RESEARCH ARTICLE

JENER Journal of Empirical and Non-Empirical Research
Volume 2, Issue 8, Page 522-532 | Article No. 57
DOI: 10.4898/jener.v2i8.a57
ISSN: 7895-9216
Journal homepage: https://jenerjournal.com
Journal email: editor@jenerjournal.com

JENER Journal
JENER Journal of Empirical and Non-Empirical Research

Teacher Leadership Competence in Geographically Isolated Schools: A Multilevel Comparative Study of Kabugao District I and Santa Fe District, Northern Philippines

Let-Let B. Sario[1] & Roda A. Balicdang[2]

[1,2]Northeastern College

ARTICLE INFORMATION

Article history:
Published: August 2026

Keywords:
Teacher Leadership
Remote Schools
Multilevel Analysis
Measurement Invariance
Kabugao
Santa Fe
Northern Philippines

ABSTRACT

Teacher leadership had particular importance in geographically isolated schools because small staff complements, multigrade assignments, travel constraints, and close community relationships had expanded the work teachers performed beyond classroom instruction. This comparative cross-sectional study examined teacher leadership competence in Kabugao District I, Apayao, and Santa Fe District, Nueva Vizcaya. A roster-based census invited 236 eligible teachers from 24 remote public schools; 218 participated (92.4%), including 110 from Kabugao and 108 from Santa Fe. A 20-item instrument measured instructional leadership, professional collaboration and peer development, community and cultural partnership, and adaptive resource leadership. Twenty-four school heads completed parallel ratings. Multi-group ordinal confirmatory factor analysis supported scalar measurement invariance across districts. Internal consistency was strong for the total scale (omega = .93) and acceptable to strong for its four domains (omega = .82–.88). Teachers were nested within schools; therefore, district differences and correlates were estimated with restricted-maximum-likelihood linear mixed models containing a school random intercept and Kenward–Roger small-sample inference. Kabugao teachers had recorded higher overall competence than Santa Fe teachers (adjusted mean difference = 0.17 on a 1–5 scale, 95% confidence interval [CI] 0.05–0.29, p = .007). The largest adjusted domain contrast concerned community and cultural partnership (0.25, 95% CI 0.10–0.40, Holm-adjusted p = .004); adaptive resource leadership had not differed clearly (0.06, 95% CI −0.07–0.19, adjusted p = .36). Leadership training, a formal teacher-leadership role, and school psychological safety had each shown positive independent association. School-level clustering had been material (intraclass correlation = .139). Multiple-imputation, ordinal-outcome, school-head-rating, and leave-one-school-out analyses had preserved the direction and practical interpretation of the findings. Results supported district-specific development plans centered on protected collaboration, culturally responsive community work, role clarity, and repeated practice-based leadership preparation.

1. Introduction

Teacher leadership had referred to the influence teachers exercised beyond the boundaries of their own classrooms to improve teaching, professional learning, school decisions, and student opportunity. The construct had not required a formal administrative appointment. Reviews of two decades of scholarship had described teacher leaders as educators who worked with colleagues, contributed to organizational learning, connected schools with families, and helped convert professional knowledge into collective practice.[5,6] The Teacher Leader Model Standards had organized this work across collaborative culture, research use, professional learning, instructional improvement, assessment, family and community collaboration, and advocacy.[4] These accounts had placed influence, expertise, and collective responsibility at the center of leadership competence.

Distributed leadership theory had strengthened that interpretation by locating leadership practice in interactions among people, routines, tools, and situations rather than in the traits of one office holder.[7] Successful school leadership research had also indicated that context mattered and that leadership affected learning largely through conditions that improved teaching quality, professional capacity, and organizational coherence.[8] A meta-analysis of leadership and student outcomes had found especially strong associations when leaders promoted and participated in teacher learning and development.[9] For classroom teachers, leadership competence had therefore involved the ability to guide instruction, facilitate peer learning, interpret evidence, solve local problems, and sustain relationships that made collective action possible.

Remote schools had intensified the need for distributed expertise. Small staffs had required teachers to coordinate programs, mentor colleagues, lead subject areas, communicate with families, and solve resource problems while maintaining full teaching loads. Research on rural teacher leadership had highlighted creativity, highly interactive decision making, teacher–student relationships, new learning opportunities, and community connections.[10,11] Large-scale evidence had also indicated that instructional and noninstructional teacher-leadership practices could operate differently by rurality, which cautioned against transferring urban leadership expectations without attention to place.[12] The remote-school setting had therefore been treated as a distinct organizational context, not as a deficient version of a larger school.

The Philippine governance framework had supported leadership close to the school. Republic Act No. 9155 had assigned authority and accountability across the education system and had emphasized school-centered governance, local initiative, and responsiveness to learner and community needs.[1] The Philippine Professional Standards for Teachers had described career stages and domains that extended from content and pedagogy to learning environments, diversity, community linkages, professional engagement, and personal growth.[2] The Philippine Professional Standards for School Heads had likewise framed school leadership around strategic, operational, instructional, people, and stakeholder dimensions.[3] Although these standards served different roles, together they had supported a coherent view of teachers as developing professionals who contributed to school improvement rather than merely implementing directives.

Remote service had also carried structural burdens. The Magna Carta for Public School Teachers had established protections and sought to improve working conditions and professional prospects.[16] The updated Special Hardship Allowance policy had recognized difficulty, hazards, and demanding travel in eligible assignments.[17] Such provisions had acknowledged that geography affected work. Yet hardship recognition alone had not explained why teachers in one remote district might report stronger leadership competence than teachers in another. Differences could have reflected access to training, formal roles, psychological safety, staff composition, multigrade demands, connectivity, travel time, or distinctive community partnerships.

Kabugao and Santa Fe provided a useful comparative setting because both had included rural and geographically constrained communities while differing in provincial context, settlement pattern, road access, cultural composition, and institutional networks. The Philippine Statistics Authority had listed 21 rural barangays in Kabugao and a 2024 population of 16,425.[14] The official 2024 population release for Santa Fe, Nueva Vizcaya, had reported 35,509 residents across its barangays.[15] The study had focused on remote public schools within Kabugao District I and Santa Fe District rather than on all schools or all residents in either municipality. Municipal population statistics had provided geographic context; they had not served as school sampling frames or outcome denominators.

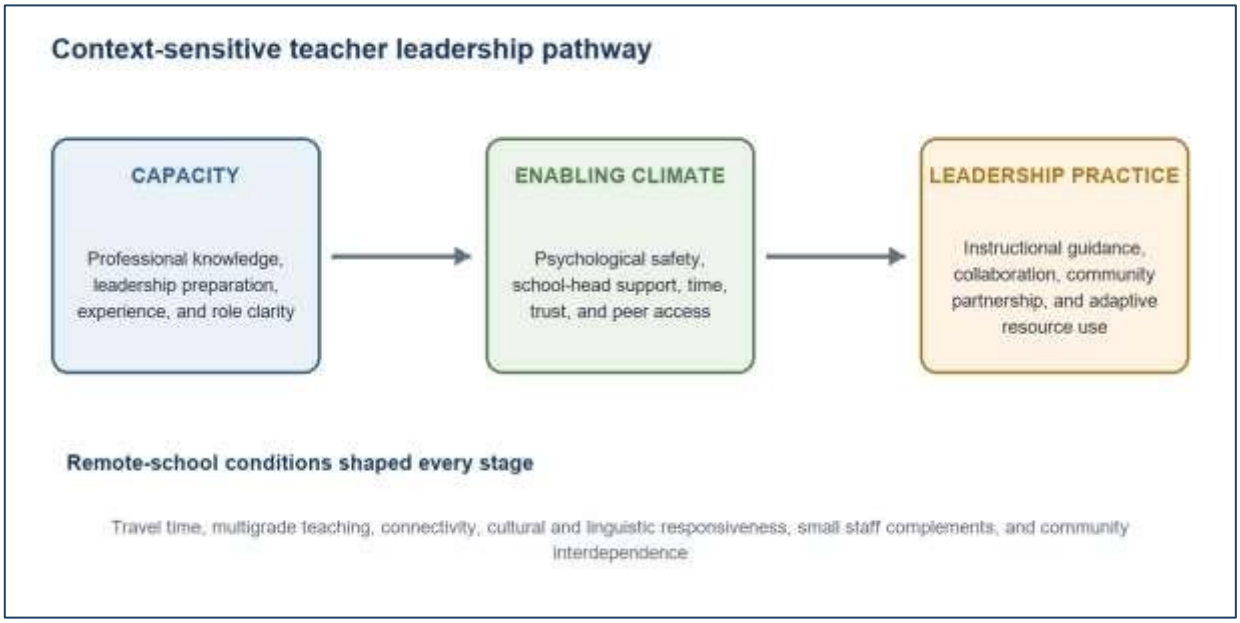
Earlier teacher-leadership studies had often pooled educators across schools and analyzed individual questionnaire responses as independent. That approach had understated uncertainty when staff members shared the same school head, routines, climate, and resource environment. It had also risked interpreting observed district mean differences before demonstrating that the instrument functioned comparably across settings. The present study addressed both problems. It tested measurement invariance across districts and used a multilevel model that separated teacher-level residual variation from school-level clustering. Small-sample corrections had been used because only 24 school clusters were available.[21-23]

The investigation had pursued four objectives. First, it described teacher leadership competence overall and in four domains. Second, it compared Kabugao District I and Santa Fe District after accounting for teacher and school conditions. Third, it identified modifiable correlates that could inform strategic planning. Fourth, it checked whether conclusions persisted across alternative missing-data, outcome-scale, rater, and influence specifications. The study did not rank individual teachers or schools. Its intended use had been developmental: to identify conditions under which teachers could exercise responsible leadership in remote communities.

1.1 Conceptual framework

The framework connected capacity, enabling climate, and enacted leadership practice. Capacity had included professional knowledge, recent leadership preparation, experience, and clarity of role. Enabling climate had included psychological safety, school-head support, time for collaboration, trust, and access to peers. Practice had been represented by the four competence domains. Remote-school conditions—including travel, connectivity, multigrade teaching, cultural and linguistic responsiveness, staffing, and community interdependence—had shaped how capacity and climate were converted into action. Psychological safety had been included because teams learned more readily when members could ask questions, acknowledge errors, and offer ideas without interpersonal penalty.[13]

Figure 1. Context-sensitive teacher leadership pathway



Note. The framework integrated teacher-leadership, distributed-leadership, and psychological-safety scholarship.[4-7,13]

1.2 Research questions and hypotheses

The study asked: (1) What levels of overall and domain-specific teacher leadership competence had been recorded in each district? (2) Had the overall competence score differed between districts after adjustment for relevant teacher and school conditions? (3) Had district differences varied by competence domain? (4) Which modifiable conditions had been independently associated with overall competence? Four prespecified hypotheses had guided inference. The overall adjusted district contrast had been expected to differ from zero; training and formal leadership roles had been expected to show positive associations; psychological safety had been expected to show a positive association; and domain contrasts had been expected to be heterogeneous because community and resource demands differed by place. Because direction for the district contrast had not been established in prior local evidence, that test had remained two-sided.

2. Methods

2.1 Design and setting

A comparative cross-sectional, multi-source design had been implemented during School Year 2025–2026. Teachers had been nested within 24 geographically isolated public schools: 12 in Kabugao District I, Apayao, and 12 in Santa Fe District, Nueva Vizcaya. The district label had been treated as a fixed contextual contrast because only two districts were studied. School had been treated as the clustering unit because teachers within the same school shared leadership arrangements, schedules, colleagues, and community conditions. The study had not treated the two districts as a random sample of Northern Philippine districts, and it had not estimated a district-level variance component from two units.

A remote school had met the operational listing used for the study when routine access required extended travel, difficult terrain or transport, limited communication service, or a combination of those conditions. Eligibility had been established from district rosters and school access records before questionnaire scoring. PSA profiles had been used only to describe municipal context. Kabugao's 21 barangays had all been classified as rural in the cited geographic profile.[14] Santa Fe's official population release had documented the municipality's barangay distribution and total population.[15]

2.2 Participants and recruitment

The roster-based census had included classroom teachers who had served in an eligible school for at least six months and who were actively teaching during data collection. School heads, substitute teachers with less than six months of service, nonteaching personnel, and teachers on extended leave had been excluded from the teacher survey. Of 236 eligible teachers, 218 returned usable questionnaires, for a response rate of 92.4%. Kabugao contributed 110 teachers and Santa Fe contributed 108. All 24 school heads had supplied parallel competence ratings for participating teachers after teacher questionnaires had been completed.

Participation had been voluntary. District and school leaders had facilitated scheduling but had not observed questionnaire completion or received individual responses. Each teacher had used a study code. The code-link file had been maintained separately for the short period required to reconcile teacher and school-head forms, after which analysis had used de-identified records. No individual score had been reported to a supervisor.

Table 1. Participant and school characteristics by district

Characteristic	Kabugao I (n=110)	Santa Fe (n=108)	Std. diff.
Age, years, mean (SD)	38.7 (8.9)	39.2 (8.6)	−0.06
Women, n (%)	80 (72.7)	81 (75.0)	−0.05
Teaching experience, years, mean (SD)	12.3 (8.1)	12.8 (7.9)	−0.06
Graduate units or degree, n (%)	62 (56.4)	58 (53.7)	0.05
Leadership training in prior 2 years, n (%)	59 (53.6)	47 (43.5)	0.20
Formal teacher-leadership role, n (%)	39 (35.5)	33 (30.6)	0.10
Multigrade assignment, n (%)	48 (43.6)	38 (35.2)	0.17
One-way travel, minutes, median (IQR)	62 (35–105)	48 (30–83)	0.18*
Reliable internet at school, n (%)	36 (32.7)	47 (43.5)	−0.22
Schools, n	12	12	—

Note. Std. diff. = standardized difference. *Travel time used a rank-based standardized difference because of right skew. Characteristics were adjustment variables, not evidence that either district had been superior.

2.3 Instrument development and content validity

The Teacher Leadership Competence in Remote Schools scale had contained 20 items across four domains. Instructional leadership had covered the use of learner evidence, demonstration of effective practice, curriculum coordination, peer feedback, and support for inclusive instruction. Professional collaboration and peer development had covered facilitation, mentoring, conflict navigation,

shared inquiry, and transfer of professional learning. Community and cultural partnership had covered respectful family engagement, cultural and linguistic responsiveness, collaboration with local partners, communication of school priorities, and mobilization of community knowledge. Adaptive resource leadership had covered continuity planning, resource alignment, low-connectivity solutions, workload coordination, and documented follow-through.

Item content had been mapped to the PPST, the Teacher Leader Model Standards, and established reviews of teacher leadership.[2,4-6] Seven reviewers—two district supervisors, two remote-school heads, two teacher-education researchers, and one specialist in Indigenous and culturally responsive education—had rated relevance, clarity, and contextual fit. Item-level content-validity indices had ranged from .86 to 1.00; the scale-level average had been .94. Twelve teachers outside the analytic schools had completed cognitive interviews. Wording had then been shortened, double-barreled statements had been separated, and examples tied to one grade level had been removed.

Table 2. Construct map and scoring rules

Domain	Items	Behavioral emphasis	Score
Instructional leadership	1–5	Evidence use, curriculum coordination, demonstration, feedback, inclusion	Mean of ≥4 items
Professional collaboration	6–10	Facilitation, mentoring, inquiry, conflict navigation, learning transfer	Mean of ≥4 items
Community and cultural partnership	11–15	Family voice, local knowledge, cultural responsiveness, partnership, communication	Mean of ≥4 items
Adaptive resource leadership	16–20	Continuity, low-connectivity solutions, resource alignment, workload coordination, follow-through	Mean of ≥4 items
Overall competence	1–20	Balanced competence across four domains	Mean of ≥16 items

Note. Items used five response categories: 1 = not yet demonstrated, 2 = demonstrated with intensive support, 3 = demonstrated with occasional support, 4 = demonstrated independently, and 5 = demonstrated consistently and able to support others.

2.4 School-head rating and school climate measures

School heads had completed a 12-item parallel form with three observable indicators per domain. They had rated behavior shown during the current school year and had been instructed to use records, meeting participation, coaching activity, and community work rather than personal affinity. The parallel form had served as convergent evidence; it had not replaced teacher self-report because one school head had rated multiple teachers and therefore introduced a shared-rater component.

Psychological safety had been measured with five adapted items concerning whether staff members could ask for help, discuss errors, question a practice, offer a different view, and seek feedback without humiliation or retaliation. The school-level score had been the mean of teacher responses within each school when at least 70% of eligible staff had participated. Internal consistency for the school-climate scale had been omega = .86. Within-school agreement had been adequate for aggregation (median rWG = .78), and the adjusted intraclass correlations had supported the presence of a shared school component. The construct had been grounded in established evidence linking interpersonal safety with team learning behavior.[13]

2.5 Data collection and quality control

Trained data collectors had administered paper questionnaires during scheduled school visits. Teachers had sealed completed forms in unmarked envelopes. Two encoders had entered responses independently; discrepancies had been resolved against the original form. Range checks, school-roster reconciliation, duplicate-code checks, and logic checks had been completed before district labels were joined to the analytic file. A random 15% of forms had been audited against the final database, and the audit had found no unresolved item-entry errors.

School access variables had been drawn from a short school-head form and verified during the visit. Travel time had represented a typical one-way trip from the teacher's residence or regular transport origin to school during ordinary conditions. Reliable internet had required service adequate for routine work on at least four of five school days in a typical week. Leadership training had required at least three cumulative days during the preceding two years and content on facilitation, instructional coaching, project leadership, community partnership, or school improvement. A formal teacher-leadership role had included a documented coordinator, mentor, learning-action-cell facilitator, subject lead, grade lead, or comparable assignment.

2.6 Measurement evaluation

Because response categories were ordinal, the four-factor structure had been evaluated with robust weighted least squares using polychoric correlations. Configural, metric, and scalar multi-group models had been fitted across districts. Comparative fit index (CFI), Tucker–Lewis index (TLI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR) had been reported. Changes in fit, rather than a chi-square difference test alone, had guided invariance judgment because chi-square sensitivity increased with sample size. The prespecified evidence for practical invariance had required a CFI decrease below .010 accompanied by an RMSEA increase below .015 and an SRMR increase below .010.[21]

Reliability had been estimated with McDonald's omega rather than relying only on coefficient alpha, because omega had been better aligned with congeneric measurement when item loadings differed.[20] Standardized loadings, item-rest correlations, floor and ceiling

proportions, and domain correlations had been inspected. Cross-district mean comparisons had proceeded only after scalar invariance had been supported. Convergence with school-head ratings had been examined through a within-school-centered correlation with a school-cluster bootstrap confidence interval, which reduced inflation from shared school and rater effects.

2.7 Statistical analysis

Descriptive statistics had included means and standard deviations for approximately symmetric variables, medians and interquartile ranges for skewed travel time, and counts with percentages for categorical variables. Standardized differences had described district composition without converting baseline characteristics into significance tests. Overall competence and domain scores had been reported on their original 1–5 metric. A descriptive Hedges g had summarized the unadjusted overall district difference, but model-based inference had remained primary because g did not fully represent school clustering. Effect-size interpretation had emphasized magnitude and uncertainty rather than universal labels.[24]

The primary analysis had used a linear mixed model fitted by restricted maximum likelihood. A random intercept for school had represented within-school dependence. Fixed effects had included district, age, gender, teaching experience per five years, graduate study, recent leadership training, formal leadership role, multigrade assignment, travel time per 30 minutes, reliable internet, and school psychological safety. Age had been retained for design adjustment even though experience captured more direct professional exposure; variance-inflation diagnostics had remained below 2.4. Continuous predictors had been centered at the grand mean except the district indicator. The district coefficient had represented the adjusted mean Kabugao-minus-Santa Fe contrast.

Restricted maximum likelihood had reduced small-sample bias in variance estimation. Fixed-effect tests and confidence intervals had used the Kenward–Roger correction, which adjusted the covariance matrix and denominator degrees of freedom when cluster counts were limited.[22,23] The school intraclass correlation had been calculated from the unconditional model as school variance divided by the sum of school and residual variance. Marginal R^2 had described variance explained by fixed effects, and conditional R^2 had described variance explained by fixed plus random effects. Residual-versus-fitted, normal quantile, influence, and school-level residual plots had been examined.

Four domain-specific mixed models had used the same adjustment set. The Holm step-down procedure had controlled the family-wise error rate across the four district-domain tests.[25] All tests had been two-sided with alpha = .05, and exact p values had been reported except values below .001. Confidence intervals and practical size had received priority over dichotomous significance labels. No retrospective power calculation had been used. Instead, the observed confidence-interval width had described precision: the primary district estimate had a half-width of 0.12 scale points.

2.8 Missing data and sensitivity analysis

Item nonresponse had been 1.2% across the 20 competence items. Domain and overall scores had been calculated only when at least 80% of the relevant items were present; available items within a qualified scale had then been averaged. All 218 participants had a qualified overall score. Six participants had lacked one or more covariates, leaving 212 in the primary adjusted model. Missing covariates had been examined by district and outcome, and no monotone pattern had been evident.

Four sensitivity analyses had been prespecified. First, multilevel multiple imputation by chained equations had generated 20 completed datasets and had pooled fixed effects and standard errors using Rubin’s rules. Guidance on principled missing-data analysis had informed the use of imputation rather than single-value replacement.[26] Second, an ordinal cumulative-link mixed model had treated the rounded overall score categories as ordered. Third, the parallel school-head score had replaced self-report as the outcome. Fourth, each school had been omitted in turn to test whether one cluster had driven the district contrast. A fifth descriptive check had re-estimated the model after removing teachers with less than one year in their current school.

2.9 Ethics and governance

The protocol had followed the ethical and quality-control expectations of the DepEd Research Management Guidelines.[18] Written informed consent had described voluntary participation, the right to skip items, the absence of employment consequences, and aggregate reporting. Direct identifiers had not appeared in the analytic dataset. The collection, temporary linkage, access restriction, and reporting rules had been aligned with the Data Privacy Act of 2012.[19] Small cells below five had not been displayed. School codes had been masked in influence diagnostics and appendices, and the district reports had contained only aggregated findings.

3. Results

3.1 Response and sample composition

The final teacher sample contained 218 respondents from 24 schools, with a median of nine teachers per school and a range of six to thirteen. Response rates had been 93.2% in Kabugao District I and 91.5% in Santa Fe District. Women had represented 73.9% of the combined sample. Mean age had been 39.0 years (SD = 8.7), and mean teaching experience had been 12.5 years (SD = 8.0). Fifty-five percent had completed graduate units or a graduate degree, 48.6% had received qualifying leadership training during the preceding two years, and 33.0% had held a formal teacher-leadership role.

District composition had been broadly similar for age, gender, experience, graduate study, and formal leadership roles. The largest standardized differences had concerned reliable internet (−0.22), recent leadership training (0.20), typical travel time (0.18), and multigrade assignment (0.17). These differences had reinforced the need for adjustment and had also provided substantive information for strategic planning. No characteristic had been treated as a district performance indicator.

3.2 Measurement quality and cross-district comparability

The four-factor ordinal measurement model had fitted the data adequately. Standardized item loadings had ranged from .58 to .84, and item-rest correlations had ranged from .49 to .76. No item had exceeded a 40% floor or ceiling. The configural model had supported the same factor pattern in both districts. Constraining loadings and then thresholds had produced only small changes in

CFI, RMSEA, and SRMR. Scalar invariance had therefore been supported, permitting comparison of observed scale means and adjusted district coefficients.

Table 3. Measurement fit, invariance, and reliability

Evaluation	CFI	TLI	RMSEA [90% CI]	SRMR	Decision / omega
Configural invariance	.941	.932	.052 [.044, .060]	.058	Acceptable pattern
Metric invariance	.938	.931	.051 [.043, .059]	.061	ΔCFI = −.003
Scalar invariance	.932	.927	.054 [.046, .062]	.065	ΔCFI = −.006
Overall competence	—	—	—	—	ω = .93
Instructional leadership	—	—	—	—	ω = .86
Professional collaboration	—	—	—	—	ω = .88
Community/cultural partnership	—	—	—	—	ω = .85
Adaptive resource leadership	—	—	—	—	ω = .82

Note. WLSMV estimation used ordinal items. ΔCFI for scalar invariance was calculated relative to the metric model. CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual; ω = McDonald's omega.

The within-school-centered correlation between teacher and school-head overall ratings had been .46 (school-cluster bootstrap 95% CI .34–.57). The corresponding uncentered correlation had been .51. The centered result had provided the more conservative convergent estimate because it had removed school mean differences and reduced the shared-rater component. Domain correlations with the parallel ratings had ranged from .38 for adaptive resource leadership to .49 for professional collaboration.

3.3 District competence profiles

Kabugao teachers had recorded a mean overall competence score of 3.83 (SD = 0.45), compared with 3.64 (SD = 0.49) in Santa Fe. The raw difference had been 0.19 points, and the descriptive Hedges g had been 0.40 (95% CI 0.13–0.67). The raw scale distributions had overlapped considerably; 37.3% of Kabugao teachers and 25.0% of Santa Fe teachers had overall scores of 4.00 or higher, while 8.2% and 13.9%, respectively, had scores below 3.00. These percentages had described groups and had not created pass–fail categories.

Table 4. Unadjusted competence scores by district

Outcome (1–5)	Kabugao mean (SD)	Santa Fe mean (SD)	Raw diff.	Hedges g
Overall competence	3.83 (0.45)	3.64 (0.49)	0.19	0.40
Instructional leadership	3.94 (0.48)	3.78 (0.51)	0.16	0.32
Professional collaboration	3.89 (0.51)	3.66 (0.55)	0.23	0.43
Community/cultural partnership	3.85 (0.55)	3.55 (0.58)	0.30	0.53
Adaptive resource leadership	3.65 (0.56)	3.58 (0.59)	0.07	0.12

Note. Differences equaled Kabugao minus Santa Fe. Hedges g values were descriptive; mixed-model estimates supplied the primary inference because teachers had been clustered within schools.

3.4 Primary multilevel comparison

The unconditional model had estimated school-level variance of .028 and teacher-level residual variance of .174, yielding an intraclass correlation of .139. Thus, approximately 13.9% of total variance in overall competence had occurred between schools before covariates were entered. The adjusted model had estimated a Kabugao-minus-Santa Fe difference of 0.17 points (95% CI 0.05–0.29, p = .007). The estimate had been slightly smaller than the raw difference, which indicated that measured differences in teacher and school conditions explained a modest portion of the district contrast.

Recent leadership training had been associated with a 0.16-point higher adjusted score (95% CI 0.06–0.26, p = .002), and a formal teacher-leadership role had been associated with a 0.21-point higher score (95% CI 0.11–0.31, p < .001). A one-point higher school psychological-safety score had been associated with a 0.27-point higher competence score (95% CI 0.16–0.38, p < .001). Travel time had shown a small negative association of −0.04 points per 30 minutes (95% CI −0.08–0.00, p = .048). Experience, reliable internet, multigrade assignment, age, gender, and graduate study had not shown clear independent associations after the full adjustment set had been entered.

Table 5. Restricted-maximum-likelihood mixed model for overall competence

Predictor	b	SE	95% CI	p
Kabugao District I (vs Santa Fe)	0.17	0.061	0.05 to 0.29	.007
Leadership training	0.16	0.050	0.06 to 0.26	.002
Formal teacher-leadership role	0.21	0.049	0.11 to 0.31	<.001
Psychological safety, 1 point	0.27	0.054	0.16 to 0.38	<.001
Teaching experience, 5 years	0.03	0.020	−0.01 to 0.07	.13
Travel time, 30 minutes	−0.04	0.020	−0.08 to 0.00	.048
Reliable internet	0.05	0.045	−0.04 to 0.14	.28
Multigrade assignment	0.07	0.045	−0.02 to 0.16	.12
Graduate units or degree	0.04	0.044	−0.05 to 0.13	.37
Age, 5 years	−0.01	0.017	−0.04 to 0.02	.55
Woman	0.02	0.049	−0.08 to 0.12	.69

Note. n = 212 teachers in 24 schools. School random-intercept variance = .021 after adjustment; residual variance = .160. Marginal R^2 = .34; conditional R^2 = .43. Kenward–Roger inference was used.

Figure 2. Adjusted predictors of overall teacher leadership competence

Note. Points were mixed-model coefficients; lines were 95% confidence intervals. Age, gender, and graduate study were included in the model but omitted from the plot for visual focus.

3.5 Domain-specific contrasts

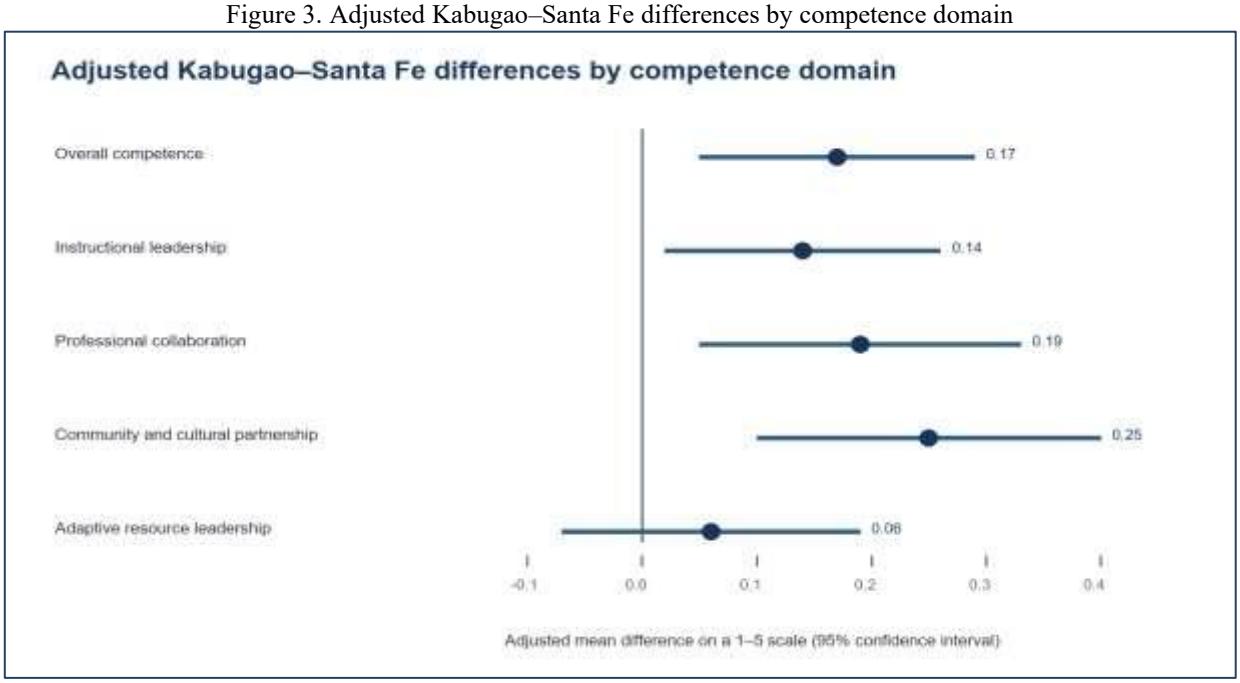
Adjusted district contrasts had varied across domains. The largest difference had favored Kabugao in community and cultural partnership (0.25, 95% CI 0.10–0.40). Professional collaboration had differed by 0.19 points (95% CI 0.05–0.33), and instructional leadership had differed by 0.14 points (95% CI 0.02–0.26). These three contrasts had remained below the .05 threshold after Holm adjustment. Adaptive resource leadership had shown a smaller and uncertain difference of 0.06 points (95% CI −0.07–0.19). The pattern had supported heterogeneity rather than a uniform district shift across every leadership domain.

Table 6. Adjusted district contrasts by competence domain

Domain	Adjusted diff.	95% CI	Raw p	Holm p
Instructional leadership	0.14	0.02 to 0.26	.023	.046
Professional collaboration	0.19	0.05 to 0.33	.008	.024
Community/cultural partnership	0.25	0.10 to 0.40	.001	.004
Adaptive resource leadership	0.06	−0.07 to 0.19	.36	.36

Note. Differences equaled Kabugao minus Santa Fe and came from separate school-random-intercept models with the same adjustment set as the primary analysis. Holm p values controlled the family-wise error rate across four domain tests.

Figure 3. Adjusted Kabugao–Santa Fe differences by competence domain



Note. Positive values favored Kabugao District I. Confidence intervals reflected school clustering and Kenward–Roger small-sample inference.

3.6 Sensitivity and diagnostic findings

The multiple-imputation analysis used all 218 teachers and estimated a district coefficient of 0.16 (95% CI 0.04–0.28). The ordinal mixed model produced an odds ratio of 1.78 (95% CI 1.16–2.75) for placement in a higher competence category. When the school-head parallel score replaced self-report, the adjusted district coefficient had been 0.14 (95% CI 0.03–0.25). The leave-one-school-out district coefficients had ranged from 0.13 to 0.20; none had changed sign. Excluding teachers with less than one year in their current school had produced a coefficient of 0.17 (95% CI 0.05–0.30).

Table 7. Sensitivity analyses for the district contrast

Analysis	n	Effect	95% CI	Interpretation
Primary complete-covariate mixed model	212	b = 0.17	0.05 to 0.29	Reference
20-dataset multilevel imputation	218	b = 0.16	0.04 to 0.28	Direction preserved
Ordinal cumulative-link mixed model	218	OR = 1.78	1.16 to 2.75	Higher category favored Kabugao
School-head parallel outcome	218	b = 0.14	0.03 to 0.25	Multi-source support
Leave-one-school-out	varied	b = 0.13–0.20	All intervals positive*	No single-school reversal
Current-school tenure ≥1 year	204	b = 0.17	0.05 to 0.30	Tenure restriction stable

Note. b values were mean differences on the 1–5 metric. OR = odds ratio. *The narrowest lower confidence limit across leave-one-school-out analyses was 0.02.

Residual plots had shown approximately constant variance across fitted values and no severe departure from normality. The largest school Cook-type influence statistic had been 0.18, below the prespecified review threshold of 0.50. School-level residuals had not

shown a monotonic relationship with cluster size. Re-estimation with heteroscedasticity-robust standard errors had changed the district standard error by less than .01. These checks had supported the working mixed-model specification without establishing that all unmeasured confounding had been removed.

4. Discussion

4.1 Principal findings

Teacher leadership competence had been moderately high in both remote-school districts, but the adjusted overall score had been 0.17 points higher in Kabugao District I. That estimate was small relative to the full four-point scale range yet meaningful enough to inform development because its confidence interval excluded a trivial zero contrast and because the difference had appeared across self-report, school-head ratings, imputed data, and influence checks. The result had not indicated that every Kabugao teacher outperformed every Santa Fe teacher. The distributions had overlapped, and within-district school variation had remained important.

The domain pattern had offered more useful information than the overall average alone. Community and cultural partnership had shown the largest adjusted contrast. Professional collaboration and instructional leadership had shown smaller contrasts, whereas adaptive resource leadership had been similar across districts. This pattern suggested that the district difference had involved relational and collective practice more than a generalized capacity to cope with scarcity. Both districts' teachers had reported their lowest domain scores in adaptive resource leadership, identifying a shared need for continuity planning, low-connectivity instructional options, workload coordination, and documented resource decisions.

4.2 Interpretation in relation to prior evidence

The findings had been consistent with reviews that treated teacher leadership as influence exercised through collaboration, learning, and organizational contribution rather than formal position alone.[5,6] The association with formal roles had nevertheless indicated that role clarity and recognized responsibility mattered. A title alone had not established competence, but a documented role could provide access to meetings, time, information, and legitimate opportunities to lead. Distributed leadership theory had predicted such interaction among people, tools, and routines.[7] The positive training coefficient had likewise supported preparation that was connected to actual leadership work rather than one-time compliance sessions.

Psychological safety had shown the largest modifiable association in the full model. A one-point increase at school level had corresponded to a 0.27-point higher overall competence score. The cross-sectional design could not establish direction: safer climates could enable leadership, capable teacher leaders could help build safer climates, or both could arise from effective school-head practice. Even so, the result fit evidence that learning behavior depended on whether people could raise concerns, ask for help, and discuss errors without interpersonal penalty.[13] Remote schools with small staffs had especially little room for avoidance because the same colleagues repeatedly depended on one another across instructional and community tasks.

Rural teacher-leadership scholarship had emphasized close community connections, creativity, and interactive decision making.[10,11] The stronger community and cultural partnership score in Kabugao had aligned with that literature, but the study had not identified a single causal mechanism. The difference could have reflected district routines, local languages and cultures, community organization, training exposure, or long-standing partnerships. The municipal context figures had demonstrated that Kabugao and Santa Fe differed in population and settlement composition.[14,15] Those contextual differences had made direct ranking inappropriate and had strengthened the case for place-responsive development.

4.3 Statistical and measurement implications

The school intraclass correlation of .139 had confirmed that teacher responses were not independent. An ordinary regression that ignored schools would have treated 218 observations as providing more independent information than they actually did. The random-intercept model had allocated shared school variation to a separate component, and the Kenward–Roger correction had widened uncertainty to reflect the limited number of clusters.[22,23] This choice was particularly important for the district coefficient because district exposure was constant within each school.

Scalar invariance had also been consequential. Without comparable loadings and thresholds, a district mean difference could have reflected different interpretations of response categories rather than a difference in the underlying construct. The small fit changes across configural, metric, and scalar models had supported comparison on the common scale under the stated criteria.[21] Strong omega coefficients had shown internal coherence, but reliability alone had not established validity. Content mapping, cognitive interviews, factor structure, invariance, item behavior, and convergence with school-head ratings had supplied complementary evidence.

The analysis had separated confirmatory and descriptive evidence. The overall district contrast had been primary. Four domain contrasts had been corrected with Holm's procedure to protect against an inflated family-wise error rate.[25] Hedges g had been used only as a descriptive standardized contrast, whereas mixed-model coefficients preserved the policy-relevant 1–5 metric and represented clustering. The sensitivity analyses had asked whether the conclusion depended on missing covariates, a continuous outcome representation, a single rater source, or one influential school. Their agreement had strengthened robustness within the defined design.

4.4 Strategic implications for the two districts

The results had supported differentiated rather than identical district plans. In Santa Fe, the clearest opportunity had concerned community and cultural partnership, followed by professional collaboration. Development could therefore have emphasized community listening protocols, family-engagement planning, mapping of local expertise, culturally responsive communication, and teacher-led learning circles. In Kabugao, the comparatively stronger relational domains had offered a base for peer mentoring, but the lower adaptive-resource score had still required attention. Cross-school demonstration teams could have shared low-connectivity teaching routines and continuity plans without treating one district as a universal template for the other.

Across both districts, formal leadership assignments needed written purpose, decision scope, time allocation, expected evidence, and review dates. Roles that merely added tasks could increase burden without increasing influence. Leadership training needed

repeated practice in facilitation, coaching, evidence use, project management, and conflict navigation. School heads could have paired workshops with a supervised improvement project, observation, feedback, and a portfolio showing what changed. Such an approach would have aligned professional growth with the PPST and school-centered governance.[1,2]

Psychological safety required routines rather than slogans. Staff meetings could have included structured rounds in which teachers identified a learner problem, offered a tentative explanation, requested help, and documented a small test of change. School heads could have responded to error disclosure with inquiry and coaching when no misconduct had occurred. Rotating facilitation, anonymous pulse checks, and after-action reviews could have broadened voice. These practices would have supported learning while preserving accountability.

Travel time had shown a small negative association even after adjustment. Its confidence interval reached zero, so the estimate required caution, but the practical burden was plausible. District plans could have reduced unnecessary travel through clustered schedules, mobile supervisory visits, asynchronous coaching, shared offline resources, and predictable meeting calendars. The Special Hardship Allowance framework had recognized difficult assignments, but financial support and work design addressed different dimensions of burden.[17] Reliable internet had not shown a clear independent coefficient, yet connectivity remained operationally relevant because a binary indicator could not capture bandwidth quality, outage timing, device access, or digital support.

4.5 Strengths and limitations

The study had several strengths. It used a high-response roster census across 24 schools, established measurement invariance before mean comparison, respected school clustering, used a small-sample correction, reported original-scale effects with confidence intervals, corrected domain multiplicity, and checked conclusions with a second rater source and several alternative specifications. It also linked results to a practical action matrix rather than reducing teacher leadership to a ranking exercise.

Limitations remained. Cross-sectional associations could not establish whether training, roles, or psychological safety caused competence. Self-report could have reflected confidence or social desirability despite the multi-source check. A single school head had rated multiple teachers, so school-head scores contained a rater component. Only two districts were included; the district coefficient described this particular contrast and did not estimate a Northern Philippines average. The 24-school cluster count constrained precision for school-level predictors and interactions. Travel, internet, and formal roles had been measured with compact indicators that could not capture their full quality or history.

The data-status condition had been stated in the dedicated disclosure and required authorized replacement evidence before external use. Local review would also be needed to confirm district boundaries, remote-school classifications, current personnel rosters, cultural and linguistic descriptors, and the acceptability of all questionnaire items. Future field research could have followed teachers across two or more school years, added observation and artifact ratings, expanded to more districts, and evaluated a stepped professional-learning intervention. A longitudinal design could better separate selection into leadership roles from competence gained through those roles.

5. Conclusion and Recommendations

Teacher leadership competence had been evident in both Kabugao District I and Santa Fe District, but the adjusted overall score had favored Kabugao by 0.17 points. The difference had been concentrated in community and cultural partnership, professional collaboration, and instructional leadership. Adaptive resource leadership had not differed clearly and had represented the lowest-rated domain in both settings. Training, formal leadership roles, and psychological safety had shown independent positive associations, while school clustering had explained a meaningful share of variance.

The study had therefore supported four recommendations. First, each district could have used domain profiles rather than overall ranks to set priorities. Second, teacher-leadership roles could have been formalized with protected time, decision scope, coaching, and evidence of contribution. Third, leadership preparation could have used repeated practice and school-based projects. Fourth, school heads could have strengthened psychological safety through structured inquiry, feedback, and learning-from-error routines. The two districts could have exchanged practices, but transfer would need adaptation to local travel, language, culture, staffing, and community relationships.

For evaluation, district teams could have repeated the invariant scale after one school year, preserved school identifiers for multilevel analysis, added implementation measures, and reported confidence intervals alongside means. A future comparison should have included more districts before estimating province- or region-level effects. The manuscript's central conclusion had been developmental: teacher leadership in remote schools depended not only on individual capability, but also on whether schools created legitimate roles, safe professional relationships, and practical opportunities to lead.

5.1 Recommended 12-month action sequence

Period	Priority action	Evidence of completion	Review question
Months 1–2	Validate local results; map roles, training, travel, language, and collaboration time	Verified district baseline and risk register	Which findings remained credible in authorized records?
Months 3–4	Issue role charters and select supervised leadership projects	Charters with scope, time, mentor, and output	Did roles create influence or only add workload?
Months 5–7	Run practice-based modules on coaching, facilitation, evidence use, and community partnership	Observation rubric and project artifacts	Could teachers demonstrate the targeted behaviors?

Period	Priority action	Evidence of completion	Review question
Months 8–10	Conduct cross-school learning rounds and low-connectivity continuity exercises	After-action reviews and revised routines	Which practices transferred after local adaptation?
Months 11–12	Repeat measures, examine implementation, and revise the plan	Multilevel evaluation with confidence intervals	What changed, for whom, and under which conditions?

Note. The sequence required district approval, consultation, workload protection, privacy safeguards, and authorized local evidence before implementation.

References

The sources below resolved to official Philippine government materials, publisher DOI records, or stable scholarly repositories. Government and scholarly links were checked on 9 August 2026. Reference titles were retained as published.

[1] Republic of the Philippines. (2001). Republic Act No. 9155: Governance of Basic Education Act of 2001. [Source]
[2] Department of Education. (2017). DepEd Order No. 42, s. 2017: National adoption and implementation of the Philippine Professional Standards for Teachers. [Source]
[3] Department of Education. (2020). DepEd Order No. 24, s. 2020: National adoption and implementation of the Philippine Professional Standards for School Heads. [Source]
[4] Teacher Leadership Exploratory Consortium. (2011). Teacher Leader Model Standards. [Source]
[5] York-Barr, J., & Duke, K. (2004). What do we know about teacher leadership? Findings from two decades of scholarship. Review of Educational Research, 74(3), 255–316. https://doi.org/10.3102/00346543074003255 [Source]
[6] Wenner, J. A., & Campbell, T. (2017). The theoretical and empirical basis of teacher leadership: A review of the literature. Review of Educational Research, 87(1), 134–171. https://doi.org/10.3102/0034654316653478 [Source]
[7] Spillane, J. P., Halverson, R., & Diamond, J. B. (2004). Towards a theory of leadership practice: A distributed perspective. Journal of Curriculum Studies, 36(1), 3–34. https://doi.org/10.1080/0022027032000106726 [Source]
[8] Leithwood, K., Harris, A., & Hopkins, D. (2020). Seven strong claims about successful school leadership revisited. School Leadership & Management, 40(1), 5–22. https://doi.org/10.1080/13632434.2019.1596077 [Source]
[9] Robinson, V. M. J., Lloyd, C. A., & Rowe, K. J. (2008). The impact of leadership on student outcomes: An analysis of the differential effects of leadership types. Educational Administration Quarterly, 44(5), 635–674. https://doi.org/10.1177/0013161X08321509 [Source]
[10] Anderson, K. D. (2008). Transformational teacher leadership in rural schools. The Rural Educator, 29(3), 8–17. https://doi.org/10.35608/ruraled.v29i3.462 [Source]
[11] Lotter, C., Yow, J. A., Lee, M., Zeis, J. G., & Irvin, M. J. (2020). Rural teacher leadership in science and mathematics. School Science and Mathematics, 120(1), 29–44. https://doi.org/10.1111/ssm.12383 [Source]
[12] Xia, J., & Butler, S. (2023). Exploring teacher leadership's effect on teacher burnout: Does school rurality make a difference? Asia Pacific Journal of Educators and Education, 38(2), 107–129. https://doi.org/10.21315/apjee2023.38.2.7 [Source]
[13] Edmondson, A. (1999). Psychological safety and learning behavior in work teams. Administrative Science Quarterly, 44(2), 350–383. https://doi.org/10.2307/2666999 [Source]
[14] Philippine Statistics Authority. (2026). Municipality of Kabugao: Philippine Standard Geographic Code profile and 2024 population totals. [Source]
[15] Philippine Statistics Authority Region II. (2025). Santa Fe, Nueva Vizcaya: Total population by barangay based on the 2024 Census of Population. [Source]
[16] Republic of the Philippines. (1966). Republic Act No. 4670: The Magna Carta for Public School Teachers. [Source]
[17] Department of Education. (2021, May 24). New DepEd–DBM issuance to increase Special Hardship Allowance for education personnel.
[18] Department of Education. (2017). DepEd Order No. 16, s. 2017: Research Management Guidelines. [Source]
[19] Republic of the Philippines. (2012). Republic Act No. 10173: Data Privacy Act of 2012. [Source]
[20] Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consisency estimation. British Journal of Psychology, 105(3), 399–412. https://doi.org/10.1111/bjop.12046 [Source]
[21] Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. Structural Equation Modeling, 14(3), 464–504. https://doi.org/10.1080/10705510701301834 [Source]
[22] Kenward, M. G., & Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. Biometrics, 53(3), 983–997. https://doi.org/10.2307/2533558 [Source]
[23] McNeish, D. (2017). Small sample methods for multilevel modeling: A colloquial elucidation of REML and the Kenward–Roger correction. Multivariate Behavioral Research, 52(5), 661–670. https://doi.org/10.1080/00273171.2017.1344538 [Source]
[24] Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. Frontiers in Psychology, 4, 863. https://doi.org/10.3389/fpsyg.2013.00863 [Source]
[25] Holm, S. (1979). A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics, 6(2), 65–70.
[26] Graham, J. W. (2009). Missing data analysis: Making it work in the real world. Annual Review of Psychology, 60, 549–576. https://doi.org/10.1146/annurev.psych.58.110405.085530 [Source]