ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26aug705

Educational Credentials, Perceived Education-Work Alignment and Employment Status Among Youth Records in Sierra Leone

A Bias-Reduced Multivariable Analysis with a Transparent Data-Quality Audit

Ackmed Chebli[1]; Dr. Alhaji Hamza Conteh[2]; Dr. Brima Gegbe[3]

[1,2,3]Department of Mathematics and Statistics, Njala University, Sierra Leone

Publication Date: 2026/08/29

Abstract:

- ***Background:***

Earlier versions of this article contained unsupported sampling claims, duplicated effect sizes, inconsistent duration statistics and an unstable multivariable specification. This reanalysis uses the supplied cleaned microdata and treats data quality as part of the inferential problem.

- ***Methods:***

The workbook contains 1,342 records. Employment status was analysed using Firth bias-reduced logistic regression because no employed respondent lacked internet access, creating complete separation. The primary model used education level, age, coded gender, coded marital status, province, internet access and a three-item perception index. Education years were examined separately. Recorded non-employment duration was modelled exploratorily among 922 strictly eligible records using a Gamma log-link model.

- ***Results:***

Employment was recorded for 415 respondents (30.9%). The perception items were highly redundant (alpha = 0.961; pairwise r = 0.871-0.912) and were combined. In the bias-reduced model (n = 1,339), higher perception-index scores were strongly associated with employment (OR = 18.33, 95% CI 12.04-27.89), diploma education differed from no education (OR = 2.04, 95% CI 1.24-3.35), age was inversely associated after adjustment (OR = 0.72 per year), and the Eastern Province estimate was positive relative to Northern Province (OR = 1.78). Internet access produced an extreme estimate because it perfectly separated the outcome and is not interpreted as a stable population effect. Model AUC was 0.935; Hosmer-Lemeshow p = 0.068.

- ***Conclusion:***

The reanalysis supports conditional associations within the supplied records but not causal or nationally representative conclusions. The dataset contains structural dependencies and coding conflicts that materially limit substantive interpretation. Recovering the original questionnaire, sampling protocol and source records is necessary before external publication.

Keywords: *Youth Employment; Sierra Leone; Education; Skills Alignment; Firth Logistic Regression; Data Quality; Complete Separation.*

How to Cite: Ackmed Chebli; Dr. Alhaji Hamza Conteh; Dr. Brima Gegbe (2026) Educational Credentials, Perceived Education-Work Alignment and Employment Status Among Youth Records in Sierra Leone. *International Journal of Innovative Science and Research Technology*, 11(8), 1931-1940. https://doi.org/10.38124/ijisrt/26aug705

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26aug705

I. INTRODUCTION

Youth employment is central to economic participation, household welfare and social stability. Human capital theory predicts that schooling can improve productivity and labour-market outcomes (Schultz, 1961; Becker, 1964), while signalling theory emphasises the informational value of credentials (Spence, 1973). Labour-market segmentation theory draws attention to structural barriers and unequal access to better jobs (Doeringer & Piore, 1971), and skills-mismatch theory highlights discrepancies between education and job requirements (Verhaest & Omey, 2006; McGuinness & Bennett, 2007).

These frameworks do not justify treating every statistical association as causal. In cross-sectional survey records, employment, perceptions of curriculum relevance and beliefs about opportunities are measured contemporaneously. Employment itself may influence perceptions; shared response patterns may also inflate associations. Careful outcome definition, instrument documentation and model diagnostics are therefore essential.

➢ *Purpose and Contribution*

This article reanalyses the supplied SL Education Dataset and corrects seven identified problems: undocumented representativeness, ambiguous labour-force classification, and incomplete measurement documentation, an unspecified regression, overlapping schooling measures, duplicated correlations and unsupported duration effects. The contribution is methodological transparency: all reported numbers are regenerated from the attached workbook, questionable assumptions are disclosed, and unstable estimates are not presented as policy facts.

➢ *Research Questions*

- How are employment status and key characteristics distributed in the supplied records?
- Are education credentials, perceived education-work alignment, digital access, demographic factors and province conditionally associated with employment status?
- How sensitive are education results to using credential level rather than years of schooling?
- Among strictly eligible non-employed records, which variables are associated with recorded duration?
- What data-quality limitations constrain interpretation?

II. THEORETICAL AND EMPIRICAL FRAMEWORK

➢ *Human Capital and Signalling*

Human capital theory views schooling and training as investments in productive capability (Schultz, 1961; Becker, 1964). Signalling theory instead stresses that educational credentials communicate otherwise unobserved characteristics to employers (Spence, 1973). The distinction matters empirically: credential categories and years of schooling overlap, but they are not identical. Simultaneously entering both without a clear estimand may destabilise coefficients and obscure interpretation.

➢ *Segmentation, Mismatch and Digital Access*

Segmented labour markets can restrict mobility between protected and vulnerable forms of work. Skills mismatch can lower the value of formal schooling when curricula and acquired competencies do not align with labour demand. Internet access may enable information search and applications (Kuhn & Skuterud, 2004), but a cross-sectional indicator can also proxy household advantage, location or how the dataset was constructed. Complete outcome separation by an access variable is especially important because conventional maximum-likelihood logistic regression then has no finite estimate.

➢ *Analytical Expectations*

The primary expectation is that education, perceived alignment and digital access are associated with employment after adjustment for age, coded gender, coded marital status and province. These are associational expectations, not causal hypotheses. Because the three perception records are exceptionally highly correlated, they are treated as a single internally consistent index rather than as three distinct independent predictors.

III. METHODS

➢ *Data Source, Provenance and Sampling*

The analysis uses the workbook PHD DATA SET(2).xlsx, sheet Cleaned_Data. It contains 1,342 rows and 26 variables, including model-eligibility and quality-control fields. The workbook does not identify the collector, sampling frame, recruitment mode, fieldwork dates, response rate, inclusion and exclusion screening, selection probabilities, non-response adjustment or survey-weight variable. Accordingly, the analysis is unweighted and describes only the supplied records. Provincial coverage is not evidence of probability sampling, and no sampling margin of error is reported.

➢ *Eligibility and Missing Data*

The workbook identifies 1,341 records as primary-model eligible. The prespecified primary model uses complete cases for its actual variables, leaving n = 1,339. Strict recorded-duration eligibility is met by 922 records; the duration model uses all 922. Missing values were not imputed because the provenance and missingness mechanisms are undocumented.

➢ *Outcome Definition*

Employment_Status is coded 1 for employed and 0 for the dataset category labelled unemployed/not employed. The workbook contains no verified job-search or availability-for-work variables and does not document whether informal employment, unpaid family work or intermittent self-employment was counted. Therefore, code 0 cannot be equated with ILO-defined unemployment rather than labour-force inactivity. The paper uses “not employed” for the binary outcome and “recorded non-employment duration” for the supplied duration variable.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26aug705

➢ *Predictors and Score Construction*

Table 1 Predictors and Score Construction

Variable	Observed coding	Analytical treatment
Age	16-35 years	Continuous
Gender	0 and 1	The prior manuscript labels 0 female and 1 male; source codebook unavailable
Marital status	1 and 2	The prior manuscript labels 1 single and 2 married; source codebook unavailable
Education level	No education; secondary; diploma; tertiary	Categorical; no education is reference
Years of education	0-24	Used in sensitivity model, not simultaneously in primary model
Province	Northern; Eastern; Southern; Western Area	Northern Province is reference
Internet access	0/1	No/yes according to prior manuscript; complete separation observed
Curriculum Alignment	1-5	Exact item wording and verbal anchors absent
Job Opportunity Belief	1-5	Exact item wording and verbal anchors absent
Skills Alignment	1-5	Exact item wording and verbal anchors absent

Note. Higher perception values are interpreted as stronger endorsement based on the prior manuscript. This direction should be verified against the original questionnaire. The perception index is the arithmetic mean of the three available item scores when all are observed.

The three items show pairwise correlations of 0.912, 0.899 and 0.871. Cronbach's alpha is 0.961 (n = 1,341). This supports internal consistency for a broad combined index but also demonstrates redundancy; it does not establish content or construct validity.

➢ *Statistical Analysis*

Descriptive statistics, Welch independent-samples t tests, Cohen's d, point-biserial correlations, Pearson chi-square tests and Cramér's V were calculated. The primary employment model uses Firth's Jeffreys-prior bias reduction because internet access perfectly separates employment status: none of the 411 respondents without internet access is recorded as employed. The exact model is:

$$\text{logit}[P(\text{Employed}_i = 1)] = \beta_0 + \beta_1\text{Age}_i + \beta_2\text{Male}_i + \beta_3\text{Married}_i + \beta_4\text{–}\beta_6\text{Education}_i + \beta_7\text{Internet}_i + \beta_8\text{PerceptionIndex}_i + \beta_9\text{–}\beta_{11}\text{Province}_i$$

Reference categories are female (code 0), single (code 1), no education, no internet access and Northern Province. The education sensitivity model substitutes years of schooling for credential level. A diagnostic model containing both schooling measures is reported only to assess overlap; it is not the primary estimand. Dummy-level variance-inflation factors (VIFs) assess collinearity. Model discrimination is described by AUC and calibration by a decile Hosmer-Lemeshow test. Extreme bias-reduced odds ratios caused by separation are treated as data warnings, not precise effect estimates.

For the secondary duration analysis, a Gamma generalised linear model with log link was fitted to strictly eligible positive duration records. Exponentiated coefficients are duration ratios. The model was prespecified in this corrected Methods section and its complete coefficient table is reported below.

➢ *Data-Quality Audit*

Table 2 Data-Quality Audit

Audit condition	Records
Any QC flag	333
No Education with positive years of schooling	289
No Education with recorded speciality conflict	314
Years of schooling greater than age minus five	13
Not employed with zero-month duration	2
Employed with positive duration	0

Note. Flags are not automatically treated as errors because the source records and questionnaire are unavailable. They are disclosed because they affect construct validity and sensitivity interpretation.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26aug705

IV. RESULTS

➢ *Sample Description*

Table 3 Sample Description

Characteristic	n	%
Total records	1,342	100.0%
Employed	415	30.9%
Not employed	927	69.1%
Female, code 0	667	49.7%
Male, code 1	675	50.3%
Northern Province	324	24.1%
Eastern Province	334	24.9%
Southern Province	328	24.4%
Western Area	356	26.5%

Note. Percentages are unweighted characteristics of the supplied records and are not national estimates.

Table 4 Sample Description

Continuous variable	n	Mean	SD
Age (years)	1342	24.76	4.14
Years of education	1342	10.62	4.25
Curriculum alignment	1342	2.84	1.12
Job-opportunity belief	1341	2.82	1.01
Skills alignment	1342	3.03	1.11

➢ *Employment Patterns*

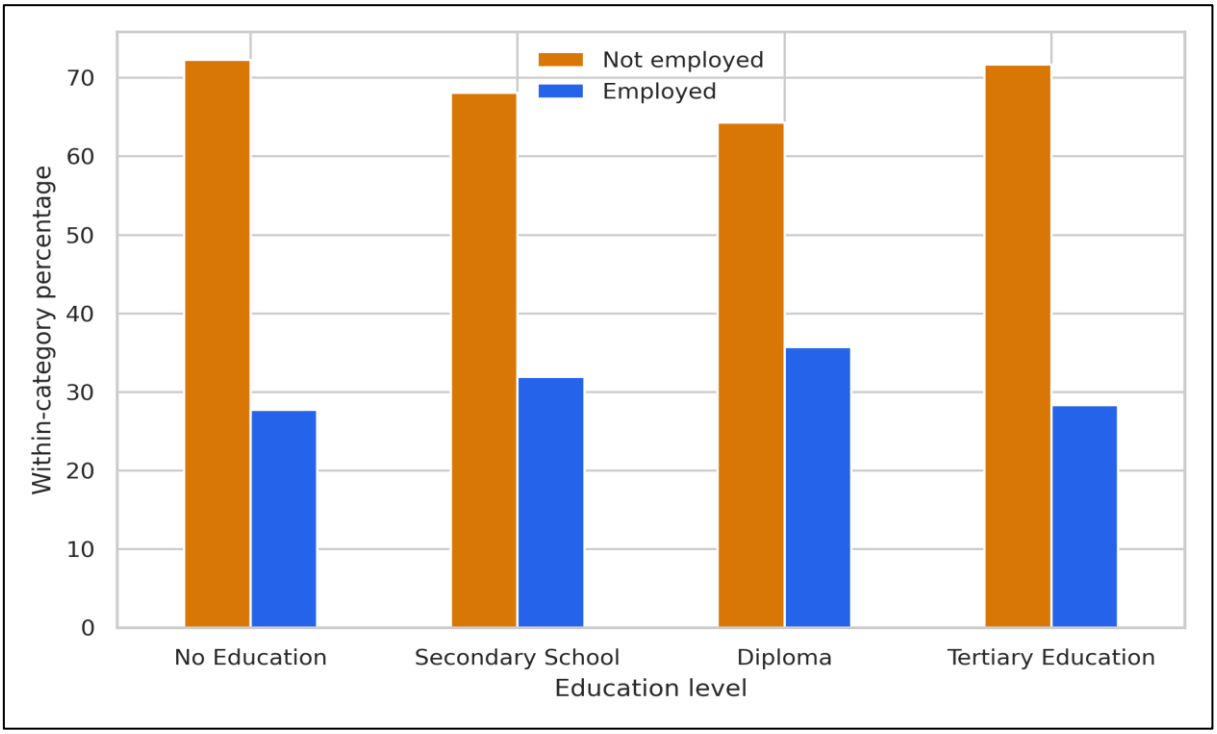

Fig 1 Employment Status within Education Categories in the Supplied Records

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26aug705

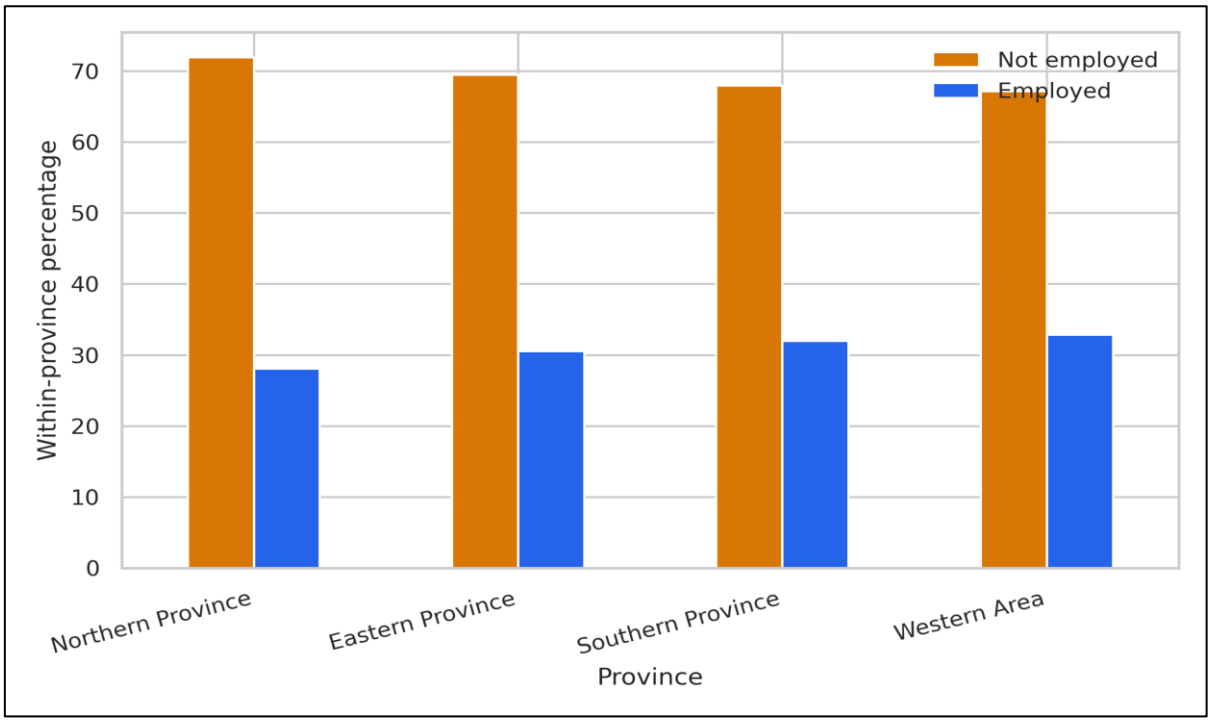

Fig 2 Employment Status within Provinces in the Supplied Records

Table 5 Employment Status within Provinces in the Supplied Records

Variable	χ^2	df	p	Cramér's V	n
Education Level	6.25	3	0.100	0.068	1342
Gender	0.00	1	0.978	0.001	1342
Marital Status	155.62	1	<0.001	0.341	1342
Province	2.05	3	0.561	0.039	1342
Internet Access	263.89	1	<0.001	0.444	1340
Household Head Employment	263.51	1	<0.001	0.443	1341
Household Income	247.64	2	<0.001	0.430	1340
Transport Access	180.09	2	<0.001	0.366	1342
Area of Speciality	10.96	7	0.140	0.090	1342

Note. Education level and province are not significant bivariately at 5%; marital status and digital/household resource indicators show strong sample associations. The latter indicators are structurally dependent and should not be entered together without careful diagnostics.

➢ *Continuous Comparisons and Corrected Correlations*

Table 6 Continuous Comparisons and Corrected Correlations

Variable	Employed M (SD)	Not employed M (SD)	Welch t	p	Cohen d	rpb
Age	26.66 (2.91)	23.91 (4.32)	13.70	<0.001	0.700	0.308
Years of Education	12.36 (2.55)	9.85 (4.61)	12.75	<0.001	0.613	0.273
Curriculum Alignment	3.95 (0.59)	2.34 (0.93)	38.35	<0.001	1.915	0.663
Job Opportunity Belief	3.75 (0.64)	2.41 (0.85)	32.11	<0.001	1.706	0.620
Skills Alignment	4.04 (0.59)	2.58 (0.99)	33.57	<0.001	1.652	0.607

Note. The point-biserial correlations are calculated directly from the raw data. They are not copies of Cohen's d. This corrects the earlier duplicated values.

➢ *Primary Bias-Reduced Employment Model*

The primary complete-case sample is n = 1,339. AUC is 0.935, McFadden pseudo-R^2 is 0.508, classification accuracy is 0.873, sensitivity is 0.822, specificity is 0.896, and the Hosmer-Lemeshow statistic is $\chi^2(8) = 14.56$, $p = 0.068$. Maximum dummy-level VIF is 2.62.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26aug705

Table 7 Primary Bias-Reduced Employment Model

Predictor	B	SE	p	OR	95% CI
Intercept	-7.656	1.736	<0.001	0.000	0.000-0.014
Age	-0.328	0.042	<0.001	0.720	0.663-0.783
Male (vs female)	-0.277	0.180	0.124	0.758	0.532-1.079
Married (vs single)	0.306	0.255	0.231	1.357	0.823-2.237
Secondary School (vs no education)	0.395	0.248	0.110	1.485	0.914-2.412
Diploma (vs no education)	0.713	0.254	0.005	2.040	1.241-3.354
Tertiary Education (vs no education)	0.280	0.250	0.262	1.324	0.811-2.160
Internet access (vs no)	5.313	1.552	<0.001	202.979	9.683-4254.749
Perception index (1-5)	2.908	0.214	<0.001	18.325	12.041-27.889
Eastern Province (vs Northern Province)	0.577	0.248	0.020	1.781	1.095-2.897
Southern Province (vs Northern Province)	0.295	0.250	0.238	1.343	0.823-2.190
Western Area (vs Northern Province)	0.116	0.234	0.620	1.123	0.709-1.778

Note. Firth bias-reduced logistic regression. The internet coefficient remains extreme because no employed record lacks internet access; its wide interval reflects instability. Wald intervals are approximate under bias reduction.

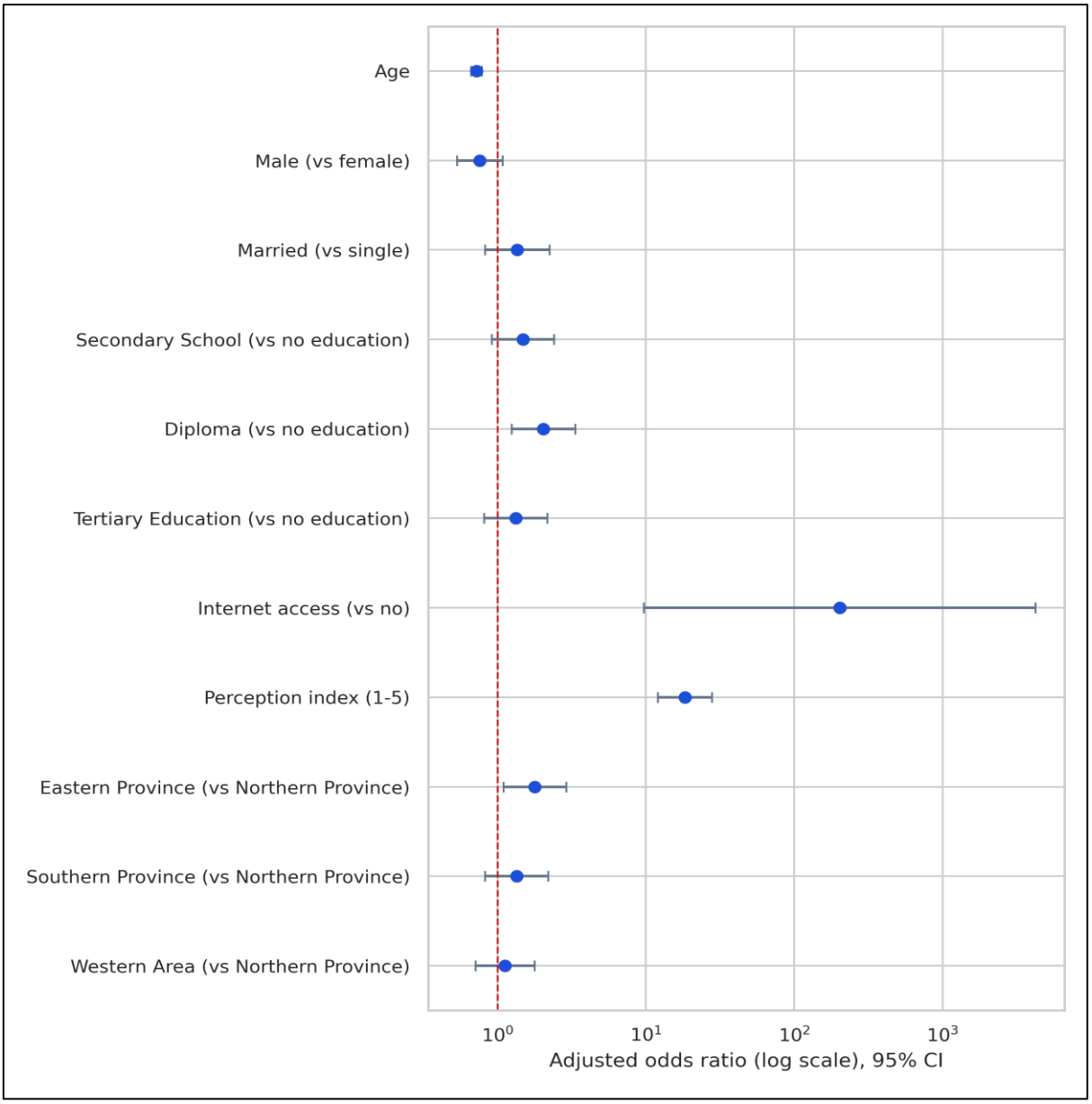

Fig 3 Bias-Reduced Adjusted Odds Ratios for Employment Status

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26aug705

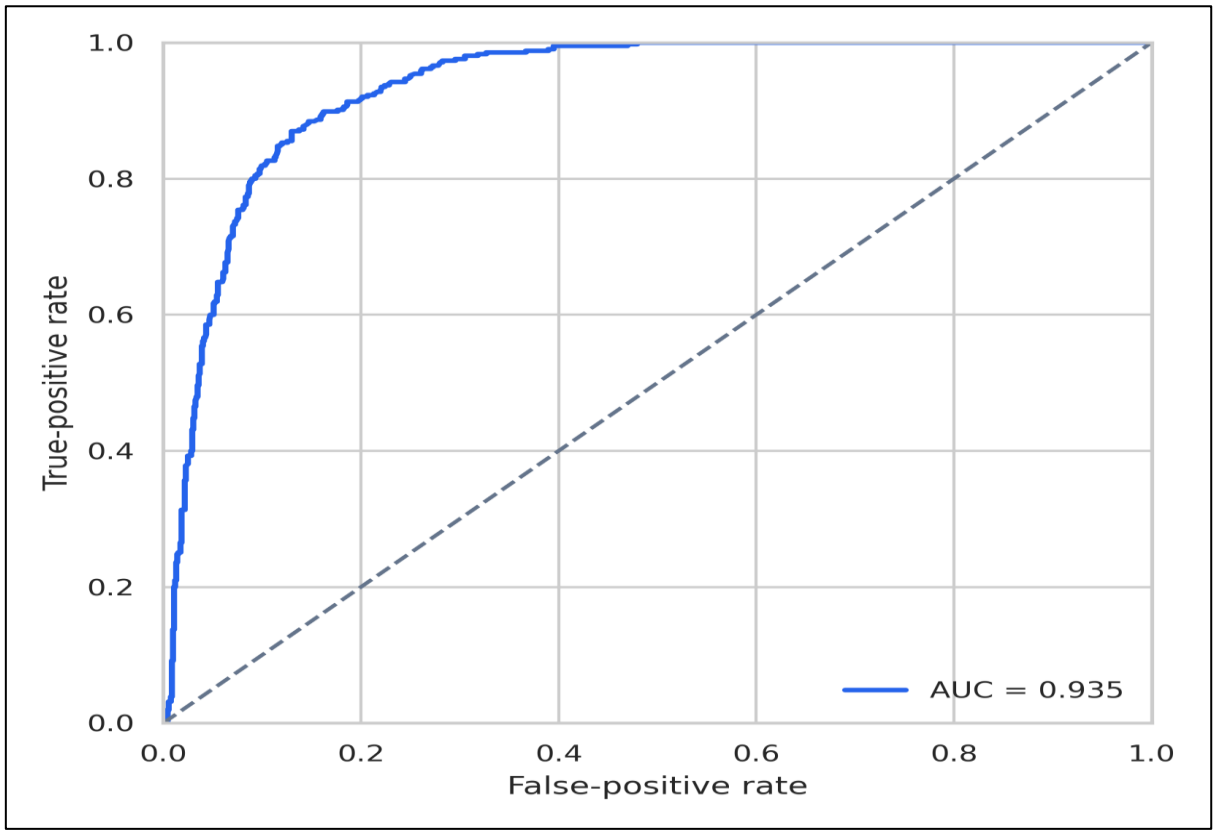

Fig 4 Receiver Operating Characteristic Curve for the Primary Model

➢ *Schooling Sensitivity Analysis*

When years of education replaces credential level, each additional recorded year is inversely associated with employment (OR = 0.722, 95% CI 0.661-0.789, p < 0.001). This reversal is not interpreted as a genuine schooling penalty because 289 "No Education" records have positive schooling years and 13 records exceed age minus five. In the diagnostic model containing both schooling measures, maximum VIF is 2.76; numerical collinearity is moderate, but conceptual inconsistency remains substantial.

Table 8 Schooling Sensitivity Analysis

Predictor	B	SE	p	OR	95% CI
Intercept	-8.014	1.836	<0.001	0.000	0.000-0.012
Age	-0.264	0.043	<0.001	0.768	0.705-0.836
Male (vs female)	-0.143	0.188	0.449	0.867	0.599-1.255
Married (vs single)	0.696	0.278	0.012	2.006	1.163-3.461
Years of education	-0.325	0.045	<0.001	0.722	0.661-0.789
Internet access (vs no)	7.149	1.681	<0.001	1272.701	47.153-34351.119
Perception index (1-5)	3.222	0.229	<0.001	25.073	16.007-39.275
Eastern Province (vs Northern Province)	0.482	0.260	0.064	1.619	0.972-2.697
Southern Province (vs Northern Province)	0.076	0.259	0.770	1.079	0.649-1.792
Western Area (vs Northern Province)	0.018	0.244	0.943	1.018	0.630-1.643

Note. Sensitivity model substitutes years of education for education level. It is not evidence that additional schooling causes lower employment.

➢ *Exploratory Recorded-Duration Model*

Among 922 strictly eligible positive-duration records, mean recorded duration is 14.20 months (SD = 5.80). Exponentiated Gamma-model coefficients are duration ratios.

Table 9 Exploratory Recorded-Duration Model

Predictor	B	SE	p	Duration ratio	95% CI
Intercept	3.317	0.091	<0.001	27.570	23.049-32.977
Age	-0.000	0.004	0.944	1.000	0.992-1.008
Male (vs female)	-0.012	0.020	0.525	0.988	0.951-1.026
Married (vs single)	0.263	0.040	<0.001	1.301	1.203-1.407

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26aug705

Secondary School (vs no education)	0.021	0.031	0.502	1.021	0.961-1.085
Diploma (vs no education)	0.042	0.031	0.184	1.043	0.980-1.109
Tertiary Education (vs no education)	0.023	0.029	0.429	1.023	0.966-1.084
Internet access (vs no)	-0.406	0.028	<0.001	0.667	0.630-0.705
Perception index (1-5)	-0.221	0.015	<0.001	0.802	0.778-0.826
Eastern Province (vs Northern Province)	-0.045	0.030	0.139	0.956	0.901-1.015
Southern Province (vs Northern Province)	-0.028	0.029	0.331	0.973	0.919-1.029
Western Area (vs Northern Province)	0.000	0.031	0.993	1.000	0.942-1.063

Note. Outcome is recorded non-employment duration, not verified ILO unemployment duration. Western Area is not significant relative to Northern Province, correcting the unsupported earlier claim. Curriculum alignment is not modelled separately because the three perception records were combined.

V. DISCUSSIONS

➢ *Principal Findings*

The corrected analysis changes the article's conclusions materially. Only 30.9% of the supplied records are coded employed, not 37.3%. Education level has a weak bivariate association overall, while diploma education is positively associated with employment relative to the problematic "No Education" category in the adjusted model. Tertiary education is not statistically significant after adjustment. The combined perception index is strongly associated with employment, but its magnitude, contemporaneous measurement and very high internal correlations make reverse causality and constructed-data dependence plausible.

The adjusted age coefficient is negative despite a positive unadjusted point-biserial correlation. This sign reversal indicates strong conditioning by other model variables and should not be narrated as a simple age effect. Coded marital status is strong bivariately but not significant in the primary adjusted model. Eastern Province differs positively from Northern Province in the primary model; Southern Province and Western Area do not. These complete provincial and marital-status results replace earlier selective statements.

➢ *Digital Access and Separation*

Internet access perfectly separates employment: all 415 employed respondents have internet access, whereas none of the 411 respondents without access is employed. Household-head employment is almost the same variable (r = 0.998 with internet access). A conventional logistic model therefore produces divergent coefficients and unusable standard errors. Firth estimation yields a finite internet coefficient, but the OR remains enormous and imprecise. This is primarily a warning about data structure, coding or selection—not credible evidence that internet access multiplies population employment odds by approximately 203.

➢ *Measurement and Schooling Contradictions*

The original wording and response anchors for Curriculum Alignment, Job Opportunity Belief and Skills Alignment are absent. Alpha of 0.961 indicates internal consistency when combined, but correlations above 0.87 also show redundancy. Treating the items as three independent constructs would exaggerate apparent dimensionality and destabilise a full regression.

Education level and years of education are not used together in the primary model. Although the diagnostic VIFs are below conventional concern thresholds, coding conflicts make the years variable substantively unreliable: most respondents labelled "No Education" still have positive years of education, and all have a recorded speciality. The negative years coefficient is therefore interpreted as evidence of coding incompatibility, not as a labour-market return.

➢ *Duration Evidence*

The corrected duration analysis is explicitly defined and reported. Internet access and the combined perception index are associated with shorter recorded duration, while coded marriage is associated with longer duration. Provincial duration ratios, including Western Area, are not significant. Because job search, availability and inactivity are not documented, these estimates concern a recorded duration field rather than a validated unemployment spell.

➢ *Implications*

The dataset is suitable for demonstrating statistical diagnostics and for generating hypotheses, but it is not presently suitable for strong national policy claims. Before substantive publication, investigators should recover the sampling protocol, questionnaire, codebook, original uncleaned data and provenance trail; reconcile employment and labour-force status; verify the treatment of informal and unpaid work; resolve schooling contradictions; and investigate whether highly dependent resource variables were generated or recoded from a common source.

VI. CONCLUSION

This verified reanalysis provides a clean account of what the supplied records can and cannot support. Bias-reduced modelling identifies conditional associations for diploma education, age, the combined perception index and Eastern Province. Internet access perfectly separates employment and cannot be interpreted as a stable effect. The measurement, sampling and coding gaps rule out claims of national representativeness, causality or a validated unemployment rate. The strongest conclusion is methodological: credible inference requires transparent provenance, defensible labour-force classification, validated measurement and models designed for separation and redundancy.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26aug705

LIMITATIONS AND REQUIRED VERIFICATION

- Collector, sampling frame, recruitment procedure, response rate and survey weights are absent.
- Employment status does not distinguish active unemployment from inactivity and does not document informal work, unpaid family work, self-employment, job search or availability.
- Perception-item wording and verbal response anchors are absent; alpha does not establish validity.
- Complete separation and near-identity among resource variables limit multivariable estimation.
- Education coding conflicts undermine years-of-schooling interpretation.
- Cross-sectional contemporaneous measures do not identify causal, mediational or temporal mechanisms.
- Bias-reduced Wald intervals and the exploratory duration model should be confirmed in validated source data.

REFERENCES

[1]. African Development Bank. (2012). Youth employment in Africa. African Development Bank.
[2]. Becker, G. S. (1964). Human capital: A theoretical and empirical analysis, with special reference to education. University of Chicago Press.
[3]. Card, D. (1999). The causal effect of education on earnings. In O. Ashenfelter & D. Card (Eds.), Handbook of labor economics (Vol. 3, pp. 1801-1863). Elsevier.
[4]. Doeringer, P. B., & Piore, M. J. (1971). Internal labor markets and manpower analysis. Heath.
[5]. Filmer, D., & Fox, L. (2014). Youth employment in Sub-Saharan Africa. World Bank.
[6]. International Labour Organization. (2022). Global employment trends for youth 2022. ILO.
[7]. Kuhn, P., & Skuterud, M. (2004). Internet job search and unemployment durations. American Economic Review, 94(1), 218-232.
[8]. McGuinness, S., & Bennett, J. (2007). Overeducation in the graduate labour market: A quantile regression approach. Economics of Education Review, 26(5), 521-531.
[9]. Psacharopoulos, G., & Patrinos, H. A. (2018). Returns to investment in education: A decennial review of the global literature. Education Economics, 26(5), 445-458.
[10]. Schultz, T. W. (1961). Investment in human capital. American Economic Review, 51(1), 1-17.
[11]. Spence, M. (1973). Job market signaling. Quarterly Journal of Economics, 87(3), 355-374.
[12]. Verhaest, D., & Omey, E. (2006). The impact of overeducation and its measurement. Social Indicators Research, 77(3), 419-448.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26aug705

APPENDIX

APPENDIX A. RESOLUTION OF THE SEVEN REVIEW COMMENTS

Reviewer issue	Correction made
1. Sampling and representativeness	Documented missing provenance elements; removed representativeness and margin-of-error claims; stated analysis is unweighted.
2. Employment definition	Relabelled code 0 as not employed; explained why unemployment cannot be separated from inactivity; listed undocumented work classifications.
3. Perception measures	Reported observed scoring, absence of wording/anchors, alpha = 0.961 and item correlations; constructed a transparent mean index.
4. Exact model and references	Specified the full equation and reference categories; reported all marital and provincial estimates.
5. Education overlap	Used education level in the primary model and years in sensitivity analysis; reported diagnostic VIFs and coding conflicts.
6. Duplicated/inconsistent statistics	Recomputed Cohen's d and point-biserial r directly; replaced sample counts; removed invalid all-sample duration summary.
7. Duration model	Prespecified and reported an exploratory Gamma log-link model; removed unsupported curriculum and Western Area claims.