

The School Heads' Supervisory Strategies in Bridging the Master Teachers' Classroom Observation Standards in the Public Elementary Schools in Buenavista Districts I And II, Division of Quezon

Amie T. Olea¹

¹School of Graduate Studies, University of Northeastern Philippines

ARTICLE INFORMATION	ABSTRACT
Article history: Published: October 2026	The study examined the extent of school heads' supervisory strategies and their reported role in bridging master teachers' classroom observation standards in public elementary schools of Buenavista Districts I and II, Division of Quezon, during school year 2025–2026. A descriptive–evaluative–inferential design used a researcher-made, two-part questionnaire and weighted means. The detailed respondent tabulation identifies 33 participants—22 school heads and 11 master teachers—across 22 schools, while a summary states 44. The overall rating for general supervisory strategies was 4.18 (Much Employed), with classroom management lowest at 3.51. Strategies for bridging observation standards averaged 4.41 (Much Evident), led by technical assistance at 4.58. Reported Mann–Whitney comparisons found a difference for instructional planning ($z = -2.02$, $p = 0.0434$) and for instructional coaching, technical assistance, and professional development support (reported $p = 0.0264$, 0.0392 , and 0.0047). Several rank totals and probability values do not reconcile with the reported tests. These descriptive ratings suggest priorities for supervision and support, but cannot show that supervision improved teaching or observation performance. The participant count, instrument mapping, and inferential calculations require confirmation before submission.
Keywords: Instructional Supervision Classroom Observation School Heads Master Teachers Buenavista Mann–Whitney U	

1. Introduction

Classroom observation can connect instructional expectations with the work teachers undertake during lesson planning, delivery, assessment, and reflection. Its practical value depends on what school leaders do before and after an observation: clarify standards, provide specific feedback, offer technical guidance, and create opportunities for professional learning. In the Philippine public school system, the Philippine Professional Standards for Teachers (PPST) provide a policy frame for professional practice and development (Department of Education [DepEd], 2017). The study focuses on the local supervisory strategies through which school heads support master teachers' observation standards in two districts of Quezon.

The phrase “bridging classroom observation standards” is used in the study for respondents' ratings of instructional coaching, technical assistance, professional development support, collaborative supervision, and resource and policy support. It does not denote a measured gap between a documented standard and a separately assessed classroom practice. This distinction matters: school leaders may report frequent guidance, while the effects of that guidance on instruction, observation scores, or pupils' learning remain unmeasured. School based professional learning policies such as the Learning Action Cell (LAC) offer one institutional route for sustained teacher collaboration (DepEd, 2016), but their presence does not itself establish implementation quality.

A wider literature treats collaboration and professional learning as intentionally organized work rather than an isolated meeting. Hargreaves and O'Connor (2018) describe purposeful designs for collaborative professionalism, while the OECD (2019) documents teacher and school leader perspectives on continuing development across systems. These works provide a framework for interpreting patterns in local ratings; they do not substitute for a direct comparison group or independent performance measures in Buenavista. The study's specific contribution is a district-level profile of perceived supervisory practices and a set of reported rank comparisons between Buenavista Districts I and II.

The research questions asked: (1) to what extent were school heads' strategies employed in instructional planning, classroom management, instructional delivery, assessment practices, and professional responsibilities; (2) whether rank orders for these dimensions differed between districts; (3) to what extent were strategies for bridging observation standards evident in five support domains; (4) whether those rank orders differed between districts; and (5) what policy recommendations followed. The manuscript preserves the reported means, labels, and test statistics. Material conflicts in participation, item mapping, and inferential calculations are identified separately for author review.

2. Methods

2.1 Design and setting

The study employed a descriptive–evaluative–inferential design in the public elementary schools of Buenavista Districts I and II, Division of Quezon, for school year 2025–2026. Its descriptive component summarized respondents' ratings of supervisory practices. Its evaluative component interpreted those practices in relation to classroom observation standards. Its inferential component compared ranks between the two districts. No intervention, observation-based outcome assessment, longitudinal follow-up, or causal model was

reported. The study timetable extended from June 2025 to May 2026; the exact dates of questionnaire administration were not specified.

2.2 Respondents and recruitment

The methods describe total enumeration of the identified school heads and master teachers. The detailed respondent table lists 22 school heads and 11 master teachers in 22 schools, divided into 18 participants in District I and 15 in District II (Table 1). This is the only school-level participant accounting reported. The preliminary abstract and final summary state 44 participants, and the final summary also describes purposive sampling. Neither the additional eleven participants nor an alternative recruitment frame is documented in the respondent table. The analysis therefore identifies counts as printed in the respondent tabulation and avoids treating either number as a reconciled analytic sample.

Table 1. Reported respondent distribution by district

District	School heads	Master teachers	Total respondents	Percentage
Buenavista District I	11	7	18	55
Buenavista District II	11	4	15	45
GRAND TOTAL	22	11	33	100

The tabulation locates slightly more respondents in District I (18) than District II (15). Each district lists eleven school heads; the district difference is in the number of master teachers (seven versus four). These composition differences matter because a comparison of district responses could conflate location with role unless raw data are separated or adjusted. The study does not report role-stratified results, eligibility criteria beyond school head or master teacher status, response rate, or demographic characteristics. Consequently, the group comparison cannot be interpreted as a clean contrast of otherwise comparable populations.

2.3 Instrument and procedure

The primary instrument was a researcher-made questionnaire with two parts and a stated total of 120 items. Part I elicited ratings of instructional planning, classroom management, instructional delivery, assessment practices, and professional responsibilities; Part II concerned instructional coaching, technical assistance, professional development support, collaborative supervision, and resource and policy support. The table of specifications states sixty items for each part. Its category item counts and printed item ranges do not consistently agree—for example, instructional planning is said to have thirteen items but is assigned items 1–12; the subsequent narrative likewise reports thirteen. The final questionnaire and coding key should establish the actual number analyzed in each domain. Both parts used five-point response scales. Part I classified 4.50–5.00 as Very Much Employed, 3.50–4.49 as Much Employed, 2.50–3.49 as Employed, 1.50–2.49 as Fairly Employed, and 1.00–1.49 as Not at All. Part II used the corresponding labels Very Much Evident, Much Evident, Evident, Fairly Evident, and Not at All. Content validation by experts in educational management, instructional supervision, and research was described, followed by revision of unclear or overlapping items. A pilot test among respondents outside the actual study was reported, but its size and a numerical Cronbach's alpha were not given. The text states that item correlations were at least 0.50 and computed *t* values exceeded a tabular value without providing item-level data or the threshold. Reliability cannot be independently evaluated from those statements.

After proposal review and approval, the instruments were administered to respondents and the results tallied and analyzed. The study lists permissions and stages of manuscript review but provides no details on mode of survey delivery, consent, anonymity, missing responses, or data management. These absent elements should be documented from study records for journal submission; none is inferred here.

2.4 Statistical analysis

Weighted means and ranks summarized the reported five-point ratings. The Mann–Whitney *U* test was designated for comparing district rank orders at $\alpha = .05$, with *z* values and associated probabilities reported in the findings. The chapter describes the two independent groups as districts, yet the appendices rank indicator-level district means; the exact units entered into the test and the treatment of ties are not clearly specified. Tables 3 and 5 reproduce the published *z* and *p* values as reported, rather than recalculating them. Some values appear incompatible with conventional two-sided normal probabilities, and selected rank totals differ across the findings narrative and computational appendices. These concerns limit inferential interpretation pending access to the analytic dataset and calculations.

4. Results

4.1 Extent of supervisory strategies

Table 2. Summary of the extent of school heads' supervisory strategies employed

Indicator	District I (Wx, Int)	District II (Wx, Int)	Average (Wx, Int)	R
Instructional Planning	4.43 ME	4.27 ME	4.35 ME	2.5
Classroom Management	3.42 E	3.60 ME	3.51 ME	5
Instructional Delivery	4.34 ME	4.36 ME	4.35 ME	2.5
Assessment Practices	4.39 ME	4.31 ME	4.35 ME	2.5
Professional Responsibilities	4.36 ME	4.34 ME	4.35 ME	2.5
AVERAGE	4.19 ME	4.18 ME	4.18 ME	

The overall rating of 4.18 falls in Much Employed. Four domains share the printed average of 4.35, while classroom management is lower at 3.51. District I's classroom-management mean of 3.42 is interpreted as Employed, whereas District II's 3.60 is Much Employed. The equality of four domain means after rounding should not be taken as evidence that their item patterns were identical. The comparatively lower classroom-management rating points to a potential priority for feedback on learner behavior, routines, and participation; it does not establish the prevalence or severity of classroom disruptions.

Within instructional planning, checking alignment among objectives, activities, and assessment was the highest reported indicator (4.60), whereas feedback on planning after observation was 4.10. That pattern suggests a distinction between checking a plan and following up with actionable guidance. In classroom management, encouraging positive learner behavior was rated 3.74, while support for handling behavior concerns was 3.21. The lower follow-through ratings are compatible with a need for targeted assistance, but the survey did not observe actual supervisory conferences or classroom behavior. The means describe perceptions of practice, not implementation fidelity.

4.2 District comparison of supervisory strategies

Table 3. Reported Mann–Whitney results for rank orders of supervisory strategies

Indicator	z	Reported p	Decision at .05
Instructional Planning	−2.02	0.0434	Significant
Classroom Management	1.89	0.0588	Not significant
Instructional Delivery	−0.40	0.6892	Not significant
Assessment Practices	−1.08	0.2801	Not significant
Professional Responsibilities	−1.70	0.0891	Not significant

Instructional planning is the only general-supervision domain reported as significant at .05 ($z = -2.02$, $p = 0.0434$). The other four p values exceed .05. A nonsignificant result does not demonstrate equivalent practice between districts; it indicates that the reported test did not reject its null hypothesis for those dimensions. The difference in planning also should not be read as a causal district effect, because role mix, school conditions, and item-level responses were not controlled. The sign of z alone cannot establish which district is higher without a confirmed rank definition and calculation.

The numerical evidence for this table needs a statistical audit. The reported rank-sum pairs differ between the findings narrative and the detailed calculations. Only the expressly reported z , p , and decisions are presented here. Independent claims about effect size, direction, or exact U statistic would exceed the available record. A corrected analysis should specify whether the observations were the 33 respondents, district-aggregated indicator means, or another set, and show the rank sums, sample sizes, tie correction, and test tail.

4.3 Strategies for bridging observation standards

Table 4. Summary of strategies in bridging master teachers' classroom observation standards

Indicator	District I (Wx, Int)	District II (Wx, Int)	Average (Wx, Int)	R
Instructional Coaching	4.60 VME	4.38 ME	4.49 ME	2
Technical Assistance	4.54 VME	4.62 VME	4.58 VME	1
Professional Development Support	4.38 ME	4.12 ME	4.25 ME	5
Collaborative Supervision	4.27 ME	4.28 ME	4.28 ME	4
Resource and Policy Support	4.48 ME	4.43 ME	4.46 ME	3
AVERAGE	4.45 ME	4.37 ME	4.41 ME	

The bridging-strategy summary averaged 4.41 (Much Evident). Technical assistance led at 4.58 (Very Much Evident), followed by instructional coaching at 4.49, resource and policy support at 4.46, collaborative supervision at 4.28, and professional development support at 4.25. The relative ranking suggests that immediate guidance on observation concerns was rated more strongly than ongoing professional-development support. Technical assistance involved clarifying standards, materials, and teaching concerns; however, the study does not show which assistance was used, whether it matched master teachers' needs, or whether subsequent observation performance changed.

The highest stated technical-assistance item was provision of technical guidance during classroom supervision (4.70). Professional-development follow-up was rated 4.54, but guidance on improving instructional competence was 4.07. Collaborative discussion of observation practices reached 4.52, whereas follow-up collaborative supervision was 4.09. These within-domain contrasts point toward continuity after the initial supervisory contact. They are descriptive ratings, so differences in scores cannot be interpreted as estimated effects of a specific practice. A log of feedback, agreed actions, and follow-up could test whether guidance is actually sustained.

4.4 District comparison of bridging strategies

Table 5. Reported Mann–Whitney results for rank orders of bridging strategies

Indicator	z	Reported p	Decision at .05
Instructional Coaching	−1.94	0.0264	Significant
Technical Assistance	−1.76	0.0392	Significant
Professional Development Support	−2.60	0.0047	Significant
Collaborative Supervision	−0.42	0.2514	Not significant
Resource and Policy Support	−0.20	0.4207	Not significant

Three bridging domains were reported as statistically different between districts: instructional coaching ($p = 0.0264$), technical assistance ($p = 0.0392$), and professional development support ($p = 0.0047$). Collaborative supervision and resource and policy support were not reported significant. The domain means in Table 4 place District I above District II for coaching (4.60 versus 4.38) and development support (4.38 versus 4.12), whereas District II is higher for technical assistance (4.62 versus 4.54). These mean contrasts describe the summaries; they do not resolve the rank-test direction or effect size.

The probabilities in Table 5 warrant particular caution. Those paired with $z = -1.94$, -1.76 , -2.60 , and -0.20 approximately correspond to one-sided normal tail areas. The reported $p = 0.2514$ for $z = -0.42$ corresponds to neither its conventional one-sided nor two-sided area. A

test direction was not specified. Thus the reported labels are retained as findings, but substantive claims of district differences should be treated as provisional until the raw scores or complete ranked dataset are reanalyzed with a declared test tail and transparent tie handling.

5. Discussion

The first research question produced a generally favorable profile of supervisory practice, yet classroom management remained the lowest domain. This is analytically useful because classroom management depends on decisions during instruction, not simply on a document checked before observation. The study's planning item pattern suggests strong attention to alignment of objectives and assessment, while lower ratings for post-observation feedback indicate that evidence from observation may not consistently return to planning. A practical supervisory cycle would make the connection explicit: identify a classroom issue, agree on a targeted action, observe its use, and revisit evidence. The present ratings suggest where to examine that cycle; they cannot establish that it is operating or improving instruction.

The PPST establishes common professional expectations, including domains relevant to teaching and professional practice (DepEd, 2017). The study's five general-supervision domains are locally operationalized ratings rather than direct measurements of every PPST strand. Alignment with the policy requires mapping the actual survey items and observation indicators to specific standards. That mapping is especially important because the instrument specification disagrees internally about item counts and ranges. Without it, a mean of 4.35 for assessment practices cannot be interpreted as attainment of a particular professional standard.

The second research question reported a district difference only in instructional planning. It is tempting to frame this as a distinct district supervisory model, but the study supplies no evidence about school size, workload, personnel, or implementation history that would explain the pattern. Because the rank test's unit is unclear, the result is best viewed as a flagged contrast for further examination. A future study could compare individual respondents by role and district, inspect variation within schools, and use the same documented item set. Qualitative accounts of planning conferences could help explain why a check on alignment is rated highly while feedback after observation is lower.

The third research question distinguished technical assistance (4.58) from professional development support (4.25). This is a meaningful conceptual difference. Immediate help with a classroom observation requirement may be easier to deliver and recognize than sustained opportunities to learn, practice, receive feedback, and refine instruction. OECD (2019) discusses professional learning as part of teachers' and school leaders' continuing development. In the Philippine context, the LAC policy provides a formal mechanism for school-based continuing professional development (DepEd, 2016). The study did not test LAC participation or quality, so these sources frame a possible response rather than verify a program effect in Buenavista.

Collaborative supervision was rated Much Evident at 4.28, but its lower follow-up item (4.09) suggests that participation in a discussion and sustained collective work should be distinguished. Hargreaves and O'Connor (2018) emphasize intentional, purposeful collaboration. In the study's setting, school heads and master teachers could specify a focus before an observation, jointly review evidence afterward, and return to the focus in later meetings. Whether that sequence produces better instruction requires observation records, teacher accounts, and pupil learning evidence. A high survey rating for discussion alone does not answer that question. The fourth research question yielded reported significant comparisons in three bridging domains, including the lowest mean domain, professional development support. The probability-to-z discrepancies and inconsistent rank sums constrain those inferences. In a Mann-Whitney analysis, specifying the observational unit is essential. If the inputs are respondent ratings, scores may be clustered within schools and affected by role composition; if inputs are domain indicator means, the test concerns differences across sets of questionnaire items, not independent people. Either choice needs a defensible explanation and an explicit method for ties. The current record does not support deciding which interpretation was intended.

Several further limitations bear directly on the conclusions. All domains were measured in one researcher-made instrument and may share response style or social desirability effects. The study did not report a numeric alpha or subscale reliability, despite describing pilot testing. No direct observation of school heads' feedback, master teachers' classroom practice, observation ratings, or student outcomes was provided. The group sizes and recruitment approach conflict across sections. The label "bridging standards" could imply an effect or closure of a gap that the design never directly measured. These features restrict generalization beyond the participating districts and preclude causal statements.

For policy, the most defensible steps are modest and measurable. School leaders could document specific post-observation feedback, follow-up on classroom-management concerns, and maintain a transparent record of technical assistance and professional-learning activities. They could align these records with identified PPST indicators and examine master teachers' own priorities in LAC sessions. Before comparing districts statistically, the author should reconcile the participant roster, finalized questionnaire, scoring rules, and inferential output. A subsequent evaluation can pair reported support with independent observations or performance indicators over time. The present data justify scrutiny of processes and targeted support, not a claim that the proposed actions will necessarily improve classroom observation scores.

The two domains with nonsignificant reported bridging comparisons—collaborative supervision and resource and policy support—remain pertinent to practice. Nonsignificance is not proof that the districts implement them equally. It may reflect the tested ranks, available sample, ties, or uncertainty in the reported probability values. District-level planning should therefore use local needs assessments rather than treating the inferential labels as an allocation rule. The district means for collaboration differ by only 0.01 after rounding, but the actual response distributions were not shown. A mean alone can conceal respondents who receive little support, especially where schools have different numbers of master teachers.

The comparatively high rating for resource and policy support (4.46) also calls for careful interpretation. Access to instructional materials and clear communication of observation procedures are tangible supports, yet the study does not provide inventories of materials or policy-compliance records. The top item on access to materials (4.71) and the lower item on administrative support during observation (4.31) could reflect different practical demands. Documenting resource requests, fulfillment, and timeliness

would test whether favorable perceptions correspond to reliable support. Similarly, asking teachers which policy guidance remained unclear would turn a broad satisfaction score into actionable evidence without assuming that a policy was implemented uniformly. The descriptive tables use weighted means for ordinal response categories. Their two-decimal precision is useful for describing the tabulated ratings but should not be interpreted as a precise interval measure of supervisory quality. The category threshold at 4.50 creates visible label changes for small rounded differences: instructional coaching at 4.49 is Much Evident, while technical assistance at 4.58 is Very Much Evident. A future report should provide the distribution of responses and uncertainty around domain estimates. That would permit readers to see whether most respondents agree or whether an average masks polarized experiences. The missing scale and respondent details currently prevent such analysis.

A more rigorous follow-up could specify a small number of observable supervisory actions, document them at baseline, and assess whether they occur after training or policy changes. Observation rubrics could record feedback specificity, alignment of lesson objectives and assessment, management strategies, and the completion of agreed follow-up. Master teachers could report whether guidance was timely and usable, while independently scored classroom observations would provide a separate outcome. Because schools and respondents are nested, analyses should account for that structure and report the number of schools and respondents contributing to each estimate. These proposals follow from the study's limitations and are not findings of the current survey.

6. Conclusion

The study reports supervisory strategies as Much Employed overall (4.18), with classroom management the lowest of five domains (3.51). Strategies for bridging master teachers' observation standards were Much Evident overall (4.41), led by technical assistance (4.58) and lowest for professional development support (4.25). The reported rank tests identified instructional planning and three bridging domains as significant at .05. These findings are descriptive and provisional: participant counts, item mappings, rank sums, and several inferential probabilities need author verification. The evidence can guide review of feedback, follow-up, and professional-learning support, but it does not establish improvements in teaching or observation outcomes.

References

- [1] Department of Education. (2016). DO 35, s. 2016: The Learning Action Cell as a K to 12 Basic Education Program school-based continuing professional development strategy for the improvement of teaching and learning. <https://www.deped.gov.ph/2016/06/07/do-35-s-2016-the-learning-action-cell-as-a-k-to-12-basic-education-program-school-based-continuing-professional-development-strategy-for-the-improvement-of-teaching-and-learning/>
- [2] Department of Education. (2017). DO 42, s. 2017: National adoption and implementation of the Philippine Professional Standards for Teachers. <https://www.deped.gov.ph/2017/08/11/do-42-s-2017-national-adoption-and-implementation-of-the-philippine-professional-standards-for-teachers-2/>
- [3] Hargreaves, A., & O'Connor, M. T. (2018). Collaborative professionalism: When teaching together means learning for all. Corwin Press. <https://www.corwin.com/books/collaborative-professionalism-247835>
- [4] Organisation for Economic Co-operation and Development. (2019). TALIS 2018 results (Volume I): Teachers and school leaders as lifelong learners. OECD Publishing. <https://doi.org/10.1787/1d0bc92a-en>