

The Utilization of the Story-Based Learning Approach in Enhancing the Reading Comprehension Skills of Key Stage 1 Learners in Buenavista District II, Division of Quezon

Andrea L. Fuenzalida¹

¹School of Graduate Studies, University of Northeastern Philippines

ARTICLE INFORMATION	ABSTRACT
Article history: Published: October 2026	The study described Key Stage 1 teachers' utilization of story-based learning and their appraisals of reading comprehension skills in eleven public elementary schools of Buenavista District II, Division of Quezon, during school year 2025–2026. A descriptive–evaluative–correlational survey included 52 teachers through complete enumeration. A 120-item questionnaire, weighted means, and Kendall concordance tests compared the item rankings of three school zones. Story-based learning use averaged 4.41 (Much Utilized), with active participation highest (4.50) and improved retention lowest (4.35). The perceived comprehension domains averaged 4.28 (Much Evident); recognizing supporting details was highest (4.37) and identifying main ideas lowest (4.22). None of the ten domain-level concordance tests was labeled significant at .05. The results describe teacher perceptions and zone rank orders, not measured gains caused by the approach.
Keywords: Story-Based Learning Reading Comprehension Key Stage 1 Teacher Perceptions Main Ideas Kendall Concordance	

1. Introduction

Stories can provide a context for discussion, sequencing, prediction, and retelling. Their usefulness for comprehension depends on the questions teachers pose, opportunities for pupils to explain, and the reading tasks used afterward. The Institute of Education Sciences (2010) recommends explicit comprehension instruction in the early grades and guided discussion of text. The study examined how Key Stage 1 teachers in Buenavista District II reported the use of story-based learning and appraised five comprehension-related skills.

The approach was operationalized through active participation, improved retention, versatile application, critical thinking, and empathy. The second set of domains concerned identifying main ideas, recognizing supporting details, sequencing, predicting outcomes, and retelling or summarizing. A story activity can elicit enthusiastic participation yet leave a child uncertain about the central idea. Thus the two sets of questionnaire ratings should be interpreted separately, and neither is a standardized reading comprehension score.

The study also asked whether three school zones agreed in their rank order of component statements. Kendall concordance examines that ordering; it cannot establish the benefit of story-based lessons compared with another approach or the amount of pupil growth. The study reports the detailed numerical summaries of the respondent groups and relevant reading-comprehension dimensions.

2. Methods

2.1 Study design and participants

The study employed a descriptive–evaluative–correlational questionnaire design for school year 2025–2026. It completely enumerated 52 Key Stage 1 teachers in eleven public elementary schools of Buenavista District II. Zone 1 comprised 20 teachers in three schools, and Zones 2 and 3 each comprised 16 teachers in four schools. The printed shares for the three zones are 33, 27, and 27 percent, totaling 87 rather than 100. Table 1 retains these shares without silently recalculating them.

No individual pupil sample, pretest, posttest, comparison group, or independent reading assessment was reported. A teacher's favorable judgment of a skill may be informed by classroom experience, but it cannot be equated with a scored change for a particular learner. Schools are nested within zones, and the descriptive profile need not generalize beyond this district.

2.2 Questionnaire and procedure

The researcher-made 120-item questionnaire had 60 items on approach utilization: active participation (12), improved retention (11), versatile application (12), critical thinking (13), and empathy (12). Another 60 items concerned the perceived comprehension domains: main ideas (12), supporting details (12), sequencing (11), predicting (12), and retelling/summarizing (13). Responses followed a five-point scale in which 4.50–5.00 was Very Much Utilized or Very Much Evident and 3.50–4.49 was Much Utilized or Much Evident.

Expert review, item analysis, and internal-consistency testing were described. The methods report all item correlations at least 0.50 and Cronbach's alpha of 0.70, but do not supply the pilot sample, domain-specific estimates, confidence intervals, or raw scores. Following permissions, the questionnaires were distributed, retrieved, and tabulated. The results do not include separately analyzed observations or pupil work. Exact data-collection dates and missing-response procedures were not specified.

2.3 Statistical treatment

Counts, percentages, weighted means, and ranks summarized the survey. Kendall's W and chi-square approximations assessed whether three zone-level indicator rankings agreed within each domain at .05. These tests use ranks of statements, not 52 independent teachers as rankers. A nonsignificant W does not demonstrate a difference in average skill level between zones; it indicates that consistent item order was not established by the reported test. All W and χ^2 values below are transcribed from the findings.

3. Results

3.1 Respondents and utilization

Table 1. Key Stage 1 teachers by school zone

Zone	Schools	Teachers	Printed %
Zone 1	3	20	33
Zone 2	4	16	27
Zone 3	4	16	27
Grand total	11	52	100

Note. Counts and percentage labels are reproduced from the respondent table. The three zone percentages sum to 87, not the printed grand total of 100.

The count distribution gives Zone 1 a larger teacher group than either other zone, while the printed percentages cannot describe the 52 respondents correctly. This matters for any interpretation of sampling representation. The Kendall analysis, however, draws on three aggregated zone rankings, and the table offers no pupil-level denominator for appraising comprehension outcomes.

Table 2. Summary of story-based-learning utilization

Domain	Overall W_x	Scale category
Active Participation	4.50	Very Much Utilized
Improved Retention	4.35	Much Utilized
Versatile Application	4.37	Much Utilized
Critical Thinking	4.37	Much Utilized
Empathy	4.44	Much Utilized
Overall average	4.41	Much Utilized

Note. Domain means and 4.41 overall mean follow the detailed summary and abstract. The 4.50 boundary places active participation in Very Much Utilized.

Active participation stood at the boundary for the highest category, while improved retention was lowest. The closeness of the other ratings indicates a generally favorable pattern with little separation among domains. A visible response during a story is easier to notice than what a child retains days later. The questionnaire did not track delayed recall or measure transfer to a new text, so the retention rating should be read as a teacher appraisal.

Table 3. Kendall concordance of story-based-use item rankings

Domain	W	χ^2	Printed threshold
Active Participation	0.48	14.40	$p > .05$
Improved Retention	0.13	3.51	$p > .05$
Versatile Application	0.36	10.80	$p > .05$
Critical Thinking	0.44	14.52	$p > .05$
Empathy	0.39	11.70	$p > .05$

Note. All five nonsignificant decisions at .05 and W - χ^2 pairs follow detailed Table 9.

No use domain showed significant concordance of component-item ranks among zones. Improved retention had the smallest W (0.13), even though its overall utilization mean was favorable. The combined pattern asks whether teachers across schools emphasize different reinforcement practices. It does not establish that story-based activities were absent in any zone or that the observed 4.35 mean represents actual memory performance.

3.2 Perceived comprehension and rank agreement

Table 4. Summary of perceived reading-comprehension domains

Domain	Overall W_x	Scale category
Identifying Main Ideas	4.22	Much Evident
Recognizing Supporting Details	4.37	Much Evident
Sequencing Events	4.32	Much Evident
Predicting Outcomes	4.26	Much Evident
Retelling or Summarizing Texts	4.24	Much Evident
Overall average	4.28	Much Evident

Note. Means follow detailed Table 15 and the findings narrative. The study abstract incorrectly labels 4.37 Very Much Evident.

All five comprehension domains were Much Evident by the stated scale. Recognizing details led, while identifying main ideas was lowest. Concrete details can be recalled without integrating the passage into a central message; this is an interpretive possibility, not a demonstrated cognitive mechanism in the study. Actual oral or written responses to matched texts would be necessary to check whether the gap reflects children's skill or the teacher's ease of observing it.

Table 5. Kendall concordance of comprehension-item rankings

Domain	W	χ^2	Printed threshold
Identifying Main Ideas	0.31	9.30	$p > .05$
Recognizing Supporting Details	0.21	6.30	$p > .05$
Sequencing Events	0.35	9.45	$p > .05$
Predicting Outcomes	0.15	4.50	$p > .05$
Retelling or Summarizing Texts	0.43	14.19	$p > .05$

Note. All five decisions were nonsignificant at .05 in detailed Table 16.

The test did not establish common ordering of comprehension indicators across zones. Prediction had the lowest W, and retelling the largest, but neither passed the stated significance criterion. The table cannot identify which zone's pupils comprehend stories better. It compares teacher-derived rankings of survey statements, while the differences in pupils' responses and the difficulty of passages were not measured.

3.3 Selected pattern for instructional review

Table 6. Selected comprehension domains for a follow-up rubric

Reported domain	Overall Wx	Printed rank
Recognizing Supporting Details	4.37	1
Sequencing Events	4.32	2
Predicting Outcomes	4.26	3
Retelling or Summarizing Texts	4.24	4
Identifying Main Ideas	4.22	5

Note. Means and rank order follow the detailed summary. This table makes the relative ordering visible; it does not add a new outcome.

The ordering suggests a practical rubric could score both the identification of an explicit detail and a summary that connects details to a main idea. The ratings do not state which text, prompting conditions, language, or scoring criteria teachers used. A common passage and clear anchors for each skill would make a later comparison between zones more interpretable than a rank ordering of differently worded questionnaire statements.

3.4 Specific comprehension behaviors

Table 7. Selected identifying-main-idea indicators

Statement in detailed findings	Overall Wx	Scale category
Learners summarize the central message after story discussions	4.25	Much Evident
The main point is explained in learners' own words	4.23	Much Evident
Central message clearly expressed in discussions	4.16	Much Evident
Essential ideas distinguished from less important details	4.11	Much Evident
Key ideas recognized after reading story texts	4.04	Much Evident

Note. Selected item means are from the detailed main-ideas findings; all fall in the stated Much Evident band. The findings narrative gives the domain mean 4.20, whereas the summary gives 4.22.

The item pattern narrows the concern behind the lowest summary domain. Repeating a central message after a discussion rated more favorably than distinguishing essential from incidental details without that prompt. A supported discussion may enable a response that a child cannot yet produce independently. The questionnaire does not specify whether teachers rated the same text, the same prompt, or the same group of children. A small assessment with both guided and independent main-idea tasks would clarify the instructional need.

The detailed main-ideas narrative gives 4.20 as its domain average, while the summary of five comprehension domains lists 4.22. The difference is small but should not be silently standardized. The study's overview retains the summary-table value, and this selected-item table uses only figures explicitly reported in the detailed findings. The author should regenerate both tables from the item dataset and explain whether a different set of statements was analyzed.

Table 8. Selected predicting and detail-recognition indicators

Statement in detailed findings	Overall Wx	Scale category
Details that explain events recognized during reading	4.43	Much Evident
Specific information identified accurately	4.42	Much Evident
Future events anticipated logically from story details	4.35	Much Evident
Possible endings suggested from story development	4.34	Much Evident
Basis of predictions explained clearly	4.24	Much Evident
Predictions adjusted as new information emerges	4.23	Much Evident

Note. Selected item means follow detailed findings. The narrative calls some 4.43 detail items Very Much Evident, conflicting with its 4.50 scale boundary; categories here apply the stated boundary.

Recognizing stated details received somewhat higher appraisals than explaining or revising a prediction. The latter acts require a child to connect evidence to a tentative idea and then reconsider it. This is an analytical interpretation of the item content, not a direct observation of reasoning. A teacher could capture initial and revised predictions from one story and score whether each is supported by text. That would provide more information than a single general rating for "predicting outcomes."

The category conflict in the detailed narrative is material for readers of the tables: 4.43 is below 4.50, so it is Much Evident under the stated bands. If a different cut point was used in scoring, the methods and all verbal interpretations need correction. Until the scale is confirmed, the numerical means are the more stable part of the descriptive record. This illustrates why journal presentation should include a transparent scoring key alongside its reported means.

4. Discussion

The first question found story-based learning Much Utilized overall, led by participation. That result is compatible with stories creating occasions for talk and shared attention, but it says little about the accuracy of a pupil's interpretation. The IES (2010) comprehension guide recommends teaching strategies explicitly and encouraging discussion of text. An instructional review should ask what questions were posed, whether pupils justified an answer from the story, and how feedback developed the explanation.

The second question found no significant concordance across the three zone rankings for any utilization domain. Teachers may select different forms of narrative activity for their learners, or the rankings may be unstable because of small numbers of indicators and compressed favorable ratings. The study did not identify the cause. Reporting the raw item ranks and tie treatment would show where the divergence occurs without treating nonsignificance as evidence that the zones have completely opposite practices.

The third question showed details higher than main ideas and retelling. It is plausible that locating explicit details is easier for children or easier for teachers to observe than summarizing, but the survey cannot distinguish those interpretations. Comprehension also depends on text difficulty, language knowledge, and decoding. A later study could use comparable texts and collect each pupil's answers, coding both literal and integrative comprehension with independent scoring.

The fourth question again concerned zone concordance, and none of the five domains was significant. This result does not compare the zones' pupil reading levels. The strongest inferential claim supported by the reported tests is the absence of demonstrated shared item order under the study's method. A causal impact claim would require defined story-based exposure, a baseline, a follow-up assessment, and an appropriate comparison that addresses differences among teachers and schools.

The study proposed activities, reading resources, mentoring, and stakeholder support. The present data identify retention and main-idea work for closer examination, but they do not establish which proposal will work. Its limitations include a single district, teacher self-report for both constructs, no direct pupil outcomes, an alpha estimate without reproducible pilot detail, and zone percentage labels that do not sum to 100. The abstract also misclassifies a 4.37 result. Those issues require author reconciliation before the findings can be presented as a fully reproducible journal study.

The approach's apparent strength in active participation should be evaluated together with the quality of comprehension responses. A classroom recording or work sample can show whether most pupils, including quieter learners, explain a main idea using text evidence. Counting raised hands alone would overstate meaningful participation. The IES (2010) guide gives a rationale for explicit comprehension strategies and sustained discussion, but the local study did not measure fidelity to a defined story lesson.

The study's percentage and mean inconsistencies do not require abandoning the study, but they affect editorial readiness. A corrected respondent table should show shares derived from 20, 16, and 16 out of 52. The main-idea domain should have one validated value across detailed and summary presentations, and verbal categories should follow the declared cut points. A reproducible analysis file with item-level, school-level, and zone-level calculations would also permit the Kendall tests to be checked. These steps would make the descriptive contribution clearer and prevent readers from mistaking perceived skill for demonstrated reading growth.

The respondent table lists 20, 16, and 16 teachers, totaling 52, but its percentages of 33, 27, and 27 sum to 87. These cannot describe a complete partition of the counts and require author verification. The three zone profiles also limit rank-concordance tests. Such tests ask whether profiles order indicators similarly, not whether any teacher's story use is linked to a child's comprehension score. Story-based learning may combine read-aloud, questioning, retelling, and discussion, but the questionnaire places several actions under broad domain ratings. It does not measure which children received which components, for how long, or at what reading level. The 4.41 utilization mean is a teacher appraisal, not exposure dose. The 4.28 comprehension mean is likewise a perception-based summary, not a standardized outcome. Their difference can motivate inquiry but does not estimate an intervention effect.

The main-idea domain was relatively low and its detailed and summary means differ, 4.20 versus 4.22. Selected items suggest guided discussion may be easier than independently selecting essential details, but no matched tasks were administered. An assessment could ask pupils to identify a central idea without prompts and justify it with story evidence. A shared rubric before and after specified instruction could clarify the need. Existing ratings do not establish the source of difficulty.

Several detailed 4.43 items are called Very Much Evident in prose although the stated threshold begins at 4.50. Numerical ratings should be retained while the scoring key is checked. Explaining or revising a prediction is more specific than guessing an ending. A classroom record could capture an initial prediction, supporting textual evidence, and revision when new details appear. This would provide better evidence of comprehension than a retrospective general rating.

IES (2010) supports explicit comprehension instruction and discussion of text but does not establish that these district practices caused the rated skills. An evaluation would specify the routine, collect baseline vocabulary and word-reading data, use comparable passages, and score independent responses at later points. It could estimate change with uncertainty and account for teacher and school clustering. Until the percentages, means, and categories are reconciled, the defensible conclusion remains descriptive: teachers rated story use favorably and perceived main-idea and evidence-based reasoning as areas for closer instruction.

The five comprehension domains should be distinguished at the task level. Recognizing a detail may require locating stated information, while a main-idea response requires selecting what is central across a passage. Prediction and retelling may additionally involve prior knowledge and oral expression. These demands help explain why the means differ, but the study contains no assessment rubric or response examples to test that explanation. A qualitative sample of pupils' answers, scored by independent raters against shared criteria, could reveal whether lower ratings reflect misunderstanding, limited language for expressing an answer, or inconsistent teacher expectations.

Stories also vary in vocabulary, length, sequence complexity, and familiarity. Without describing the texts, the same rating phrase may represent quite different tasks among teachers. A follow-up could select comparable passages, record whether the text was read aloud or independently, and use a consistent prompt. It could score literal details, central idea, evidence-backed prediction, and retelling separately. The distinction between teacher support and independent comprehension is important: an answer after guiding questions is useful evidence of participation, but it is different from an unaided response. The current questionnaire does not make that distinction.

The stated alpha of .70 requires more context. The study should report the pilot sample, the precise scale or subscales used, and whether the coefficient was calculated on the full instrument or on domains. One coefficient for many items does not establish that use of story-based learning and perceived comprehension are distinct measures. Content review by reading specialists and cognitive interviewing with teachers could clarify ambiguous statements. A refined instrument could then be paired with pupil tasks. This would improve construct validity as well as the credibility of a later association analysis.

The high reported use of the approach may create a ceiling effect. If most teachers choose favorable response options, the instrument has little spread to distinguish instructional patterns or relate them to outcomes. Reporting the frequency of each response category, the number of teachers at the maximum, and item-level variability would expose this limitation. Observation could capture differences that a short rating scale misses, such as the quality of questions and the amount of pupil explanation. The existing overall mean of 4.41 should therefore be read as a positive appraisal, not a precise measure of how much instruction occurred.

The three zones are not equivalent to three independent replications of an intervention. Their teacher counts differ, their schools may have distinct contexts, and the ranking tests summarize item profiles. A comparison of comprehension performance would require a defined learner sample, baseline data, and a model reflecting pupils within classrooms. The proposed story-based activities could be assessed prospectively with a pre-specified primary skill, such as independent identification of a central idea. Results could then include uncertainty intervals and implementation records. Even a positive change without a comparison would require caution about maturation and other instruction during the same period.

These measurement steps have immediate instructional relevance. Teachers can ask pupils to point to evidence for a detail, explain why it supports a prediction, and retell the central sequence in their own words. Recording responses across a few common texts can show which part of comprehension needs support. The study's lower appraisal for identifying main ideas makes that a reasonable starting question, but its inconsistent domain mean and percentage table should be corrected from the underlying records. Until then, the study presents the numerical results as printed and limits its conclusion to perceived practice and skills.

5. Conclusion

Story-based learning use was rated Much Utilized overall (4.41), and comprehension-related skills Much Evident overall (4.28). Participation led approach ratings (4.50); recognition of details led comprehension ratings (4.37), while main ideas was lowest (4.22). None of the ten reported zone rank-concordance tests was significant. The findings support a focused review of reinforcement and main-idea instruction, but they do not demonstrate reading gains attributable to stories. Percentage and verbal-category inconsistencies require correction.

References

- [1] Institute of Education Sciences. (2010). Improving reading comprehension in kindergarten through 3rd grade. U.S. Department of Education. <https://ies.ed.gov/ncee/wwc/PracticeGuide/14>