

The Effectiveness of the Strategies Employed by the Key Stage 2 Teachers in Teaching Physical Education in Buenavista District II, Division of Quezon

Jerwin F. Marcaida¹

¹School of Graduate Studies, University of Northeastern Philippines

ARTICLE INFORMATION	ABSTRACT
Article history: Published: October 2026 Keywords: Physical Education Key Stage 2 Teaching Strategies Movement Skills Teacher Perceptions Kendall Concordance	The study described Key Stage 2 teachers' Physical Education strategies and perceived effectiveness in eleven public elementary schools of Buenavista District II, Division of Quezon, during school year 2025–2026. Forty teachers were included through complete enumeration. A descriptive–evaluative–correlational survey used a 120-item questionnaire, weighted means, and Kendall concordance of three school zones' item rankings. Overall strategy use was 4.38 (Much Employed), led by gamification (4.61). Perceived effectiveness averaged 4.26 (Much Effective), with skill performance highest (4.54) and movement skills lowest (3.42, Effective under the stated scale). None of five strategy domains showed significant rank concordance at .05; only the physical-fitness effectiveness domain did ($W = 0.66$, $\chi^2 = 19.80$). The survey provides teacher appraisals rather than independently measured learner fitness or achievement. These results support closer observation of movement development and differentiated practice, but do not show that the reported strategies caused improvement.

1. Introduction

Quality Physical Education calls for inclusive participation and the development of movement competence, rather than merely enjoyable activity. UNESCO (2015) outlines a policy approach to meaningful participation, equality, and safeguarding. Teachers in Buenavista District II reported how they used gamification, differentiated instruction, cooperative learning, movement exploration, and peer teaching, and rated the effectiveness of their practice for five learner-related domains.

The two sets of ratings serve different purposes. One describes perceived use of teaching strategies; the other is a judgment about movement skills, participation, physical fitness, skill performance, and learning achievement. A high rating in the second set does not directly measure a pupil's performance or change from baseline. The study also tested whether three school zones ranked the component statements in each domain similarly, a question about agreement in priorities rather than a treatment effect.

The Philippine Professional Standards for Teachers provide a local frame for lesson design, differentiation, and assessment (Department of Education [DepEd], 2017). Applying this frame would require observation and evidence of learner work, which the survey did not report. The study's descriptive and concordance evidence can identify practices worth checking during lesson review. The study retains the reported statistics and distinguishes perceived effectiveness from observed physical and learning outcomes.

2. Methods

2.1 Design and participants

The study employed a descriptive–evaluative–correlational survey in eleven public elementary schools of Buenavista District II for school year 2025–2026. All 40 Key Stage 2 teachers handling Physical Education were selected through complete enumeration. Zone 1 comprised eleven teachers in three schools, Zone 2 fourteen in four schools, and Zone 3 fifteen in four schools. No pupil sample, direct fitness assessment, or baseline and follow-up performance record was included in the reported analysis.

Teachers' ratings were the unit for descriptive questionnaire summaries; the concordance analysis compared three school-zone rankings of component indicators. These are distinct levels of analysis. The zone counts are small and schools are nested in zones, so a claim about equal district-wide implementation should not be based on the W tests alone.

2.1 Instrument and data collection

The two-part researcher-made questionnaire specified 120 items. Part I had 60 strategy-use items across gamification (11), differentiated instruction (12), cooperative learning (13), movement exploration (12), and peer teaching (12). Part II had 60 effectiveness items across movement skills (11), participation (13), physical fitness (12), skill performance (11), and learning achievement (13). Five-point categories classified 4.50–5.00 as Very Much Employed or Very Much Effective, 3.50–4.49 as Much Employed or Much Effective, and 2.50–3.49 as Employed or Effective.

The instrument was reviewed by experts and piloted according to the methods. An item correlation threshold of 0.50 and Cronbach's alpha of 0.70 are stated, without the pilot sample, confidence interval, domain-specific estimates, or item data. The questionnaire was administered and tabulated, but dates and missing-response rules are not detailed. The term "effectiveness" is therefore retained as the respondents' rating construct, not an externally verified impact.

2.2 Statistical treatment

Frequency, percentage, weighted mean, and rank described responses. Kendall's W and chi-square tests evaluated concordance of

indicator order among three zone profiles within each domain at .05. A nonsignificant test fails to establish common ordering; it is not evidence of a difference in the overall mean. Neither *W* nor the survey's design establishes a causal effect of strategy use on fitness, performance, or achievement. Statistics in Tables 3 and 5 are printed values, not recomputations from raw responses.

3. Results

3.1 Respondents

Table 1. Key Stage 2 Physical Education teachers by zone

Zone	Schools	Teachers	Printed %
Zone 1	3	11	28
Zone 2	4	14	35
Zone 3	4	15	37
Grand total	11	40	100

Note. Numbers and rounded shares follow the respondent table.

The three groups supplied comparable numbers of teachers, but Zone 1 had one fewer school. Table 1 helps establish the setting of the perceptions; it does not describe pupil age, physical readiness, resources, or class size. Those contextual variables could be important for interpreting strategy selection and actual movement performance, but were not entered into the reported tests.

3.2 Strategy use

Table 2. Summary of teaching-strategy ratings

Domain	Overall <i>W_x</i>	Scale category
Gamification	4.61	Very Much Employed
Differentiated Instruction	4.47	Much Employed
Cooperative Learning	4.31	Much Employed
Movement Exploration	4.31	Much Employed
Peer Teaching	4.21	Much Employed
Overall average	4.38	Much Employed

Note. Domain means and overall mean follow the detailed Table 8 narrative; cooperative learning and movement exploration are two domains with the same 4.31 mean.

Gamification was the only strategy in the Very Much Employed band. Peer teaching was the lowest of the five, although it remained Much Employed. The difference suggests an opportunity to inspect how students receive feedback and practice a skill with classmates. It does not mean that games were more effective than peer teaching, because the table measures reported use rather than comparative effects. The two domains tied at 4.31 should also be distinguished in lesson observation: cooperative tasks and exploration may involve quite different opportunities to practice.

Table 3. Kendall concordance of strategy-item rankings

Domain	<i>W</i>	χ^2	Printed threshold
Gamification	0.37	9.99	$p > .05$
Differentiated Instruction	0.08	2.40	$p > .05$
Cooperative Learning	0.12	3.96	$p > .05$
Movement Exploration	0.28	8.40	$p > .05$
Peer Teaching	0.45	13.50	$p > .05$

Note. All five nonsignificant decisions at .05 and figures follow detailed Table 9 and the study abstract.

There was no significant agreement in the ordering of strategy items across zones. Differentiation and cooperative learning had especially low reported *W* values, indicating that the same component practices were not ranked consistently under the test. This finding should motivate examination of the exact item ranks and local lesson constraints. It is not a test showing that one zone uses differentiation less overall; the reported means and the *W* statistics answer different questions.

3.3 Perceived effectiveness

Table 4. Summary of perceived-effectiveness ratings

Domain	Overall <i>W_x</i>	Scale category
Movement Skills	3.42	Effective
Learner Participation	4.36	Much Effective
Physical Fitness	4.46	Much Effective
Skill Performance	4.54	Very Much Effective
Learning Achievement	4.53	Very Much Effective
Overall average	4.26	Much Effective

Note. Means are from the detailed Table 15 narrative. The study's 2.50–3.49 band makes 3.42 Effective, although one passage calls it Much Effective.

Skill performance and learning achievement crossed the Very Much Effective cut point, but movement skills was substantially lower and in the Effective band. These broad constructs may overlap in everyday language; the item definitions and assessment rubrics would be needed to know why teachers rated skill performance at 4.54 but movement skills at 3.42. Without direct pupil tests, the contrast identifies a measurement and teaching question, not a confirmed discrepancy in learners' actual capabilities.

Table 5. Kendall concordance of perceived-effectiveness item rankings

Domain	W	χ^2	Printed threshold	Decision
Movement Skills	0.49	13.23	$p > .05$	Not significant
Learner Participation	0.12	3.60	$p > .05$	Not significant
Physical Fitness	0.66	19.80	$p < .050$	Significant
Skill Performance	0.25	6.75	$p > .05$	Not significant
Learning Achievement	0.32	10.56	$p > .05$	Not significant

Note. Figures and decisions follow detailed Table 16. The reported significance strings are thresholds, not exact probabilities. Only physical fitness showed significant rank concordance across the three zones. The zones may agree on which physical-fitness statements are most salient while still offering different activities and resources; that cannot be distinguished from W alone. Nonsignificant agreement for movement and participation should not be described as proof of disagreement. The inference is limited to the ordering of the questionnaire indicators, and the test did not evaluate learner-level improvement.

3.4 Movement-skill detail

Table 6. Selected movement-skill effectiveness items

Statement in detailed findings	Overall Wx	Scale category
Movement tasks are performed according to instructions	3.59	Much Effective
Physical movements are executed with confidence	3.58	Much Effective
Better coordination and balance are evident during activities	3.50	Much Effective
Movement skills are applied in different physical activities	3.48	Effective

Note. Selected means and labels are from the movement-skills findings. These are teacher-rated indicators.

Performance according to instruction rated above transfer of a movement to different activities. The distinction is pedagogically useful: following a prompt in a familiar drill and applying a skill in a new game are not identical achievements. A targeted classroom assessment could document both under common scoring criteria. The survey does not reveal whether lower transfer ratings reflect practice time, task complexity, or assessment expectations.

3.5 What the strategy items may capture

Table 7. Selected gamification indicators

Statement in findings	Overall Wx	Scale category
Game-based activities are incorporated into PE lessons	4.70	Very Much Employed
Game elements increase learners' interest in PE	4.67	Very Much Employed
Interactive challenges are included in PE activities	4.66	Very Much Employed
Learners develop movement skills through games	4.65	Very Much Employed
Points or scores are used to motivate learners	4.61	Very Much Employed
Educational games make learning enjoyable	4.53	Very Much Employed

Note. Item means and labels follow the detailed gamification findings. PE = Physical Education; wording is condensed.

Every selected game item was in the highest use category. Yet these statements span teacher actions, inferred learner interest, and perceived movement development. Combining them into one utilization domain can make it difficult to distinguish what teachers did from what pupils achieved. The high 4.70 rating for incorporating games is an implementation report; the 4.65 rating that learners developed skills is already an outcome judgment. A subsequent instrument could separate these constructs and use observed tasks to score actual movement.

A game may give repeated movement opportunities or may emphasize winning and scoring. The questionnaire does not specify time active, inclusion of children at different skill levels, or feedback on safe technique. UNESCO (2015) highlights equitable, meaningful participation as a quality concern. In an observational follow-up, teachers could record how many pupils practiced the target movement and whether the game rules allowed all of them to participate. Such records would turn the high utilization rating into an inspectable practice.

Table 8. Selected learning-achievement and movement indicators

Statement in detailed findings	Overall Wx	Scale category
Learning objectives are attained during PE lessons	4.71	Very Much Effective
Concepts taught in PE are understood by learners	4.62	Very Much Effective
Learners follow and apply instructions correctly	4.62	Very Much Effective
Movement tasks are performed according to instructions	3.59	Much Effective
Movement skills are applied in different physical activities	3.48	Effective

Note. Selected reported item means are from the learning-achievement and movement-skills findings; categories follow the study's stated scale. The contrast between high judgments about attainment of objectives and lower ratings for movement application in different activities deserves careful interpretation. The statements may refer to distinct goals, and teachers may have different evidence for a written concept, a prompted skill, and transfer to another game. The survey does not show the criteria used to decide that an objective was attained. A common assessment task could score knowledge and motor performance separately and avoid treating a high rating for one as proof of the other.

The selected movement items also suggest a progression for instruction: demonstrate a task under clear guidance, practice it with feedback, then apply it in a new setting. The data do not establish that this sequence occurred or that a new strategy would improve it. Nonetheless, the lower transfer item offers a precise question for a lesson study or coaching cycle, with independent scoring of pupils before and after the cycle.

4. Discussion

The research question about strategy use showed broad self-reported uptake and a lead for gamification. UNESCO's (2015) quality Physical Education guidelines emphasize meaningful, inclusive participation. A game format can provide participation, but its instructional quality depends on the movement objective, feedback, safety, and inclusion. The study did not observe lessons or compare pupils exposed to particular strategies, so the 4.61 gamification rating is a utilization result rather than evidence of greater effectiveness.

The perceived-effectiveness question revealed a notable distinction between skill performance and movement skills. The domains may have been interpreted differently by teachers, or the items may ask about practice under varying conditions. An independent rubric for balance, coordination, transfer, and skill execution would make the claims testable. Comparing a baseline to later demonstrations across multiple tasks would also reduce the risk that high questionnaire ratings reflect a general favorable response style.

The concordance findings are about zone rank orders. Physical-fitness items showed the sole significant agreement among effectiveness domains, while no strategy domain did. This does not provide a mechanism linking a specific strategy to fitness. The questionnaire's descriptive-evaluative-correlational label does not substitute for a paired correlation or an intervention comparison. A subsequent study could match each teacher's lesson practices to direct pupil measures and account for pupils clustered within classes and schools.

The study's practical recommendations—differentiation, cooperative learning, guided practice, and teacher coaching—are reasonable areas to consider, but none was tested as an intervention. PPST can guide a focused observation on objectives, adaptation for learners, and assessment (DepEd, 2017). An observation cycle could specify a movement competency, record tasks and feedback, score performance independently, and revisit the same skill. That sequence would produce evidence of instructional change and learner development beyond present perceptions.

The findings are limited to 40 teachers in a single district, with no direct learner sample or outcome. The instrument's reported alpha of 0.70 lacks pilot details and subscale estimates. The study's category wording for 3.42 conflicts with its scale and another summary; the scale-based label is preserved with this conflict disclosed. Exact test p values and raw item rankings were not provided. The study can guide local inquiry, but it cannot support a causal effectiveness claim or an effect size for Key Stage 2 pupils.

The contrast between reported strategy use and perceived effectiveness should not be treated as a dose–response relationship. The domain means summarize different sets of statements; no teacher-level correlation between, for example, gamification use and a pupil's movement result was calculated. Even if such a correlation were found, more motivated teachers or better-resourced schools could influence both measures. A future evaluation should define a particular game-based lesson and compare equivalent learning opportunities, not treat “games” as a uniform intervention.

The high learning-achievement mean may reflect a broad favorable judgment on lesson objectives, whereas the movement-skills mean captures more demanding applications. Clarifying each construct's item content is essential before journal submission. The study's policy suggestions for coaching and differentiated practice are practical hypotheses. They would gain support from observation records, scored skill demonstrations, pupil participation data, and an analysis of how facilities and class size shape opportunities for practice.

High strategy-frequency ratings describe teachers' reports rather than time pupils spent practicing. Incorporating games could mean a brief warm-up, an extended movement sequence, or a competitive activity in which participation is uneven. The questionnaire does not identify lesson duration or the proportion of children active. An observation schedule could record the target skill, practice opportunities, feedback, safety, and inclusion. This would provide a more interpretable exposure measure than a general five-point utilization statement.

Knowledge and motor performance should be distinguished for the effectiveness question. Selected items rated objective attainment and concept understanding highly, while movement application in varied activities was lower. A pupil might explain a rule yet need practice to execute a movement in a new setting, or perform a movement without articulating the formal concept. These are possible interpretations of the constructs, not observed pupil profiles. A balanced PE assessment could include a concept check, a scored skill demonstration, and a transfer task.

The five outcome domains were based on teacher perceptions, with movement skills averaging 3.42. Calling that score Much Effective in one passage conflicts with the stated category boundary, which places it in Effective. The numerical value and band should be checked before submission. The tables suggest uniformly high reported strategy use and more varied perceived effectiveness. They do not reveal which strategy produced which outcome or whether a pupil's ability changed over time.

Rank agreement across three zones was generally strong in strategy domains, but the significant physical-fitness result should not be generalized to every outcome. Its printed W of 0.66 and chi-square of 19.80 stand apart from weaker or nonsignificant outcome-domain results. Agreement concerns relative item ordering, not whether one zone's children were fitter or whether strategies caused change. Statement counts, ties among rounded means, and thresholds should be transparent for each test. Teacher and pupil units need separation in subsequent analysis.

A next study could define one coherent strategy, train teachers in the same lesson sequence, and assess a specific movement before and after implementation. Attendance, facilities, class size, and opportunities for practice would make exposure differences interpretable. A comparison or staggered implementation would strengthen causal analysis. UNESCO (2015) emphasizes quality and participation in PE; recording who participates alongside aggregate performance would address that concern. The present survey offers a descriptive starting point, not evidence of a tested intervention effect.

The measurement scale merits closer definition. The study treats broad strategy categories such as gamification as if the category denotes a uniform method, yet games differ in rules, feedback, practice repetitions, and physical demands. A construct map could separate the teacher's instructional action from the pupil's participation and the skill demonstrated. Independent observers could then score the action and participation, while a separate rubric scores movement. This would reduce the risk that the same teacher's favorable impression drives both the strategy-use and effectiveness ratings. Agreement between observers could be assessed before comparing classrooms.

The 40 teachers represent a local set of Key Stage 2 classrooms, not an individually randomized intervention. Classroom context, available space, weather, equipment, and pupil experience may influence how a strategy operates. The survey does not report these variables in a linked dataset, so its descriptive means cannot be adjusted for them. A useful follow-up could compare a defined movement skill across lessons with different practice structures while documenting those conditions. A stated analysis plan would define the unit of analysis, primary outcome, handling of absent pupils, and time interval. This would make an effectiveness claim testable rather than inferred from high agreement with questionnaire statements.

The weaker movement-skills rating can support reflective professional development even without an effect estimate. Teachers could review video or observation notes on demonstration, feedback, active time, and the opportunity to repeat a skill. They could then examine the distribution of pupil performance rather than only a class average. A high mean can hide children who participate less or need adaptations. The study did not collect such observations, so these are proposed applications of its descriptive pattern. The reported outcome-domain tests should not be simplified into a blanket finding that all physical education strategies are effective.

5. Conclusion

Teachers rated their Physical Education strategies Much Employed overall (4.38), with gamification highest (4.61). They rated perceived effectiveness Much Effective overall (4.26), although movement skills stood at 3.42 (Effective). Only the physical-fitness domain showed significant agreement in zone item rankings. The ratings identify a need to examine movement transfer, differentiation, and feedback with direct pupil evidence; the study did not demonstrate learner improvement caused by the strategies.

References

- [1] 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/>
- [2] United Nations Educational, Scientific and Cultural Organization. (2015). Quality physical education: Guidelines for policy-makers. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000231101>