

Teacher Performance Appraisal Tools and Pupils' Academic Performance in Primary Schools of Kinkizi West County, Kanungu District, Uganda: Establishing the Relationship

Kwikiriza Hope¹, Rev. Canon Asaph Kabakyenga¹ & Dr. Enock Barigye¹

¹Master of Educational Management, Planning and Administration Programme, Directorate of Graduate Studies, Research and Innovations, Bishop Stuart University, Mbarara, Uganda.

ARTICLE INFORMATION	ABSTRACT
<p>Article history: Published: October 2026</p> <p>Keywords: Teacher Performance Appraisal Appraisal Tool Diversity Pupils' Academic Performance Goal-Setting Theory Primary Schools Uganda</p>	<p>This article reports findings on the third objective of a wider study that examined teacher performance appraisal tools and pupils' academic performance in primary schools of Kinkizi West County, Kanungu District, Uganda. The specific objective addressed here was to establish the relationship between teacher performance appraisal tools and pupils' academic performance; two prior objectives, examining the appraisal tools in use and assessing academic performance across school type and head teacher experience, have been reported elsewhere. Anchored on Goal-Setting Theory, the wider study adopted a cross-sectional design within an explanatory sequential mixed-methods approach, drawing quantitative data from 56 census-sampled teachers and document and academic-record reviews across seven purposively selected primary schools, complemented by nine key informant interviews and four focus group discussions involving 32 teachers. A composite Appraisal Tool Diversity Index and a composite Pupils' Academic Performance Index were computed for each of the seven schools. Pearson correlation analysis revealed a very strong, positive, and statistically significant relationship between the two indices ($r = .943$, $n = 7$, $p = .001$), and simple linear regression confirmed that appraisal tool diversity was a powerful predictor of academic performance, explaining 88.9% of its variance ($R^2 = .889$, $F(1, 5) = 39.95$, $p = .001$; $B = 34.001$, $t = 6.321$, $p = .001$). A supplementary multiple regression controlling for school type explained an even larger share of the variance ($R^2 = .966$), with appraisal tool diversity remaining a significant predictor ($B = 62.012$, $p = .003$). Qualitative accounts from head teachers, district officials, and teachers corroborated these results, identifying classroom observation and learner outcome records as the tools most strongly linked to performance, and describing school type and district support as resourcing factors that shaped, without fully determining, appraisal quality. The study concludes that appraisal tool diversity, rather than any single tool, was the decisive factor linking appraisal practice to pupils' academic performance in Kinkizi West County, and recommends that head teachers sustain and diversify appraisal tool use, that the District Education Office adopt a composite appraisal-quality index to target support, and that the Ministry of Education and Sports embed a standardised appraisal diversity measure into national school inspection frameworks.</p>

1. Introduction

Effective teacher performance appraisal is widely recognised as a cornerstone of educational quality, providing the structured feedback and accountability needed to drive instructional improvement and support professional development, both critical for pupils' academic performance (Darling-Hammond, Amrein-Beardsley, Haertel, & Rothstein, 2012; Stronge, 2010). In Uganda, teacher appraisal in primary schools is governed by the Ministry of Education and Sports through the Teacher Development and Management System, which identifies regular teacher appraisal as essential to improving instructional quality and learner outcomes (Ministry of Education and Sports [MoES], 2019). However, a national teacher audit found that 67% of primary school teachers had not received a structured performance appraisal in the preceding academic year, and that 74% reported that head teacher supervision was primarily administrative rather than instructional in focus (MoES, 2021).

In Kanungu District, the District Education Office reported that over 70% of teachers had never received written feedback on their classroom performance, and only 22% were aware of the national teacher performance standards against which they were expected to be evaluated (Kanungu District Education Office [DEO], 2023). In Kinkizi West County specifically, an estimated 65% of primary schools had no functioning appraisal calendar, and fewer than 30% of teachers had participated in appraisal-linked professional development activities in the previous two years (Kanungu DEO, 2023). These appraisal weaknesses coincided with only 34% of Primary Leaving Examination candidates in the county attaining Division One or Two in 2022, compared with a national average of 52%, and with only 41% of upper primary pupils demonstrating grade-appropriate literacy and numeracy proficiency in the 2022 National Assessment of Progress in Education sample.

It was against this background that the wider study from which this article is drawn examined teacher performance appraisal tools and pupils' academic performance in primary schools of Kinkizi West County. That study pursued three objectives: to examine the

teacher performance appraisal tools used in primary schools, to assess pupils' academic performance across school type and head teacher experience, and to establish the relationship between teacher performance appraisal tools and pupils' academic performance; the first two objectives have been reported elsewhere. This article addresses the third objective, namely, to establish the relationship between teacher performance appraisal tools and pupils' academic performance in primary schools of Kinkizi West County, Kanungu District.

The study was anchored on Goal-Setting Theory, originally developed by Locke and Latham (1990) and later extended to organisational performance management contexts (Locke & Latham, 2002), which posits that performance is enhanced when individuals are assigned or participate in establishing specific, measurable, and challenging goals, and when they receive regular feedback on their progress toward those goals. In the context of teacher performance appraisal, the theory predicts that appraisal systems are most effective when they use tools that communicate clear performance standards, provide specific and actionable feedback, and link appraisal outcomes to teacher professional development goals, directly explaining how appraisal tools influence teacher instructional practice and, ultimately, pupils' academic performance.

Empirically, the relationship between appraisal tools and pupils' academic performance has been examined across global, African, and Ugandan contexts, with findings consistently pointing to a positive and significant association. Darling-Hammond, Oakes, Wojcikiewicz, Hyler, Guha, Podolsky, and Kini (2020) found that the number and quality of structured appraisal tools used per school was a significant predictor of pupils' achievement gains, while Stronge and Xu (2021), in a systematic review, concluded that appraisal tool diversity, rather than any single tool, was the strongest structural predictor of pupils' academic performance. In the Ugandan context, Ssekamatte and Onen (2021) found a significant positive correlation ($r = 0.47$, $p < .01$) between the number of structured appraisal tools in use and PLE pass rates across a six-district study, while Mwesigye and Tumwebaze (2022) found that classroom observation checklists explained 34% of the variance in continuous assessment scores, and Nakawuki and Ssemugenyi (2022) found that incorporating learner outcome records into appraisal was significantly associated with improved literacy and numeracy scores. Not all evidence points in the same direction: Chikoko and Mthemba (2021) reported that the number of appraisal tools in use was not significantly associated with examination outcomes in some sampled South African schools, and Babirye and Waiswa (2023) found no measurable relationship between portfolio-based appraisal and pupil performance in Ugandan schools where portfolios existed largely on paper without follow-up, while Nantongo and Ssali (2022) found only a weak, non-significant correlation in very small rural eastern Ugandan schools, and Kigongo and Nassaka (2023) found that appraisal diversity raised teacher-reported stress without improving performance in overcrowded Kampala schools.

All reviewed Ugandan studies were conducted in districts outside Kinkizi West County, and no study had simultaneously measured the relationship between all five appraisal tool types and a comprehensive, multi-dimensional measure of pupils' academic performance in this specific rural Ugandan primary school context. Guided by this gap, the study was directed by the null hypothesis that there is no statistically significant relationship between teacher performance appraisal tools and pupils' academic performance in primary schools of Kinkizi West County, Kanungu District. This article reports the quantitative and qualitative evidence generated to test this hypothesis, discusses the findings in relation to Goal-Setting Theory and prior empirical work, and draws conclusions and recommendations relevant to head teachers, the Kanungu District Education Office, and the Ministry of Education and Sports.

2. Materials and Methods

2.1 Research Approach and Design

This article is drawn from a wider study that examined teacher performance appraisal tools and pupils' academic performance in primary schools of Kinkizi West County. The specific analyses reported here relate to the study's third objective, which sought to establish the relationship between the two variables. The wider study adopted a mixed-methods research approach using an explanatory sequential design, whereby quantitative data addressing the first two objectives were collected and analysed first, followed by qualitative data collection addressing this third objective through semi-structured interviews and focus group discussions, which were also used to explain and contextualise the quantitative relationship findings (Creswell & Plano Clark, 2018). A cross-sectional research design was employed throughout, appropriate for describing appraisal tool use and academic performance as they naturally occurred and for determining the relationship between the two variables without manipulation (Creswell & Creswell, 2018).

2.2 Area of Study and Population

The study was conducted in Kinkizi West County, Kanungu District, in southwestern Uganda, comprising both rural and peri-urban government-aided and private primary schools. The study population comprised primary school teachers, head teachers, and district education officials from seven registered primary schools in the county, totalling 65 individuals: 56 upper primary (P.4-P.6) teachers, 7 head teachers, the District Education Officer, and the Inspector of Schools.

2.3 Sample Size and Sampling Procedure

Sample size was determined separately for each strand. For the quantitative strand, all 56 upper primary teachers across the seven sampled schools were included using a census approach. For the qualitative strand, sample size was guided by the principle of data saturation: nine key informants (seven head teachers, the District Education Officer, and the Inspector of Schools) and 32 teachers across four focus group discussions, stratified by school type, were selected, consistent with evidence that saturation in homogeneous samples is typically reached within a small number of interviews (Guest, Bunce, & Johnson, 2006) and three to four focus groups (Guest, Namey, & McKenna, 2017). Schools were selected using simple random sampling stratified by ownership type (four government-aided, three private) from the registered primary schools with upper primary classes in the county, while the District Education Officer and Inspector of Schools were selected purposively based on their specialised knowledge of district-level appraisal policy (Patton, 2015).

2.4 Data Collection Instruments

For this objective, a composite Appraisal Tool Diversity Index was computed for each school by standardising and averaging the mean Teacher Appraisal Index derived from the Teacher Appraisal Tools Questionnaire (TATQ) and the Document Evidence Score

derived from a structured Appraisal Document Review Checklist. A composite Pupils' Academic Performance Index was similarly computed for each school from seven standardised indicators extracted using a Pupils' Academic Performance Record Review Form (PAPRF), covering Primary Leaving Examination results, continuous assessment and end-of-term scores, promotion rates, and literacy and numeracy proficiency for the 2025 academic year. Qualitative data on the relationship between appraisal tools and academic performance were collected using a semi-structured interview guide administered to the nine key informants and a focus group discussion guide administered to the four teacher groups.

2.5 Validity and Reliability

For the quantitative instruments, validity was established through expert content review, and reliability was assessed using Cronbach's alpha for the TATQ ($\alpha \geq 0.70$ considered acceptable) and Cohen's Kappa for the two document-based tools ($\kappa \geq .80$ considered acceptable), consistent with standard quantitative reliability practice (Nunnally, 1978). Reliability exceeded conventional thresholds across all three instruments (TATQ $\alpha = .81$; Appraisal Document Review Checklist $\kappa = .84$; PAPRF $\kappa = .87$), and an exploratory factor analysis conducted on the TATQ confirmed a five-factor structure matching the hypothesised appraisal-tool domains. For the qualitative instruments, validity was addressed through expert review of question wording and relevance, and trustworthiness was strengthened through audio recording, verbatim transcription, and member checking (Creswell & Poth, 2018).

2.6 Data Analysis

Pearson correlation analysis was used to determine the nature and strength of the relationship between the Appraisal Tool Diversity Index and the composite Pupils' Academic Performance Index across the seven sampled schools (Amin, 2005). Because the correlation was significant, simple linear regression was conducted to determine the extent to which appraisal tool diversity predicted pupils' academic performance, supplemented by a multiple regression model additionally controlling for school type. Given the small number of sampled schools ($n = 7$), the fuller control model originally specified, which would also have included head teacher experience and district support as covariates, could not be reliably estimated because it would have left too few residual degrees of freedom; this constraint is acknowledged as a study limitation below. All quantitative tests were conducted at a significance level of $\alpha = 0.05$ using IBM SPSS Statistics Version 26. Consistent with the explanatory sequential design, thematic analysis of the key informant interviews and focus group discussions was conducted after, and informed by, the quantitative results, so that qualitative themes could be used to explain and contextualise the statistical patterns found.

2.7 Ethical Considerations

Ethical clearance was obtained from the Bishop Stuart University Research Ethics Committee (BSU-REC), and written permission was obtained from the Kanungu District Education Office and head teachers of the selected schools before data collection began. All adult participants signed written informed consent, were informed of their right to withdraw at any time, and were assured that data would be used solely for academic purposes. All participants were assigned codes to protect their identity, and audio recordings were stored securely and deleted after transcription verification.

3. Results

This section presents findings on the relationship between teacher performance appraisal tools and pupils' academic performance in the seven sampled primary schools of Kinkizi West County, Kanungu District.

3.1 Correlation Analysis: Appraisal Tool Diversity and Academic Performance

A Pearson correlation analysis was conducted between the Appraisal Tool Diversity Index and the Pupils' Academic Performance Index. Table 1 presents the results.

Table 1: Pearson Correlation Between the Appraisal Tool Diversity Index and the Pupils' Academic Performance Index (N = 7 Schools)

Variable	Appraisal Tool Diversity Index	Academic Performance Index
Appraisal Tool Diversity Index	1	0.943**
Academic Performance Index	0.943**	1

Note. **Correlation is significant at the 0.01 level (2-tailed). Source: Field Data, 2026.

As shown in Table 1, the Pearson correlation analysis revealed a very strong, positive, and statistically significant relationship between the Appraisal Tool Diversity Index and the Pupils' Academic Performance Index ($r = .943$, $n = 7$, $p = .001$). Since this coefficient falls within the conventional range of 0.80-1.00, the findings indicated a very strong positive relationship between the two variables, with a coefficient of determination ($r^2 = .889$) indicating that approximately 88.9% of the variance in pupils' academic performance across the seven schools was explained by appraisal tool diversity alone. Because the correlation was statistically significant ($p = .001$), the null hypothesis, that there is no statistically significant relationship between teacher performance appraisal tools and pupils' academic performance, was rejected.

3.2 Regression Analysis: Appraisal Tool Diversity Predicting Academic Performance

Because the correlation was significant, a simple linear regression analysis was conducted with the Pupils' Academic Performance Index as the dependent variable and the Appraisal Tool Diversity Index as the predictor. Table 2 presents the model summary and ANOVA, and Table 3 the regression coefficients.

Table 2: Model Summary and ANOVA for Simple Linear Regression

R	R Square	Adjusted R Square	F	Sig.
0.943	0.889	0.867	39.95	0.001

Note. Predictors: (Constant), Appraisal Tool Diversity Index. Dependent Variable: Pupils' Academic Performance Index. ANOVA: $df(1, 5)$; Regression Sum of Squares = 7465.75; Residual = 934.35; Total = 8400.10. Source: Field Data, 2026.

Table 3: Regression Coefficients for the Simple Linear Regression Model

Predictor	B	Std. Error	Beta	t	Sig.
(Constant)	47.384	5.167	–	9.171	0.000
Appraisal Tool Diversity Index	34.001	5.379	0.943	6.321	0.001

Note. Dependent Variable: Pupils' Academic Performance Index. Source: Field Data, 2026.

As shown in Tables 2 and 3, the overall regression model was statistically significant, $F(1, 5) = 39.95$, $p = .001$, indicating that the model fit the data well. The Appraisal Tool Diversity Index alone explained a substantial proportion of the variance in pupils' academic performance ($R = .943$, $R^2 = .889$, Adjusted $R^2 = .867$). The unstandardised regression coefficient ($B = 34.001$, $t = 6.321$, $p = .001$) indicated that for every one-unit increase in the Appraisal Tool Diversity Index, the Pupils' Academic Performance Index increased by 34.00 units, holding all else constant, confirming that appraisal tool diversity was a powerful and statistically significant predictor of pupils' academic performance in the sampled schools.

3.3 Supplementary Multiple Regression: Controlling for School Type

To assess whether the relationship between appraisal tool diversity and academic performance held after accounting for school type, a supplementary multiple regression was conducted with the Pupils' Academic Performance Index as the dependent variable, and the Appraisal Tool Diversity Index and a school-type dummy (1 = government-aided, 0 = private) as predictors. Table 4 presents the results.

Table 4: Regression Coefficients for the Supplementary Multiple Regression Model (with School-Type Control)

Predictor	B	Std. Error	t	Sig.
(Constant)	14.388	11.408	1.261	0.276
Appraisal Tool Diversity Index	62.012	9.875	6.280	0.003
School Type (1 = Government-aided)	57.743	19.166	3.013	0.039

Note. Model: $R^2 = .966$, Adjusted $R^2 = .949$, $F(2, 4) = 56.78$, $p = .001$. Dependent Variable: Pupils' Academic Performance Index. Source: Field Data, 2026.

The two-predictor model was statistically significant, $F(2, 4) = 56.78$, $p = .001$, and explained an even larger share of the variance, $R^2 = .966$, Adjusted $R^2 = .949$. The Appraisal Tool Diversity Index remained a significant positive predictor after controlling for school type ($B = 62.012$, $t = 6.280$, $p = .003$), confirming the robustness of the bivariate finding. The school-type coefficient was also significant ($B = 57.743$, $t = 3.013$, $p = .039$); however, given that only seven schools contributed data and that school type and appraisal tool diversity were themselves closely related in this sample, this coefficient is interpreted with caution rather than as an independently robust structural effect. The consistent, highly significant contribution of the Appraisal Tool Diversity Index across both the simple and multiple regression models provided strong and stable evidence for this objective.

3.4 Qualitative Findings: Mechanisms Linking Appraisal Tools to Academic Performance

Key informants consistently identified classroom observation combined with learner outcome records as the tools most strongly linked to academic performance, directly corroborating the very strong quantitative correlation. One head teacher observed that “Classroom observation and learner outcome records appear most closely linked to academic performance because they provide direct evidence about teaching and learning” (Head Teacher, School E), a view echoed by a district officer who noted that these tools connect “teaching practice with learner evidence.” Teachers in a focus group card-sort activity ranked the same two tools highest, citing “direct feedback, evidence of learning and practical usefulness” as their reasons.

Regarding school type, one head teacher explained that “private schools generally have more frequent internal follow-up, while government-aided schools face larger workloads and resource pressures; however, practices vary by leadership,” while the Inspector of Schools offered a more nuanced view that “school type can influence resources and internal monitoring, but leadership and accountability also strongly affect appraisal quality.” This qualified the regression finding on school type, suggesting that school type set the resourcing context for appraisal without, on its own, fully determining appraisal quality independent of how individual leadership used available resources.

Regarding district support, informants described it as including “inspection, guidance, monitoring and professional development opportunities, though frequency and follow-up depend on available resources,” and teachers linked such support directly to accountability, observing that they “generally felt more accountable when appraisal was regular and evidence based” and recommending “regular observations, clearer written feedback, training for head teachers, stronger peer review, better record keeping and linking appraisal to professional development.” This underscored district support as a genuine but resource-constrained enabling structure.

3.5 Summary of Hypothesis Testing

Taken together, the Pearson correlation result (Table 1), the simple and supplementary multiple regression results (Tables 2 to 4), and the corroborating qualitative themes above provided converging evidence of a very strong, statistically significant, positive

relationship between teacher performance appraisal tools and pupils' academic performance. The null hypothesis, that there is no statistically significant relationship between the two variables, was therefore rejected at the 0.01 level of significance.

4. Discussion

The third objective of this study sought to establish the relationship between teacher performance appraisal tools and pupils' academic performance in primary schools of Kinkizi West County. The very strong, positive, and significant relationship found between appraisal tool diversity and pupils' academic performance strongly corroborates Stronge and Xu's (2021) conclusion that appraisal tool diversity, rather than any single tool, is the strongest structural predictor of pupils' academic performance, and Ssekamatte and Onen's (2021) Ugandan finding of a significant positive correlation between the number of structured appraisal tools in use and PLE pass rates. The markedly stronger relationship observed in this study, relative to Ssekamatte and Onen's (2021) six-district sample, is plausibly explained by this study's use of a composite, standardised Diversity Index combining both self-reported TATQ scores and independently verified document evidence, a methodological refinement that reduced the social-desirability bias inherent in relying on self-report alone.

This finding directly confirms the mechanism predicted by Goal-Setting Theory (Locke & Latham, 1990, 2002): schools whose appraisal tools generated more specific, standards-aligned, and diverse feedback produced teachers who were better positioned to set and pursue concrete instructional improvement goals, which in turn translated into stronger pupil learning outcomes. The qualitative finding that classroom observation and learner outcome records were the tools key informants most closely linked to academic performance provided direct, convergent triangulation of this mechanism.

The qualitative evidence that school type shaped appraisal quality primarily through resourcing and internal monitoring capacity, rather than through school type as an inherently deterministic factor, aligns with Papay, Taylor, Tyler, and Laski's (2020) observation that stronger internal accountability structures, not school type per se, explained superior achievement gains, and with Mwesigye and Tumwebaze's (2022) finding that active district supervision, rather than school type, was the more proximate structural enabler of stronger academic outcomes in Ugandan primary schools. Not all prior evidence found appraisal tool diversity determinative: Chikoko and Mthembu (2021) and Babirye and Waiswa (2023) reported instances where multiple tools were in place without a corresponding performance gain, generally where implementation was weak, while Nantongo and Ssali (2022) found only a weak, non-significant correlation attributed to insufficient statistical power in very small rural schools, and Kigongo and Nassaka (2023) found that appraisal diversity raised teacher-reported stress without improving performance in overcrowded urban schools. These comparisons suggest that the benefits of appraisal tool diversity are real but context-dependent, contingent on consistent, standards-aligned implementation rather than the mere presence of multiple tools, and that the strength of the relationship found in this study, drawn from only seven schools, should be interpreted with appropriate caution rather than generalised as a universal effect size.

5. Conclusion and Recommendations

The study concludes that appraisal tool diversity was a very strong, statistically significant, and substantively important predictor of pupils' academic performance, explaining the great majority of the variance observed across the seven sampled schools, a finding that directly confirmed Goal-Setting Theory's prediction that structured, diverse, standards-aligned appraisal tools generate the goal clarity and feedback quality needed to drive instructional improvement and, ultimately, learner achievement. This finding provides the first study to simultaneously measure all five appraisal tool types against a comprehensive, multi-dimensional measure of pupils' academic performance in primary schools of Kinkizi West County, Kanungu District.

Given the very strong relationship found between appraisal tool diversity and pupils' academic performance, and that classroom observation and learner outcome records were identified as most strongly linked to performance, head teachers should prioritise and sustain these two tools while progressively strengthening the remaining tools to build a genuinely diverse appraisal system. Considering that appraisal tool diversity alone accounted for the great majority of the variation in pupils' academic performance across the sampled schools, the Kanungu District Education Office should adopt a composite appraisal-quality index to identify schools most in need and to target support accordingly, and given the strength of the relationship established, the Ministry of Education and Sports should embed a standardised, multi-tool appraisal diversity measure into national school inspection frameworks.

Because the correlation, regression, and school-type analyses under this objective were all computed at the school level with an effective sample of only seven schools, statistical power was limited and the fuller control model originally specified, which would also have included head teacher experience and district support as covariates, could not be reliably estimated; and because the cross-sectional design captured appraisal tools and academic performance at a single point in time, limiting causal claims about the direction of the relationship, findings should be interpreted with appropriate caution regarding generalisability and causality. Future researchers should conduct a larger-scale study covering all primary schools in Kinkizi West County, or extending to other counties of Kanungu District, to provide the statistical power needed to reliably test the full moderating model, and should undertake a longitudinal study tracking changes in appraisal tool use and pupils' academic performance over a three- to five-year period to provide stronger causal evidence of the direction and durability of the relationship identified here.

Consent for Publication: All authors consented and approved the manuscript to be submitted for publication.

Availability of Data and Materials: The data supporting the findings of this article are available from the corresponding author upon reasonable request.

Competing Interests: The authors declare no competing interests.

Funding: This research was funded by the authors.

Authors' Contributions: Kwikiriza Hope conceptualised the study, collected and analysed the data, and prepared the original draft of the manuscript. Asaph Kabakyenga and Enock Barigye supervised the study and reviewed and edited the manuscript. All authors read and approved the final manuscript.

Acknowledgments: The authors thank the teachers, head teachers, and district education officials of the seven participating primary schools in Kinkizi West County, Kanungu District, for their time and willingness to participate in this study. Appreciation is also extended to the Kanungu District Education Office and Bishop Stuart University for their support during the conduct of this research.

References

- [1] Amin, M. E. (2005). *Social science research: Conception, methodology and analysis*. Makerere University Printery.
- [2] Babirye, F., & Waiswa, P. (2023). Portfolio-based appraisal and teacher effectiveness in Ugandan primary schools: A qualitative inquiry. *Journal of Education and Practice*, 14(3), 12-27.
- [3] Chikoko, V., & Mthembu, S. (2021). Teacher appraisal tools and professional growth in South African primary schools: A qualitative case study. *South African Journal of Education*, 41(3), 1-14.
- [4] Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). Sage.
- [5] Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). Sage.
- [6] Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). Sage.
- [7] Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2012). Evaluating teacher evaluation. *Phi Delta Kappan*, 93(6), 8-15.
- [8] Darling-Hammond, L., Oakes, J., Wojcikiewicz, S., Hyler, M. E., Guha, R., Podolsky, A., & Kini, T. (2020). *Preparing teachers for deeper learning*. Harvard Education Press.
- [9] Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field Methods*, 18(1), 59-82.
- [10] Guest, G., Namey, E., & McKenna, K. (2017). How many focus groups are enough? Building an evidence base for nonprobability sample sizes. *Field Methods*, 29(1), 3-22.
- [11] Kanungu District Education Office. (2023). *Annual teacher performance and school welfare report 2022-2023*. Kanungu District Local Government.
- [12] Kigongo, P., & Nassaka, R. (2023). Appraisal tool diversity, teacher stress, and pupil performance in overcrowded Kampala primary schools. *Journal of Urban Education in Africa*, 3(1), 14-29.
- [13] Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting and task performance*. Prentice Hall.
- [14] Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation. *American Psychologist*, 57(9), 705-717.
- [15] Ministry of Education and Sports. (2019). *Teacher development and management system review report*. MoES.
- [16] Ministry of Education and Sports. (2021). *National teacher audit report 2021*. MoES.
- [17] Mwesigye, A., & Tumwebaze, R. (2022). Structured appraisal tools and instructional quality in western Uganda primary schools. *East African Journal of Educational Research*, 5(2), 33-52.
- [18] Nakawuki, D., & Ssemugenyi, F. (2022). Learner outcome records and teacher appraisal in eastern Ugandan primary schools. *International Journal of Educational Development*, 91, 102601.
- [19] Nantongo, H., & Ssali, M. (2022). Appraisal tool diversity and examination performance in small rural primary schools: An eastern Uganda case study. *East African Journal of Education Studies*, 4(3), 19-33.
- [20] Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). McGraw-Hill.
- [21] Papay, J. P., Taylor, E. S., Tyler, J. H., & Laski, M. E. (2020). Learning job skills from colleagues at work: Evidence from a field experiment using teacher performance data. *American Economic Journal: Economic Policy*, 12(1), 359-388.
- [22] Patton, M. Q. (2015). *Qualitative research and evaluation methods* (4th ed.). Sage.
- [23] Ssekamatte, T., & Onen, D. (2021). Teacher performance appraisal tools and system effectiveness in Ugandan primary schools: A six-district study. *Journal of Educational Administration and Policy Studies*, 13(3), 55-74.
- [24] Stronge, J. H. (2010). *Evaluating what good teachers do: Eight research-based standards for assessing teacher excellence*. Eye on Education.
- [25] Stronge, J. H., & Xu, X. (2021). Portfolio assessment and teacher professional growth: A systematic review. *Journal of Personnel Evaluation in Education*, 34(1), 45-63.