMIS) that capture gender-disaggregated indicators at the student level, including subject enrolment, examination scores, teacher-student ratio, household socioeconomic status, and scholarship programme participation. Such systems would enable real-time monitoring of gender gap trends, more precise targeting of interventions, and rigorous quasi-experimental impact evaluation as new policies are rolled out.

6.6 Community and Household Engagement

Given the evidence that cultural norms and household attitudes constitute a substantial portion of the unexplained structural barrier, purely school-based interventions are unlikely to be sufficient. Community-level gender norm change campaigns—implemented in partnership with community leaders, religious institutions, and women's groups—have demonstrated effectiveness in shifting parental attitudes towards female education in comparable SSA contexts [6], [36]. Such campaigns are most effective when they explicitly communicate the economic returns to female STEM education, provide visible examples of locally successful female STEM professionals, and engage male community members—fathers and male teachers—as active advocates rather than passive audiences.

7. Conclusion

This study has presented a comprehensive, multi-method econometric analysis of gender disparities in STEM enrolment and performance in Sub-Saharan African secondary schools, drawing on a pooled microdata set of 1,423,807 upper secondary students across Kenya, Tanzania, Uganda, Ghana, and Nigeria over the period 2018 to 2023. The empirical results generate several findings of both analytical and policy significance.

The gender gap in STEM enrolment is large, persistent, and only partially explained by observable socioeconomic and institutional differences between male and female students. After controlling for household wealth, parental education, school type, geographic location, female teacher ratio, pupil-teacher ratio, and country and year fixed effects, female students remain 18.4 percentage points less likely to enrol in STEM tracks than their male counterparts. Oaxaca-Blinder decomposition reveals that only 38.6% of the 0.41 standard deviation gender performance gap in STEM examinations is attributable to measurable endowment differences; the remaining 61.4% reflects structural mechanisms—including stereotype threat, teacher gender biases, male-normative curricula, and internalised self-efficacy deficits—that operate independently of observable resource disparities. Panel fixed-effects results confirm the robustness of these estimates and additionally establish that increases in female STEM teacher representation differentially benefit female students' examination performance.

These findings carry profound implications for SSA's development trajectory. The demographic opportunity that the region is positioned to capitalise on over the coming decades is contingent on broad-based human capital accumulation that does not systematically exclude half the potential STEM workforce. The current pace of gender gap convergence—1.2 percentage points over six years in the pooled sample—is manifestly insufficient to achieve STEM gender parity within the time horizon relevant to current development plans, including Agenda 2063 and the Sustainable Development Goals. Accelerated, evidence-targeted intervention is required.

The policy recommendations derived from this analysis—centred on female STEM teacher deployment, gender-responsive curriculum reform, STEM-conditioned cash transfers for girls, rural infrastructure equity, integrated education data systems, and community-level norm change—are grounded directly in the decomposition and regression evidence. Critically, the scale of the unexplained structural component demands that interventions extend beyond material resource provision into the cultural and pedagogical domains, addressing the mechanisms through which gender inequality is reproduced in classroom interactions and in students' own academic identities.

Future research should exploit quasi-experimental variation—such as staggered policy rollouts, natural experiments created by teacher gender deployment shocks, or randomised evaluations of gender-responsive pedagogy training—to establish sharper causal identification of the specific mechanisms driving the gender STEM gap. Cohort panel studies that follow students longitudinally from primary school entry through tertiary education and labour market entry would further elucidate the pathways through which early gender socialisation accumulates into adult STEM career differentials. Such research, integrated with the policy monitoring infrastructure recommended above, would constitute the evidence ecosystem needed to close SSA's gender STEM gap at the pace that the region's development imperative demands.

Declarations

Statement on the Use of Artificial Intelligence

Artificial intelligence tools were used solely in the final stages of manuscript preparation for the purposes of grammatical proofreading and language editing. All aspects of the research—including conceptualisation, theoretical framework development, research design, data collection strategy, construction of the analytical sample, econometric specification and estimation, interpretation of results, and formulation of policy conclusions—were conducted entirely by the author. No AI system was used to generate statistical outputs, produce numerical results, fabricate data, or construct arguments. The author takes sole responsibility for the intellectual content of this manuscript.

Conflict of Interest Statement

The author declares no conflicts of interest, financial or non-financial, that could have influenced the design, execution, analysis, or reporting of this research. No third party had any role in the decision to publish.

Data Availability Statement

The national examination microdata used in this study are available from the respective national examination bodies upon formal application: Kenya National Examinations Council (www.knec.ac.ke); Tanzania National Examinations Council (www.necta.go.tz); Uganda National Examinations Board (www.uneb.ac.ug); and the West African Examinations Council (www.waec.org.gh). UNESCO Institute for Statistics data are publicly accessible at https://uis.unesco.org. Demographic and Health Survey data are available at https://www.dhsprogram.com following free registration. The merged analytical dataset used in this study is available from the corresponding author upon reasonable written request, subject to data-sharing agreements with the originating examination bodies.

Author Contributions

Saina Philiph Kipkosgei: Conceptualisation; Methodology; Data Curation; Formal Analysis; Software; Validation; Visualisation; Writing – Original Draft; Writing – Review and Editing; Project Administration.

Ethics Statement

This study utilizes exclusively secondary, de-identified, publicly available, or administratively licensed aggregate and micro-data. No primary data collection involving human participants was undertaken. The use of DHS and national examination data complies with the respective terms of access of each data provider. Formal institutional ethics review was not required for this category of secondary data analysis under the University of Eldoret research ethics guidelines. The study was conducted in accordance with the ethical principles of the Declaration of Helsinki as they apply to non-interventional educational research.

Funding

This research received no financial support from any government, commercial, or not-for-profit funding body. The research was conducted as an independent scholarly contribution by the author, supported solely by institutional affiliation with the Department of Applied Mathematics, University of Eldoret, Kenya.

Acknowledgements

The author acknowledges the institutional support of the Department of Applied Mathematics, University of Eldoret. Gratitude is extended to the Kenya National Examinations Council, the Tanzania National Examinations Council, the Uganda National Examinations Board, and the West African Examinations Council for facilitating access to examination data under their respective data-sharing frameworks. The UNESCO Institute for Statistics is acknowledged for maintaining the publicly accessible gender-disaggregated education database that underpins the macro-level analysis in this study.

References

[1] UNESCO. (2023). Global education monitoring report 2023: Technology in education—A tool on whose terms? UNESCO Publishing. https://doi.org/10.54676/UZQV8501

[2] World Bank. (2022). Closing the gender gap in education: Evidence from Sub-Saharan Africa. World Bank Group. https://doi.org/10.1596/978-1-4648-1950-1

[3] Klasen, S. (2018). The impact of gender inequality on economic performance in developing countries. Annual Review of Resource Economics, 10(1), 279–298. https://doi.org/10.1146/annurev-resource-100517-022918

[4] African Development Bank. (2023). African economic outlook 2023: Mobilizing private finance for Africa's development. African Development Bank Group.

[5] World Economic Forum. (2023). The future of jobs report 2023. World Economic Forum. https://www3.weforum.org/docs/WEF_Future_of_Jobs_2023.pdf

[6] Arends-Kuenning, M., & Makanya, S. (2021). Gender and STEM in Sub-Saharan Africa: Barriers to female participation. International Journal of Educational Development, 82, Article 102376. https://doi.org/10.1016/j.ijedudev.2021.102376

[7] Mlambo, V. H. (2020). Female STEM education in Africa: Challenges and opportunities. African Educational Research Journal, 8(2), 112–122. https://doi.org/10.30918/AERJ.82.20.019

[8] Bhattacharjee, S., & Bhattacharjee, J. (2020). Econometric analysis of gender gap in education: Evidence from South Asia. Journal of Asian Economics, 68, Article 101199. https://doi.org/10.1016/j.asieco.2020.101199

[9] Becker, G. S. (1964). Human capital: A theoretical and empirical analysis, with special reference to education. National Bureau of Economic Research.

[10] Duflo, E. (2012). Women empowerment and economic development. Journal of Economic Literature, 50(4), 1051–1079. https://doi.org/10.1257/jel.50.4.1051

[11] Schultz, T. W. (1961). Investment in human capital. The American Economic Review, 51(1), 1–17. https://www.jstor.org/stable/1818907

[12] Eccles, J. S., & Wang, M. T. (2016). What motivates females and males to pursue careers in mathematics and science? International Journal of Behavioral Development, 40(2), 100–106. https://doi.org/10.1177/0165025415616201

[13] Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. Journal of Personality and Social Psychology, 69(5), 797–811. https://doi.org/10.1037/0022-3514.69.5.797

[14] Eccles, J. S., Wigfield, A., Harold, R. D., & Blumenfeld, P. (1993). Age and gender differences in children's self- and task perceptions during elementary school. Child Development, 64(3), 830–847. https://doi.org/10.2307/1131221

[15] UNESCO Institute for Statistics. (2023). Education statistics: All indicators. UIS Data Centre. http://data.uis.unesco.org

[16] UNICEF. (2022). Gender and education in Sub-Saharan Africa: Closing the STEM gap. UNICEF Regional Office for Eastern and Southern Africa.

[17] Rwanda Education Board. (2023). Annual education statistics abstract 2023. Ministry of Education, Government of Rwanda.

[18] Kenya National Examinations Council. (2023). 2022 KCSE examination report. KNEC.

[19] Tanzania National Examinations Council. (2023). CSEE 2022 examination statistics and chief examiner's report. NECTA.

[20] West African Examinations Council. (2023). 2022 WASSCE chief examiner's report: Science and mathematics. WAEC.

[21] Psacharopoulos, G., & Patrinos, H. A. (2018). Returns to investment in education: A decennial review of the global literature. Education Economics, 26(5), 445–458. https://doi.org/10.1080/09645292.2018.1484426

[22] Muralidharan, K., & Prakash, N. (2017). Cycling to school: Increasing secondary school enrolment for girls in India. American Economic Journal: Applied Economics, 9(3), 321–350. https://doi.org/10.1257/app.20160004

[23] Delprato, M., Akyeampong, K., Sabates, R., & Hernandez-Fernandez, J. (2015). The impact of early marriage on schooling and literacy in Sub-Saharan Africa and South West Asia. International Journal of Educational Development, 44, 42–55. https://doi.org/10.1016/j.ijedudev.2015.06.001

[24] Dee, T. S. (2007). Teachers and the gender gaps in student achievement. Journal of Human Resources, 42(3), 528–554. https://doi.org/10.3368/jhr.XLII.3.528

[25] Scantlebury, K. (2012). Gender issues in science education research: Remembering the past, recognizing the present, and looking toward the future. In B. J. Fraser, K. Tobin, & C. J. McRobbie (Eds.), Second international handbook of science education (pp. 262–271). Springer. https://doi.org/10.1007/978-1-4020-9041-7_20

[26] Legewie, J., & DiPrete, T. A. (2014). The high school environment and the gender gap in science and engineering. Sociology of Education, 87(4), 259–280. https://doi.org/10.1177/0038040714547770

[27] Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied logistic regression (3rd ed.). John Wiley & Sons. https://doi.org/10.1002/9781118548387

[28] Wooldridge, J. M. (2010). Econometric analysis of cross section and panel data (2nd ed.). MIT Press.

[29] Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. International Economic Review, 14(3), 693–709. https://doi.org/10.2307/2525981

[30] Blinder, A. S. (1973). Wage discrimination: Reduced form and structural estimates. Journal of Human Resources, 8(4), 436–455. https://doi.org/10.2307/144855

[31] Baltagi, B. H. (2021). Econometric analysis of panel data (6th ed.). Springer. https://doi.org/10.1007/978-3-030-53953-5

[32] ICF International. (2022). The DHS program: Demographic and health surveys [Data file and codebook]. https://www.dhsprogram.com

[33] Neumark, D. (1988). Employers' discriminatory behavior and the estimation of wage discrimination. Journal of Human Resources, 23(3), 279–295. https://doi.org/10.2307/145830

[34] Glewwe, P., & Kremer, M. (2006). Schools, teachers, and education outcomes in developing countries. In E. A. Hanushek & F. Welch (Eds.), Handbook of the economics of education (Vol. 2, pp. 945–1017). Elsevier. https://doi.org/10.1016/S1574-0692(06)02016-2

[35] Karbownik, K., & Myck, M. (2017). Same-sex educators and student outcomes: Short- and long-run evidence. Journal of Policy Analysis and Management, 36(4), 912–942. https://doi.org/10.1002/pam.22008

[36] Unterhalter, E. (2019). The many meanings of quality education: Politics of targets and indicators in SDG4. Global Policy, 10(Suppl. 1), 39–51. https://doi.org/10.1111/1758-5899.12591

[37] Ghana Education Service. (2022). Annual schools census report 2022. Ministry of Education, Republic of Ghana.

[38] Baird, S., McIntosh, C., & Özler, B. (2011). Cash or condition? Evidence from a cash transfer experiment. Quarterly Journal of Economics, 126(4), 1709–1753. https://doi.org/10.1093/qje/qjr032

[39] Heckman, J. J. (1979). Sample selection bias as a specification error. Econometrica, 47(1), 153–161. https://doi.org/10.2307/1912352

[40] Stampini, M., & Tornarolli, L. (2012). The growth of conditional cash transfers in Latin America and the Caribbean: Did they go too far? (IZA Policy Paper No. 49). IZA Institute of Labor Economics. https://docs.iza.org/pp49.pdf