Self-Assessed and Observed Instructional Competence of Teachers: Basis for Targeted Technical Assistance

A basic research proposal · Basic Education Research Fund 2026 (school level)

[Researcher 1], School Principal IV · [Researcher 2], Assistant School Principal II
[Researcher 3], Teacher III, School Research Coordinator
[Name of school], [Schools Division Office], Department of Education

A sample with invented data, for teaching: nash.com.ph/blog

Introduction and rationale: are we helping teachers where they need it?

17%

of school division superintendents said division training was based on teachers' actual needs

World Bank (2016), p. 46

65% vs 25%

of Grade 1 to 3 teachers who believed they check understanding throughout the lesson, against those observed doing it

Molina et al. (2018), p. 37

114

Teachers I to VII in our school, each with an e-SAT on 37 indicators and full-period observations this school year

School records; DM 089, s. 2025



Our school plans its technical assistance from the e-SAT consolidation and the IDPs. No one has checked whether teachers' self-ratings point to the same needs as their observed lessons.

Literature review: what self-assessments tell, and miss

Dunning, Heath & Suls (2004)

Self-views hold "only a tenuous to modest relationship" with actual performance; people tend to overrate themselves (p. 69).

World Bank (2016), PETS-QSDS

Philippine teachers rated their subject knowledge very highly, in "stark contrast" with their test results; self-ratings may not be a good basis for planning PD (p. 43).

Molina et al. (2018)

In Mindanao classrooms, 72% of teachers rated themselves effective at adjusting to learners' level; 10% scored high when observed (p. 38).

Van Zoest et al. (2002)

Teachers saw themselves closer to exemplary than observers did, and the gap shrank as their practice improved (p. 245).

Literature review: observing, ranking needs, and targeting help

Kane & Staiger (2012)

A single observation is a noisy measure: reliabilities of 0.14 to 0.37 in the MET project (p. 8).

Borich (1979)

A training need is a discrepancy, weighted by its importance; check self-ratings against observation before acting on them (pp. 3, 7–8).

Kraft, Blazar & Hogan (2018)

Coaching raised instruction by 0.49 SD and achievement by 0.18 SD across 60 causal studies; more when paired with group training.

Desimone & Garet (2015)

Aim for "collective experiences that are targeted to groups of teachers with similar needs" (p. 255).

Four questions, four null hypotheses

1. How do our teachers rate their capability and priorities on the 37 PPST indicators (e-SAT), and what ratings did the first full-period observation give on its six indicators?
2. How closely do self-assessed and observed needs agree, and are years in teaching and career stage related to needs the e-SAT does not flag?
3. What practices do the observers' notes record, and do lessons with each practice receive different ratings?
4. Which needs should our technical assistance address first, and in what form; and how do teachers rate the assistance they received?

Null hypotheses (α = .05)

- H01: For each indicator, the shares with a self-assessed and an observed need are equal.
- H02: Years in teaching and career stage are not related to unflagged needs.
- H03: Ratings do not differ with and without each coded practice.
- H04: Satisfaction with the post-observation conference does not differ by its delay.

What the study covers, and what it cannot show

Scope

All 114 Teachers I to VII of our school who consent; their e-SAT of November 2025; their first full-period observation of SY 2025–2026, its notes and its conference date; a questionnaire after the school year.

Limitations

One observation per teacher measures one lesson; the e-SAT's four levels are undefined; only six indicators are observed in the first observation; one school; master teachers and head teachers are the raters, not participants.

Research design: four records the school already keeps, one code

e-SAT 2025–2028
November 2025 · Capability and priority, 37 indicators

COT, first observation
October 2025 onward · Ratings on six indicators

Observation notes
Same lesson · What the rater saw

PMCF
Same lesson · Date of the post-observation conference

The custodian
Links records under random codes; replaces raters' names; keeps the key

Coded data set
What the researchers see

+ Questionnaire
after the school year

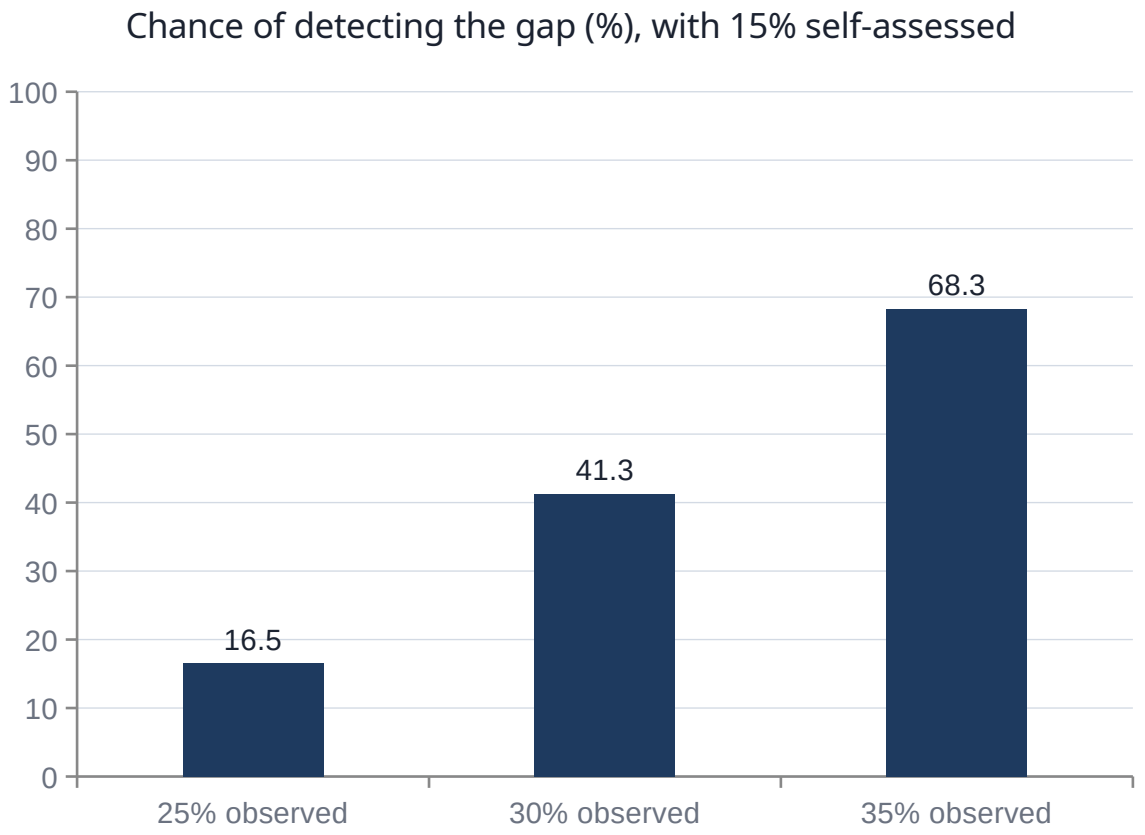
No new test and no extra observation: the study reads records the PMES already produces, linked by a code that only the custodian can trace, and adds one short questionnaire.

Sampling: every Teacher I to VII; what 100 teachers can detect

114 Teachers I to VII invited: no sampling

~100 expected to consent; master teachers and head teachers are raters, not participants

6 indicators in the first observation, compared with the same six in the e-SAT



Power of McNemar's exact test at Holm's strictest threshold (.05 / 6), 100 teachers, 15% flagging themselves, 20,000 simulations.

Data collection: the school's records, the notes, one questionnaire

The school's PMES records

The e-SAT (capability and priority on the 37 PPST indicators), the first observation's ratings converted with DM 089's Table 13, and the PMCF date of the post-observation conference. A teacher's IPCRF rating is not used.

The observation notes and a codebook

Ten codes for practices tied to four indicators (higher-order questions, tiered tasks, group roles, reading and numeracy strategies), coded by two master teachers of a neighboring school who do not see the ratings.

The technical assistance questionnaire

Six forms of help received, each rated on six statements adapted from ARTA's Client Satisfaction Measurement; three preferred forms for next year; one open question. About ten minutes.

Data collection: validity, reliability, schedule

Codebook pilot

Tried on ten redacted notes from SY 2024–2025 and revised before the formal coding.

Thresholds now

Krippendorff's alpha .800 for a conclusion, .667 for a tentative one, fixed in this proposal.

Questionnaire review

Two teachers outside the study read every item for clarity.

Consent sessions after the MOA, February 2026. Records copied by the custodian after the school year ends and the teachers' break, May 2026; the questionnaire in the same weeks. Analysis May to June; the plan ready for the raters' IS Plans in June.

Ethical issues: consent, privacy, no pressure, review



Consent without supervisors

The research coordinator, who rates no one, takes consent with no rater or school head present; sealed forms; the researchers never learn who agreed (PHREB, 2022).



A custodian

The e-SAT focal person, who already holds the files, links records under random codes and replaces raters' names; the e-SAT itself says the school head may not see a teacher's results.



No decision about anyone

Results are for groups; the plan names indicators and counts, never a teacher. Refusing changes nothing in a teacher's PMES.



Review

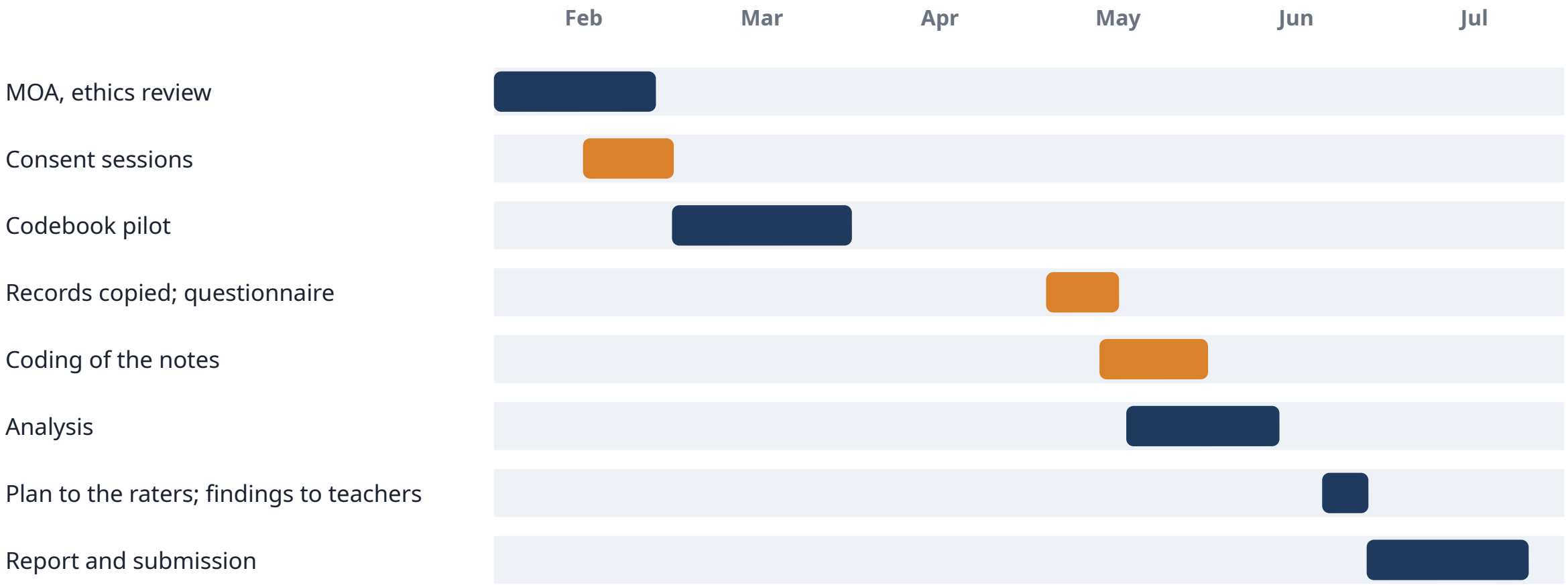
A PHREB-accredited research ethics committee; consent specific to the records named (NPC Circular No. 2023-04).

Plan for data analysis: one named test for each question

Q1	Shares at each e-SAT level; observed ratings by indicator; Friedman test across the six.
Q2	Four groups per indicator; McNemar's exact test with Holm's adjustment; kappa with bootstrap intervals, PABAK, Spearman's rho; GEE logistic model of unflagged needs; raters' variance share.
Q3	Krippendorff's alpha per code; Mann-Whitney tests of the ratings with and without each practice; rank-biserial r; Holm.
Q4	Borich-type discrepancy scores for all 37 indicators, checked by a ranked discrepancy model; the plan's rule (40%, 15%); satisfaction as ARTA scores it, with Wilson intervals; Fisher's test and logistic regression of satisfaction on the conference's delay.

Python (pandas, SciPy, statsmodels) and Excel, both free or already licensed; every statistic Excel can compute will be a formula in the executive data set.

Timetable: February to July 2026 (six months)



Orange: work with teachers and records. All months are of 2026; the school year ends March 31, and the teachers' break runs through April.

Cost estimates: PhP 18,000 in eligible items

Particulars	Category	Amount (PhP)
Food during coding and consent sessions	Food	4,620.00
Questionnaire copies, report copies	Reproduction	3,600.00
Dissemination: LAC session, raters' meeting	Dissemination	3,000.00
Coders' fares, six coding days	Travel	2,880.00
Consent forms, envelopes, coding sheets	Supplies	2,400.00
Prepaid load: custodian, coders, online form	Communication	1,500.00
Total		**18,000.00**

Not charged to the grant

- No equipment, software or honoraria: none is eligible under DO 16, s. 2017
- Tranches 80% and 20% (school level)
- Coders' time through the district's partnership between schools

Dissemination and advocacy: who needs the findings

Our teachers

A LAC session: the findings for groups, the practices in the codebook, the plan

Our fourteen raters

A planning meeting: the plan's rows for their IS Plans; the codebook as walkthrough look-fors

The Schools Division Office

The report, the data set and a two-page brief to the SDRC

Other school heads

The district school heads' meeting: comparing e-SAT and observation records with consent

References

Borich, G. D. (1979). A needs assessment model for conducting follow-up studies of teacher education and training. University of Texas at Austin.

Department of Education. (2017). Research management guidelines (DO 16, s. 2017).

Department of Education. (2025). Guidelines on the multi-year PMES for teachers from SY 2025–2026 to 2027–2028 (DM 089, s. 2025).

Desimone, L. M., & Garet, M. S. (2015). Best practices in teachers' professional development in the United States. Psychology, Society, & Education, 7(3), 252–263.

Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment. Psychological Science in the Public Interest, 5(3), 69–106.

Kane, T. J., & Staiger, D. O. (2012). Gathering feedback for teaching. Bill & Melinda Gates Foundation.

Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement. Review of Educational Research, 88(4), 547–588.

Krippendorff, K. (2004). Reliability in content analysis. Human Communication Research, 30(3), 411–433.

Molina, E., Fatima, S. F., Trako, I., & Wilichowski, T. (2018). Teacher practices in Mindanao: Results of the Teach classroom observation study. World Bank.

Philippine Health Research Ethics Board. (2022). National ethical guidelines for research involving human participants 2022.

Van Zoest, L. R., Breyfogle, M. L., & Ziebarth, S. W. (2002). Self-perceived and observed practices of secondary school mathematics teachers. Teacher Development, 6(2), 245–268.

World Bank. (2016). Assessing basic education service delivery in the Philippines (Report No. AUS6799).

Thank you

[Researcher 1] · [DepEd email address] · [mobile number]
[Researcher 2] · [DepEd email address]
[Researcher 3] · [DepEd email address]
[Schools Division Office], [address]