

The Effectiveness of the Utilization of the Artificial Intelligence (AI) Powered Chatbots in Enhancing the Vocabulary Acquisition of the Grade 7 Learners in the Public Junior High Schools in Calabanga Districts, Division of Camarines Sur

Samuel T. Palivino¹

¹School of Graduate Studies, University of Northeastern Philippines

ARTICLE INFORMATION	ABSTRACT
Article history: Published: October 2026	The study examined teachers' ratings of AI-powered chatbot utilization and perceived effectiveness for Grade 7 vocabulary acquisition in public junior high schools in Calabanga Districts during school year 2025–2026. Thirty Grade 7 teachers completed a researcher-made rating scale. Frequency, percentage, weighted mean, rank and Kendall's W with chi-square were used. Overall chatbot utilization was rated Utilized (3.37), with accessibility and ease of use highest (3.67) and learner engagement lowest (3.07). Teachers rated perceived effectiveness for vocabulary acquisition Much Effective overall (3.72); comprehension was highest (3.88), while word recognition was lowest (3.56). The study characterizes teacher perceptions and concordance of indicator ordering rather than measured vocabulary test gains or an experimental effect of AI chatbot exposure.
Keywords: AI-Powered Chatbots Philippines School-Based Surve Kendall's Coefficient of Concordance Teacher Perceptions	

1. Introduction

In the Philippine basic education system, AI-powered chatbots and Grade 7 vocabulary acquisition is situated at the intersection of classroom practice, school policy and local resource conditions. The way educational practices are organized within a district cannot be inferred from national policy statements alone; information on local implementation and the views of the personnel responsible for delivery is also necessary. The study therefore focused on public junior high schools in Calabanga Districts, Division of Camarines Sur, a setting in which the routines supporting accessibility and ease of use, immediate feedback and support and word recognition can vary across schools.

The phenomenon examined by the study involves more than the presence of a program or institutional requirement. What respondents recognize as frequent, substantial or effective practice may depend on exposure, school size, resource access, the specific indicator being rated and the standards used for judging it. For this reason, evidence about AI-powered chatbots and Grade 7 vocabulary acquisition must distinguish the extent of reported implementation from measurable changes in teachers' work or learners' outcomes. The ratings obtained in a cross-sectional questionnaire provide one form of evidence, but not every form required for a causal claim. The study organized its inquiries around substantive domains rather than a single unqualified effectiveness score. The domains included accessibility and ease of use, frequency of use, interaction with chatbots, immediate feedback and support. Considering these areas separately makes it possible to identify strengths and weaknesses that may be hidden by a favorable overall mean. It also allows school leaders to direct follow-up observation and support toward specific activities instead of treating all facets of AI-powered chatbots and Grade 7 vocabulary acquisition as interchangeable.

A second analytical concern was whether respondent groups ranked the indicators similarly. Concordance across group rankings addresses the consistency of priorities or perceptions, not the magnitude of implementation or a relationship between an intervention and an educational outcome. Accordingly, the study treated means and ranks as descriptive indicators while interpreting the significance tests within the narrower meaning of Kendall's coefficient of concordance.

The study is also relevant to established educational guidance. The relevant work by Miao & Holmes, 2023 sets out a policy or research framework against which reported practices may be considered. Such guidance offers analytical context without transforming survey responses into direct observational evidence. The research questions therefore concerned the extent of practices in the specified setting, respondent-group agreement about their ordering, and implications for local educational improvement.

2. Methods

2.1 Design and setting

The study adopted a descriptive–evaluative–correlational survey of Grade 7 teachers using a researcher-made rating scale. It investigated AI-powered chatbots and Grade 7 vocabulary acquisition in public junior high schools in Calabanga Districts, Division of Camarines Sur during the study period. The design was observational. Although the term “correlational” appeared in the design description, the inferential results reported in the study primarily concerned agreement among ranked indicators; the resulting statistics did not estimate causal effects or the strength of a learner-level association.

2.2 Participants and sampling

The reported study participants consisted of 30 Grade 7 teachers. Respondents were grouped by the school classifications or school groupings shown in the results. Survey estimates should be read in light of the actual respondent composition rather than as

probability-sample estimates for all Philippine schools. Where the study did not provide distinct pupil assessments or repeated measures, teachers' accounts of student development were interpreted as perceptions.

2.3 Instrument and procedure

The study used a structured questionnaire or researcher-made rating scale encompassing the stated domains of AI-powered chatbots and Grade 7 vocabulary acquisition. Item-level statements were evaluated through the verbal-response categories and weighted-mean interpretation bands displayed with the study tables. The study described instrument preparation and administration through participating schools and used grouped ratings to address the research questions. Reported validity, reliability and response-handling details were not independently reconstructed or substituted where documentation was incomplete.

2.4 Data treatment and interpretation

The study reported frequency, percentage, ranks, weighted mean, Kendall's W and chi-square. Weighted means described central tendencies in ordinal rating responses, while reported ranks established the relative order of indicators. Kendall's W described concordance in ranks across the specified school groups; related chi-square results were treated according to the reported thresholds. No new significance calculations, sample-based confidence intervals or unreported multivariable estimates were introduced. Tables reproduce printed figures, including any discrepancies requiring verification.

2.5 Ethical and reporting considerations

The study was reported using aggregated responses. Individual-level datasets, consent documentation, and institutional ethics reference numbers were not reproduced in the available study chapters, and none is supplied here as though verified. Interpretation avoids identifying individual teachers or learners and refrains from attributing perceived outcomes to the surveyed practices without appropriate longitudinal or comparative evidence.

3. Results

The tables present the numerical entries and response-scale labels reported for each domain. Domain-level means describe respondent ratings; significance tests refer to group rank concordance.

Table 1. The Extent of the Utilization of the Artificial Intelligence (AI) Powered Chatbots Employed by the Grade 7 Teachers along Accessibility and Ease

Indicators	Big Schools			Medium Schools			Small Schools			Average		
	Wx	Int	R	Wx	Int	R	Wx	Int	R	Wx	Int	R
1. Accesses the AI-powered chatbot easily using available devices.	3.80	MU	4	3.67	MU	5	3.60	MU	6.5	3.69	MU	5
2. Opens the chatbot platform without technical difficulty.	3.70	MU	8	3.64	MU	7.5	3.63	MU	5	3.66	MU	7
3. Navigates the chatbot interface with ease.	3.83	MU	3	3.77	MU	3	3.64	MU	4	3.75	MU	3
4. Understands the instructions provided by the chatbot.	4.00	MU	2	3.88	MU	2	3.90	MU	2	3.93	MU	2
5. Enters questions or vocabulary tasks in the chatbot correctly.	3.73	MU	6.5	3.64	MU	7.5	3.66	MU	3	3.68	MU	6
6. Follows the chatbot prompts during vocabulary activities.	3.60	MU	10	3.51	MU	10	3.54	MU	10	3.55	MU	10
7. Receives clear responses from the chatbot.	3.67	MU	9	3.61	MU	9	3.58	MU	8	3.62	MU	9
8. Locates vocabulary learning features within the chatbot.	3.73	MU	6.5	3.66	MU	6	3.55	MU	9	3.65	MU	8
9. Operates the chatbot independently during learning activities.	3.77	MU	5	3.72	MU	4	3.60	MU	6.5	3.70	MU	4

Note. Numerical values, group categories, ranks, and verbal labels are reproduced as presented in the study; no recomputation or silent correction was applied.

The reported indicators in this table examine a specific dimension of AI-powered chatbots and Grade 7 vocabulary acquisition, including operational aspects associated with accessibility and ease of use. Their weighted means and respondent-group ranks describe how participants judged the prevalence, visibility or perceived effectiveness of the activity in their own schools.

The range of responses across individual statements should be interpreted as a pattern of relative strengths and areas for attention, not as direct evidence that any one activity produced an improvement. Rating scales capture perceptions that may also depend on school conditions, the opportunities respondents had to observe a practice and their interpretation of the response categories.

For practice, the findings point to the value of verifying the relevant activity through lesson plans, logs, classroom observations or other direct indicators. A focused review of accessibility and ease of use would help establish whether the reported ratings correspond with the consistency and quality of actual implementation.

Table 2. The Extent of the Utilization of the Artificial Intelligence (AI) Powered Chatbots Employed by the Grade 7 Teachers along Frequency of Use

Indicators	Big Schools			Medium Schools			Small Schools			Average		
	Wx	Int	R	Wx	Int	R	Wx	Int	R	Wx	Int	R
1. Uses the chatbot frequently while studying English vocabulary.	3.67	MU	2	3.61	MU	2	3.58	MU	2	3.62	MU	2
2. Practices vocabulary exercises using the chatbot on a regular basis.	3.17	U	8.5	3.10	U	8	3.00	U	9	3.09	U	8.5
3. Engages with the chatbot during classroom vocabulary activities.	2.57	U	10	2.66	U	10	2.50	U	10	2.58	U	10
4. Utilizes the chatbot when completing vocabulary learning tasks.	3.33	U	6.5	3.25	U	7	3.30	U	5	3.29	U	6
5. Consults the chatbot when encountering unfamiliar English words.	3.73	MU	1	3.68	MU	1	3.62	MU	1	3.68	MU	1
6. Reviews vocabulary lessons using the chatbot repeatedly.	3.60	MU	3	3.54	MU	3	3.49	MU	3	3.54	MU	3
7. Interacts with the chatbot throughout vocabulary learning sessions.	3.17	U	8.5	3.00	U	9	3.10	U	8	3.09	U	8.5

Note. Numerical values, group categories, ranks, and verbal labels are reproduced as presented in the study; no recomputation or silent correction was applied.

The reported indicators in this table examine a specific dimension of AI-powered chatbots and Grade 7 vocabulary acquisition, including operational aspects associated with frequency of use. Their weighted means and respondent-group ranks describe how participants judged the prevalence, visibility or perceived effectiveness of the activity in their own schools.

The range of responses across individual statements should be interpreted as a pattern of relative strengths and areas for attention, not as direct evidence that any one activity produced an improvement. Rating scales capture perceptions that may also depend on school conditions, the opportunities respondents had to observe a practice and their interpretation of the response categories.

For practice, the findings point to the value of verifying the relevant activity through lesson plans, logs, classroom observations or other direct indicators. A focused review of frequency of use would help establish whether the reported ratings correspond with the consistency and quality of actual implementation.

Table 3. The Extent of the Utilization of the Artificial Intelligence (AI) Powered Chatbots Employed by the Grade 7 Teachers along Interaction with the Ch

Indicators	Big Schools			Medium Schools			Small Schools			Average		
	Wx	Int	R	Wx	Int	R	Wx	Int	R	Wx	Int	R
1. Initiates conversations with the chatbot during vocabulary learning activities.	2.95	U	9	2.90	U	9	2.88	U	9	2.91	U	9
2. Asks questions about unfamiliar English words.	3.19	U	7	3.13	U	7	3.10	U	7.5	3.14	U	7
3. Responds to the prompts given by the chatbot.	3.41	U	5	3.33	U	5	3.37	U	4	3.37	U	5
4. Clarifies the meanings of vocabulary words through chatbot interaction.	3.14	U	8	3.00	U	8	3.10	U	7.5	3.08	U	8
5. Requests examples of words used in sentences.	2.90	U	10	2.77	U	10	2.84	U	10	2.84	U	10
6. Explores vocabulary topics through continuous interaction with the chatbot.	3.56	MU	3	3.60	MU	3	3.54	MU	3	3.57	MU	3
7. Follows the chatbot's instructions during vocabulary exercises.	3.50	MU	4	3.46	U	4	3.25	U	5	3.40	U	4

Note. Numerical values, group categories, ranks, and verbal labels are reproduced as presented in the study; no recomputation or silent correction was applied.

The reported indicators in this table examine a specific dimension of AI-powered chatbots and Grade 7 vocabulary acquisition, including operational aspects associated with interaction with chatbots. Their weighted means and respondent-group ranks describe how participants judged the prevalence, visibility or perceived effectiveness of the activity in their own schools.

The range of responses across individual statements should be interpreted as a pattern of relative strengths and areas for attention, not as direct evidence that any one activity produced an improvement. Rating scales capture perceptions that may also depend on school conditions, the opportunities respondents had to observe a practice and their interpretation of the response categories.

For practice, the findings point to the value of verifying the relevant activity through lesson plans, logs, classroom observations or other direct indicators. A focused review of interaction with chatbots would help establish whether the reported ratings correspond with the consistency and quality of actual implementation.

Table 4. The Extent of the Utilization of the Artificial Intelligence (AI) Powered Chatbots Employed by the Grade 7 Teachers along Immediate Feedback and

Indicators	Big Schools			Medium Schools			Small Schools			Average		
	Wx	Int	R	Wx	Int	R	Wx	Int	R	Wx	Int	R
1. Receives Immediate responses to vocabulary questions.	3.52	MU	8	3.56	MU	4	3.44	U	8	3.51	MU	7
2. Obtains instant explanations of unfamiliar words.	3.70	MU	1.5	3.63	MU	1	3.57	MU	1	3.63	MU	1
3. Checks the correctness of vocabulary answers through chatbot feedback.	3.57	MU	4.5	3.51	MU	6.5	3.48	U	6	3.52	MU	5
4. Identifies errors in vocabulary usage through chatbot responses.	3.33	U	11	3.29	U	11	3.25	U	11	3.29	U	11
5. Improves vocabulary answers based on the chatbot's suggestions.	3.57	MU	4.5	3.54	MU	5	3.49	U	4.5	3.53	MU	3
6. Understands correct word meanings through instant feedback.	3.57	MU	4.5	3.50	MU	8	3.49	U	4.5	3.52	MU	5
7. Corrects mistakes in spelling with the help of the chatbot.	3.50	MU	9	3.40	U	10	3.38	U	10	3.43	U	10
8. Revises vocabulary responses after receiving chatbot feedback.	3.53	MU	7	3.43	U	9	3.50	MU	2.5	3.49	U	8.5

Note. Numerical values, group categories, ranks, and verbal labels are reproduced as presented in the study; no recomputation or silent correction was applied.

The reported indicators in this table examine a specific dimension of AI-powered chatbots and Grade 7 vocabulary acquisition, including operational aspects associated with immediate feedback and support. Their weighted means and respondent-group ranks describe how participants judged the prevalence, visibility or perceived effectiveness of the activity in their own schools.

The range of responses across individual statements should be interpreted as a pattern of relative strengths and areas for attention, not as direct evidence that any one activity produced an improvement. Rating scales capture perceptions that may also depend on school conditions, the opportunities respondents had to observe a practice and their interpretation of the response categories.

For practice, the findings point to the value of verifying the relevant activity through lesson plans, logs, classroom observations or other direct indicators. A focused review of immediate feedback and support would help establish whether the reported ratings correspond with the consistency and quality of actual implementation.

Table 5. The Extent of the Utilization of the Artificial Intelligence (AI) Powered Chatbots Employed by the Grade 7 Teachers along Learner Engagement

Indicators	Big Schools			Medium Schools			Small Schools			Average		
	Wx	Int	R	Wx	Int	R	Wx	Int	R	Wx	Int	R
1. Participates actively in vocabulary learning activities using the chatbot.	3.19	U	3	3.13	U	4	3.00	U	7	3.11	U	5
2. Focuses on vocabulary tasks while interacting with the chatbot.	3.11	U	5	3.04	U	6	3.16	U	5	3.10	U	6
3. Shows interest in learning new vocabulary through the chatbot.	3.17	U	4	3.10	U	5	3.21	U	2	3.16	U	4
4. Enjoys vocabulary practice while using the chatbot.	3.00	U	10	2.94	U	9	3.05	U	6	3.00	U	7.5
5. Explores new words through chatbot-guided activities.	3.00	U	10	2.94	U	9	2.80	U	10	2.91	U	10
6. Remains attentive during chatbot-assisted learning sessions.	3.09	U	7	3.00	U	7	2.91	U	8	3.00	U	7.5
7. Interacts willingly with the chatbot during vocabulary exercises.	3.10	U	6	3.14	U	3	3.29	U	1	3.18	U	3
8. Motivates themselves to learn more vocabulary through the chatbot.	3.00	U	10	2.88	U	11	2.79	U	11	2.89	U	11

Note. Numerical values, group categories, ranks, and verbal labels are reproduced as presented in the study; no recomputation or silent correction was applied.

The reported indicators in this table examine a specific dimension of AI-powered chatbots and Grade 7 vocabulary acquisition, including operational aspects associated with learner engagement. Their weighted means and respondent-group ranks describe how participants judged the prevalence, visibility or perceived effectiveness of the activity in their own schools.

The range of responses across individual statements should be interpreted as a pattern of relative strengths and areas for attention, not as direct evidence that any one activity produced an improvement. Rating scales capture perceptions that may also depend on school conditions, the opportunities respondents had to observe a practice and their interpretation of the response categories.

For practice, the findings point to the value of verifying the relevant activity through lesson plans, logs, classroom observations or other direct indicators. A focused review of learner engagement would help establish whether the reported ratings correspond with the consistency and quality of actual implementation.

Table 6. The Summary of the Extent of the Utilization of the Artificial Intelligence (AI) Powered Chatbots Employed by the Grade 7 Teachers

Indicators	Big Schools			Medium Schools			Small Schools			Average		
	Wx	Int	R	Wx	Int	R	Wx	Int	R	Wx	Int	R
1. Accessibility and Ease of Use	3.74	MU	1	3.65	MU	1	3.61	MU	1	3.67	MU	1
2. Frequency of Use	3.34	U	4	3.28	U	4	3.24	U	4	3.29	U	4
3. Interaction with the Chatbot	3.35	U	2.5	3.33	U	3	3.31	U	3	3.33	U	3
4. Immediate Feedback and Support	3.55	MU	2.5	3.50	MU	2	3.45	U	2	3.50	MU	3
5. Learner Engagement	3.11	U	5	3.05	U	5	3.04	U	5	3.07	U	5
Average	3.41	U		3.36	U		3.33	U		3.37	U	

Legend:

4.50 - 5.00	Very Much Utilized (VMU)	Wx	- Weighted Mean
3.50 - 4.49	Much Utilized (MU)	Int	- Interpretation
2.50 - 3.49	Utilized (U)	R	- Rank
1.50 - 2.49	Fairly Utilized (FU)		
1.00 - 1.49	Not at All (NA)		

Note. Numerical values, group categories, ranks, and verbal labels are reproduced as presented in the study; no recomputation or silent correction was applied.

The summary consolidates the domain-level ratings and makes visible the relative standing of major components of AI-powered chatbots and Grade 7 vocabulary acquisition. The relative order of these reported domains is more informative for local prioritization than the overall mean alone, because the domains may involve distinct resources and implementation processes.

An apparently favorable average does not mean every school implemented each activity equally or that learners experienced the same educational benefits. The reported categories reflect a survey scale; they should not be read as externally verified proficiency standards or comparable units of learning gain.

The practical use of this finding is diagnostic. Components with lower ratings, including potential constraints affecting vocabulary comprehension, may warrant closer monitoring, better implementation documentation and targeted professional support before a district-level improvement claim is made.

Table 7. The Test of Significant Agreement on the Rank Orders of the Extent of the Utilization of the Artificial Intelligence (AI) Powered Chatbots Employ

Indication	Aspects				
	Accessibility and Ease of Use	Frequency of Use	Interaction with the Chatbot	Immediate Feedback and Support	Learner Engagement
Summation of the Squared Deviation from the mean	1195	726.50	736	738.50	844
No. of Groups	3	3	3	3	3
No. of Cases or Items	12	10	10	11	11
Coefficient of Concordance W	0.93	0.98	0.99	0.75	0.85
Computed chi-square	30.69	26.46	26.73	22.50	25.50
Degree of Freedom	11	9	9	10	10
Tabular χ^2					
0.05	19.68	16.92	16.92	18.31	18.31
0.025	22.62	19.68	19.68	21.16	21.16
0.01	24.72	21.67	21.67	23.21	23.21
0.005	26.76	23.59	23.59	25.19	25.19
0.001	31.26	27.88	27.88	29.59	29.59
Decision on the Alternative Hypothesis (H ₁)	Accepted	Accepted	Accepted	Accepted	Accepted
Significance of Agreement	0.005	0.005	0.005	0.025	0.005

Note. Numerical values, group categories, ranks, and verbal labels are reproduced as presented in the study; no recomputation or silent correction was applied.

This table addresses the study question concerning agreement in the ordering of the reported indicators. The statistical decision must be understood alongside the coefficient, chi-square value, number of ranked cases and stated significance level; it does not evaluate how strongly AI-powered chatbots and Grade 7 vocabulary acquisition changed in practice.

Concordance among school-group rank profiles is analytically different from agreement among all individual respondents. Particularly when some groups are smaller, highly similar aggregate rankings can coexist with unreported within-group variation. Conversely, a nonsignificant concordance result does not establish the absence of an educational practice.

The immediate implication is to examine whether the observed group differences reflect school circumstances, access to support or indicator interpretation. Follow-up work on vocabulary retention could use school-level records, interviews, observation protocols or direct learner measures, depending on the outcome being investigated.

Table 8. The Extent of the Effectiveness of the Utilization of the Artificial Intelligence (AI) Powered Chatbots in Enhancing the Vocabulary Acquisition o

Indicators	Big Schools			Medium Schools			Small Schools			Average		
	Wx	Int	R	Wx	Int	R	Wx	Int	R	Wx	Int	R
1. Identifies the meanings of unfamiliar vocabulary words.	3.93	ME	7.5	3.88	ME	5.5	3.76	ME	10.5	3.86	ME	8
2. Understands the definitions of new English words.	4.00	ME	2.5	3.93	ME	3	3.84	ME	5	3.92	ME	3
3. Recognizes vocabulary words encountered in reading passages.	4.00	ME	2.5	3.96	ME	2	3.90	ME	1.5	3.95	ME	2
4. Interprets the meaning of words based on chatbot explanations.	3.90	ME	10	3.87	ME	7	3.80	ME	7.5	3.86	ME	8
5. Differentiates the meanings of similar or confusing words.	3.93	ME	7.5	3.81	ME	10	3.76	ME	10.5	3.83	ME	10
6. Associates new vocabulary words with their correct meanings.	3.93	ME	7.5	3.85	ME	8	3.80	ME	7.5	3.86	ME	8
7. Explains the meaning of vocabulary words after interacting with the chatbot.	3.97	ME	4.5	3.91	ME	4	3.86	ME	3.5	3.91	ME	4
8. Relates newly learned words to previously learned vocabulary.	3.93	ME	7.5	3.88	ME	5.5	3.83	ME	6	3.88	ME	5.5
9. Analyzes word meanings based on examples provided by the chatbot.	3.83	ME	11	3.79	ME	11	3.77	ME	9	3.80	ME	11
10. Clarifies the meaning of difficult words through chatbot assistance.	4.03	ME	1	4.00	ME	1	3.90	ME	1.5	3.98	ME	1

Note. Numerical values, group categories, ranks, and verbal labels are reproduced as presented in the study; no recomputation or silent correction was applied.

The reported indicators in this table examine a specific dimension of AI-powered chatbots and Grade 7 vocabulary acquisition, including operational aspects associated with vocabulary usage in context. Their weighted means and respondent-group ranks describe how participants judged the prevalence, visibility or perceived effectiveness of the activity in their own schools.

The range of responses across individual statements should be interpreted as a pattern of relative strengths and areas for attention, not as direct evidence that any one activity produced an improvement. Rating scales capture perceptions that may also depend on school conditions, the opportunities respondents had to observe a practice and their interpretation of the response categories.

For practice, the findings point to the value of verifying the relevant activity through lesson plans, logs, classroom observations or other direct indicators. A focused review of vocabulary usage in context would help establish whether the reported ratings correspond with the consistency and quality of actual implementation.

Table 9. The Extent of the Effectiveness of the Utilization of the Artificial Intelligence (AI) Powered Chatbots in Enhancing the Vocabulary Acquisition o

Indicators	Big Schools			Medium Schools			Small Schools			Average		
	Wx	Int	R	Wx	Int	R	Wx	Int	R	Wx	Int	R
1. Recalls the meanings of previously learned vocabulary words.	3.70	ME	3	3.66	ME	3.5	3.68	ME	3	3.68	ME	3
2. Remembers new vocabulary words after interacting with the chatbot.	3.77	ME	1.5	3.73	ME	1	3.69	ME	1.5	3.73	ME	1
3. Retains the meanings of words learned through chatbot activities.	3.50	ME	10	3.54	ME	10	3.46	ME	10	3.50	ME	10
4. Recognizes previously learned vocabulary during later lessons.	3.63	ME	7	3.61	ME	6	3.58	ME	5	3.61	ME	6
5. Retrieves vocabulary meanings when needed in learning tasks.	3.67	ME	5	3.60	ME	7	3.54	ME	7	3.60	ME	7
6. Reinforces memory of vocabulary through repeated chatbot practice.	3.67	ME	5	3.63	ME	5	3.61	ME	4	3.64	ME	4
7. Maintains understanding of vocabulary words over time.	3.67	ME	5	3.66	ME	3.5	3.55	ME	6	3.63	ME	5
8. Strengthens memory of vocabulary words through chatbot exercises.	3.77	ME	1.5	3.70	ME	2	3.69	ME	1.5	3.72	ME	2
9. Stores newly learned vocabulary in long-term memory.	3.57	ME	9	3.59	ME	8	3.51	ME	9	3.56	ME	9

Note. Numerical values, group categories, ranks, and verbal labels are reproduced as presented in the study; no recomputation or silent correction was applied.

The reported indicators in this table examine a specific dimension of AI-powered chatbots and Grade 7 vocabulary acquisition, including operational aspects associated with spelling accuracy. Their weighted means and respondent-group ranks describe how participants judged the prevalence, visibility or perceived effectiveness of the activity in their own schools.

The range of responses across individual statements should be interpreted as a pattern of relative strengths and areas for attention, not as direct evidence that any one activity produced an improvement. Rating scales capture perceptions that may also depend on school conditions, the opportunities respondents had to observe a practice and their interpretation of the response categories. For practice, the findings point to the value of verifying the relevant activity through lesson plans, logs, classroom observations or other direct indicators. A focused review of spelling accuracy would help establish whether the reported ratings correspond with the consistency and quality of actual implementation.

Table 10. The Extent of the Effectiveness of the Utilization of the Artificial Intelligence (AI) Powered Chatbots in Enhancing the Vocabulary Acquisition o

Indicators	Big Schools			Medium Schools			Small Schools			Average		
	Wx	Int	R	Wx	Int	R	Wx	Int	R	Wx	Int	R
1. Uses new vocabulary words correctly in sentences.	3.99	ME	2	3.90	ME	2	3.93	ME	2	3.94	ME	2
2. Applies learned words in meaningful contexts.	3.64	ME	9	3.56	ME	8	3.61	ME	8	3.60	ME	8
3. Incorporates vocabulary words into written exercises.	3.50	ME	11	3.45	E	11	3.58	ME	9	3.51	ME	10
4. Constructs sentences using newly learned words.	3.92	ME	3	3.83	ME	4.5	3.80	ME	3.5	3.85	ME	3
5. Demonstrates understanding of word meanings through context.	3.66	ME	8	3.47	E	10	3.38	E	11	3.50	ME	11
6. Integrates vocabulary words in classroom discussions.	3.91	ME	4	3.85	ME	3	3.70	ME	5	3.82	ME	5
7. Formulates responses using appropriate vocabulary.	3.61	ME	10	3.50	ME	9	3.55	ME	10	3.55	ME	9
8. Expresses ideas clearly using learned words.	3.70	ME	7	3.67	ME	6	3.62	ME	6.5	3.66	ME	6
9. Adapts vocabulary to different situations or topics.	3.75	ME	6	3.58	ME	7	3.62	ME	6.5	3.65	ME	7

Note. Numerical values, group categories, ranks, and verbal labels are reproduced as presented in the study; no recomputation or silent correction was applied.

The reported indicators in this table examine a specific dimension of AI-powered chatbots and Grade 7 vocabulary acquisition, including operational aspects associated with word recognition. Their weighted means and respondent-group ranks describe how participants judged the prevalence, visibility or perceived effectiveness of the activity in their own schools.

The range of responses across individual statements should be interpreted as a pattern of relative strengths and areas for attention, not as direct evidence that any one activity produced an improvement. Rating scales capture perceptions that may also depend on school conditions, the opportunities respondents had to observe a practice and their interpretation of the response categories.

For practice, the findings point to the value of verifying the relevant activity through lesson plans, logs, classroom observations or other direct indicators. A focused review of word recognition would help establish whether the reported ratings correspond with the consistency and quality of actual implementation.

Table 11. The Extent of the Effectiveness of the Utilization of the Artificial Intelligence (AI) Powered Chatbots in Enhancing the Vocabulary Acquisition o

Indicators	Big Schools			Medium Schools			Small Schools			Average		
	Wx	Int	R	Wx	Int	R	Wx	Int	R	Wx	Int	R
1. Spells vocabulary words correctly after chatbot practice.	3.81	ME	7	3.69	ME	9	3.63	ME	9	3.71	MU	9
2. Writes words accurately in exercises provided by the chatbot.	3.85	ME	4	3.77	ME	5	3.71	ME	3	3.70	MU	4
3. Checks the spelling of unfamiliar words using the chatbot	3.80	ME	8	3.74	ME	7	3.66	ME	7	3.73	MU	7
4. Corrects misspelled words based on chatbot feedback.	3.61	ME	12	3.50	ME	12	3.44	E	12	3.52	MU	12
5. Recognizes correct spelling patterns through chatbot activities.	3.66	ME	11	3.58	ME	11	3.60	ME	10	3.61	MU	11
6. Memorizes the correct spelling of newly learned words.	3.95	ME	1	3.87	ME	2	3.70	ME	5	3.84	MU	2
7. Applies proper spelling in written tasks and exercises.	3.85	ME	4	3.75	ME	6	3.80	ME	1.5	3.80	MU	3
8. Distinguishes between similarly spelled words.	3.70	ME	9.5	3.62	ME	10	3.56	ME	11	3.63	MU	10
9. Practices spelling repeatedly using chatbot exercises.	3.85	ME	4	3.73	ME	8	3.70	ME	5	3.76	MU	6
10. Identifies errors in spelling through chatbot guidance.	3.70	ME	9.5	3.81	ME	3	3.64	ME	8	3.72	MU	8

Note. Numerical values, group categories, ranks, and verbal labels are reproduced as presented in the study; no recomputation or silent correction was applied.

The reported indicators in this table examine a specific dimension of AI-powered chatbots and Grade 7 vocabulary acquisition, including operational aspects associated with accessibility and ease of use. Their weighted means and respondent-group ranks describe how participants judged the prevalence, visibility or perceived effectiveness of the activity in their own schools.

The range of responses across individual statements should be interpreted as a pattern of relative strengths and areas for attention, not as direct evidence that any one activity produced an improvement. Rating scales capture perceptions that may also depend on school conditions, the opportunities respondents had to observe a practice and their interpretation of the response categories.

For practice, the findings point to the value of verifying the relevant activity through lesson plans, logs, classroom observations or other direct indicators. A focused review of accessibility and ease of use would help establish whether the reported ratings correspond with the consistency and quality of actual implementation.

Table 12. The Extent of the Effectiveness of the Utilization of the Artificial Intelligence (AI) Powered Chatbots in Enhancing the Vocabulary Acquisition o

Indicators	Big Schools			Medium Schools			Small Schools			Average		
	Wx	Int	R	Wx	Int	R	Wx	Int	R	Wx	Int	R
1. Identifies vocabulary words when they encounter them in reading passages.	3.57	ME	9	3.40	E	10	3.44	E	9	3.47	E	9
2. Recognizes new words learned through the chatbot.	3.63	ME	4.5	3.52	ME	8	3.47	E	8	3.54	ME	8
3. Distinguishes familiar words from unfamiliar words.	3.47	E	10	3.44	E	9	3.40	E	10	3.44	E	10
4. Matches words with their correct meanings.	3.67	ME	2.5	3.59	ME	3	3.52	ME	7	3.59	ME	3.5
5. Selects the correct word in multiple-choice exercises.	3.70	ME	1	3.66	ME	1	3.63	ME	1	3.66	ME	1
6. Locates target vocabulary words in sentences or paragraphs.	3.60	ME	7	3.53	ME	6.5	3.57	ME	3.5	3.57	ME	6
7. Detects spelling differences between similar words.	3.60	ME	7	3.57	ME	5	3.57	ME	3.5	3.58	ME	5
8. Recalls previously learned words during reading activities.	3.67	ME	2.5	3.61	ME	2	3.59	ME	2	3.62	ME	2

Note. Numerical values, group categories, ranks, and verbal labels are reproduced as presented in the study; no recomputation or silent correction was applied.

The reported indicators in this table examine a specific dimension of AI-powered chatbots and Grade 7 vocabulary acquisition, including operational aspects associated with frequency of use. Their weighted means and respondent-group ranks describe how participants judged the prevalence, visibility or perceived effectiveness of the activity in their own schools.

The range of responses across individual statements should be interpreted as a pattern of relative strengths and areas for attention, not as direct evidence that any one activity produced an improvement. Rating scales capture perceptions that may also depend on school conditions, the opportunities respondents had to observe a practice and their interpretation of the response categories.

For practice, the findings point to the value of verifying the relevant activity through lesson plans, logs, classroom observations or other direct indicators. A focused review of frequency of use would help establish whether the reported ratings correspond with the consistency and quality of actual implementation.

Table 13. The Summary of the Extent of the Effectiveness of the Utilization of the Artificial Intelligence (AI) Powered Chatbots in Enhancing the Vocabulary

Indicators	Big Schools			Medium Schools			Small Schools			Average		
	Wx	Int	R	Wx	Int	R	Wx	Int	R	Wx	Int	R
1. Vocabulary Comprehension	3.95	ME	1	3.88	ME	1	3.83	ME	1	3.88	ME	1
2. Vocabulary Retention	3.66	ME	4	3.63	ME	4	3.58	ME	4	3.62	ME	4
3. Vocabulary Usage in Context	3.79	ME	2.5	3.70	ME	3	3.69	ME	2	3.83	ME	2
4. Spelling Accuracy	3.79	ME	2.5	3.73	ME	2	3.66	ME	3	3.73	ME	3
5. Word Recognition	3.61	ME	5	3.54	ME	5	3.53	ME	5	3.56	ME	5
Average	3.76	ME		3.70	ME		3.66	ME		3.72	ME	

Legend:

4.50 - 5.00	Very Much Effective (VME)	Wx	- Weighted Mean
3.50 - 4.49	Much Effective (ME)	Int	- Interpretation
2.50 - 3.49	Effective (E)	R	- Rank
1.50 - 2.49	Fairly Effective (FE)		
1.00 - 1.49	Not at All (NA)		

Note. Numerical values, group categories, ranks, and verbal labels are reproduced as presented in the study; no recomputation or silent correction was applied.

The summary consolidates the domain-level ratings and makes visible the relative standing of major components of AI-powered chatbots and Grade 7 vocabulary acquisition. The relative order of these reported domains is more informative for local prioritization than the overall mean alone, because the domains may involve distinct resources and implementation processes.

An apparently favorable average does not mean every school implemented each activity equally or that learners experienced the same educational benefits. The reported categories reflect a survey scale; they should not be read as externally verified proficiency standards or comparable units of learning gain.

The practical use of this finding is diagnostic. Components with lower ratings, including potential constraints affecting interaction with chatbots, may warrant closer monitoring, better implementation documentation and targeted professional support before a district-level improvement claim is made.

Table 14. The Test of Significant Agreement on the Rank Orders of the Extent of the Effectiveness of the Utilization of the Artificial Intelligence (AI) Po

Indication	Aspects				
	Vocabulary Comprehension	Vocabulary Retention	Vocabulary Usage in Context	Spelling Accuracy	Word Recognition
Summation of the Squared Deviation from the mean	815	652	943.50	1085	616.50
No. of Groups	3	3	3	3	3
No. of Cases or Items	11	10	11	12	10
Coefficient of Concordance W	0.82	0.88	0.95	0.84	0.83
Computed chi-square	24.60	23.76	28.50	27.72	22.41
Degree of Freedom	10	9	10	11	9
Tabular χ^2					
0.05	18.31	16.92	18.31	19.68	16.92
0.025	21.16	19.60	21.16	22.62	19.60
0.01	23.21	21.67	23.21	24.72	21.67
0.005	25.19	23.59	25.19	26.76	23.59
0.001	29.59	27.88	29.59	31.26	27.88
Decision on the Alternative Hypothesis (H_1)	Accepted	Accepted	Accepted	Accepted	Accepted
Significance of Agreement	0.01	0.005	0.005	0.005	0.01

Note. Numerical values, group categories, ranks, and verbal labels are reproduced as presented in the study; no recomputation or silent correction was applied.

This table addresses the study question concerning agreement in the ordering of the reported indicators. The statistical decision must be understood alongside the coefficient, chi-square value, number of ranked cases and stated significance level; it does not evaluate how strongly AI-powered chatbots and Grade 7 vocabulary acquisition changed in practice.

Concordance among school-group rank profiles is analytically different from agreement among all individual respondents. Particularly when some groups are smaller, highly similar aggregate rankings can coexist with unreported within-group variation. Conversely, a nonsignificant concordance result does not establish the absence of an educational practice.

The immediate implication is to examine whether the observed group differences reflect school circumstances, access to support or indicator interpretation. Follow-up work on immediate feedback and support could use school-level records, interviews, observation protocols or direct learner measures, depending on the outcome being investigated.

4. Discussion

Taken together, the findings suggest a differentiated profile of AI-powered chatbots and Grade 7 vocabulary acquisition within public junior high schools in Calabanga Districts, Division of Camarines Sur. Some domains were rated more favorably than others, but the study's comparisons concerned respondents' evaluations of program activity or practice. The interpretation should therefore emphasize the organization and visibility of those practices rather than unobserved changes in educational outcomes.

The domains must also be considered conceptually. For example, accessibility and ease of use and word recognition involve different operational demands even when they appear in the same questionnaire. An overall rating cannot establish that schools have equivalent capacity, resources or instructional conditions. Follow-up should identify the precise activities behind each item and test whether respondents applied a shared meaning to the measurement scale.

The study's rank-concordance analysis makes an additional contribution by examining alignment in the order assigned to indicators. Kendall's W is sensitive to patterns of ranking among the selected groups; it does not assess causation, pupil-level correlations or improvements over time. A statistically significant result should therefore be explained as concordance of ordered profiles, while nonsignificant outcomes should not be called evidence that a practice was ineffective.

The cited scholarship and policy provide context but should not be treated as validating the local mean scores. Guidance from Miao & Holmes, 2023, Department of Education, 2017 highlights the importance of clearly specified practices, appropriate pedagogical support or relevant educational standards. Applying such guidance requires attention to local implementation, documentation and direct assessment rather than reliance on favorable response categories alone.

Implications for school administrators include establishing observable criteria for high-priority activities, checking participation and resource access across schools, and using professional conversations to determine why some dimensions received lower assessments.

If the objective is to understand effects on actual teacher or learner performance, a subsequent design should add independent outcome data, repeated observations and an appropriate comparison framework.

The main limitations arise from the local scope of the respondents, reliance on reported ratings, the unequal or limited respondent groups in some analyses, and incomplete access to item-level observations for independent verification. In addition, the underlying survey cannot by itself separate the contribution of word recognition from other school, family, learner or policy influences. The conclusions therefore remain descriptive and context-specific; extension beyond the study setting requires additional evidence.

5. Conclusion

The study documented the reported status of AI-powered chatbots and Grade 7 vocabulary acquisition in public junior high schools in Calabanga Districts, Division of Camarines Sur. Its domain-level results and agreement tests offer an account of local respondent perceptions, areas of comparatively stronger implementation and aspects warranting closer attention. Reported means and statistical decisions were retained as presented, without reinterpretation as experimental effects.

The findings support evidence-informed review of accessibility and ease of use, vocabulary comprehension and word recognition. Implementation monitoring, clearer criteria for rating performance, and direct observational or learner-level evidence are suitable priorities for further research. The study does not demonstrate that an individual practice caused an improvement in school, teacher or learner outcomes.

References

- [1] Miao, F., & Holmes, W. (2023). Guidance for generative AI in education and research. UNESCO. <https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research>
- [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/>