

Appendix 1. User Experience Questionnaire (UEQ) – Original English Version

No.	Adjective Pair (English)	Dimension
1	Annoying – Enjoyable	Attractiveness
2	Not understandable – Understandable	Perspiciuity
3	Difficult to learn – Easy to learn	Perspiciuity
4	Inefficient – Efficient	Efficiency
5	Obstructive – Supportive	Dependability
6	Unpredictable – Predictable	Dependability
7	Dull – Exciting	Stimulation
8	Boring – Interesting	Stimulation
9	Conventional – Inventive	Novelty
10	Usual – Leading-edge	Novelty

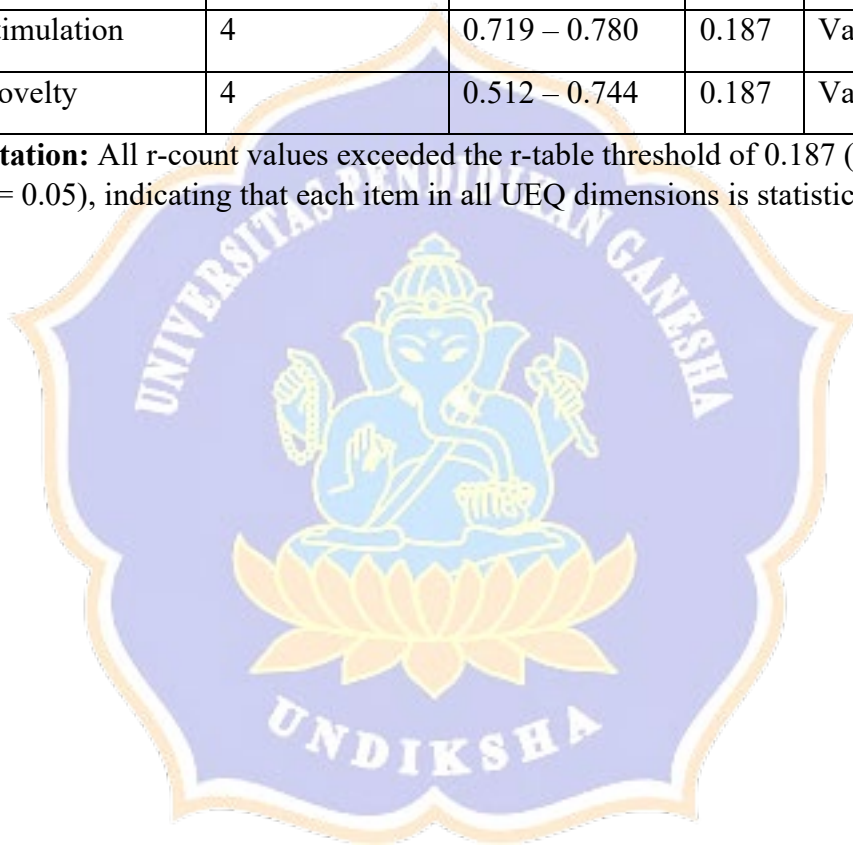
Scale: 7-point bipolar Likert scale

1 = fully agree with left adjective, 7 = fully agree with right adjective

Appendix 2. Instrument Validity Test Results

No.	UEQ Dimension	Number of Items	Range of r-count	r-table	Conclusion
1	Attractiveness	6	0.705 – 0.819	0.187	Valid
2	Perspiciuity	4	0.720 – 0.823	0.187	Valid
3	Efficiency	4	0.715 – 0.792	0.187	Valid
4	Dependability	4	0.583 – 0.743	0.187	Valid
5	Stimulation	4	0.719 – 0.780	0.187	Valid
6	Novelty	4	0.512 – 0.744	0.187	Valid

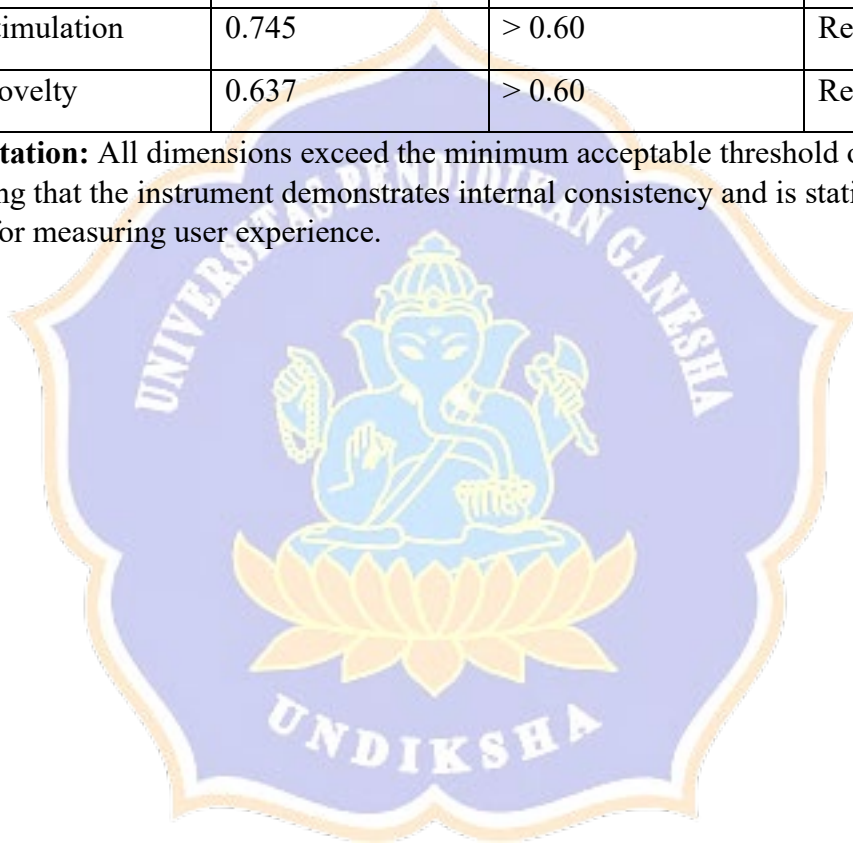
Interpretation: All r-count values exceeded the r-table threshold of 0.187 (for $n = 108$ at $\alpha = 0.05$), indicating that each item in all UEQ dimensions is statistically valid.



Appendix 3. Instrument Reliability Test Results

No.	UEQ Dimension	Cronbach's Alpha	Reliability Threshold	Conclusion
1	Attractiveness	0.857	> 0.60	Reliable
2	Perspiciuity	0.785	> 0.60	Reliable
3	Efficiency	0.753	> 0.60	Reliable
4	Dependability	0.688	> 0.60	Reliable
5	Stimulation	0.745	> 0.60	Reliable
6	Novelty	0.637	> 0.60	Reliable

Interpretation: All dimensions exceed the minimum acceptable threshold of 0.60, confirming that the instrument demonstrates internal consistency and is statistically reliable for measuring user experience.



Appendix 4. Descriptive Statistical Results of Each UEQ Dimension

The following table presents the average (mean) scores obtained for each of the six UEQ dimensions based on students' responses after using ChatGPT for English learning. These values offer insight into how the system was perceived across various aspects of user experience.

No.	Dimension	Mean Score	Interpretation
1	Attractiveness	1.52	Positive
2	Perspiciuity	1.47	Positive
3	Efficiency	1.38	Positive
4	Dependability	1.14	Positive
5	Stimulation	1.35	Positive
6	Novelty	1.11	Positive

These results indicate that students consistently rated their experience with ChatGPT as favourable across all six dimensions. The strongest impressions were found in “Attractiveness” and “Perspiciuity,” which reflect that students found the platform enjoyable and easy to understand. Meanwhile, dimensions like “Efficiency” and “Stimulation” showed that ChatGPT helped students complete learning tasks more quickly and remained engaging throughout their usage.

The mean score for each dimension was computed by averaging all responses on a 7-point Likert scale. The formula used is:

$$Mean = \frac{\sum X}{n}$$

Where:

- $\sum X$ is the total score given by all respondents for each dimension
- n is the total number of valid responses (in this case, 108)

Mean Range	Interpretation
> 0.8	Positive
-0.8 to 0.8	Neutral
< -0.8	Negative

Based on this scale, all six UEQ dimensions achieved scores well above 0.8, meaning students' experiences with ChatGPT were not only functional but also

emotionally engaging and enjoyable. These outcomes support the platform's relevance and effectiveness in higher education language learning settings.



Appendix 5. UEQ Benchmark Comparison for ChatGPT

To contextualize the user experience scores obtained in this study, each UEQ dimension was compared against the official UEQ benchmark dataset. This benchmark is derived from evaluations of more than 250 digital products and services worldwide, including software tools, web applications, and AI-based platforms.

The comparison allows us to assess how students' experience using ChatGPT for English learning ranks relative to other systems.

Dimension	Mean Score	Benchmark Category	Interpretation
Attractiveness	1.52	Excellent	Highly engaging and positively received overall
Perspicuity	1.47	Good	Easy to understand and learn
Efficiency	1.38	Good	Efficient in supporting learning tasks
Dependability	1.14	Above Average	Reliable and predictable
Stimulation	1.35	Good	Emotionally engaging and motivating
Novelty	1.11	Above Average	Perceived as innovative and refreshing

Interpretation:

1. Attractiveness received the highest possible benchmark category: Excellent, suggesting that students were strongly drawn to and emotionally satisfied with the platform.
2. Perspicuity, Efficiency, and Stimulation all scored in the Good category, which reflects consistent usability, clarity, and motivational value.
3. Dependability and Novelty were categorized as Above Average, indicating that while the system was generally trusted and perceived as creative, these aspects were slightly less prominent compared to others.

These benchmark results confirm that ChatGPT offers a user experience that is not only functional and efficient but also enjoyable and emotionally meaningful—a crucial combination for sustaining motivation in language learning.

Appendix 6. Respondents' Academic Semester Profile

To better understand the diversity of perspectives represented in this study, participants were categorized based on their current academic semester. This distribution is relevant because students at different stages of their academic journey may engage with ChatGPT differently—whether for foundational tasks or more complex assignments.

Semester	Number of Respondents	Percentage (%)
2nd Semester	14	12.96%
4th Semester	37	34.26%
6th Semester	19	17.59%
8th Semester	38	35.19%
Total	108	100%

Interpretation:

1. The largest groups of respondents came from the 8th semester (35.19%) and 4th semester (34.26%), indicating strong representation from both senior and mid-level students.
2. Students in the 2nd semester (12.96%) represent those with limited university experience but already familiar with digital tools like ChatGPT.
3. This distribution ensures that the data reflects varied usage habits and expectations, from entry-level users to those applying ChatGPT to thesis writing or final projects.

The range of academic levels included in the sample enhances the credibility and generalizability of the findings, particularly regarding how ChatGPT is used across different stages of English language education.

Appendix 7. Frequency and Purpose of ChatGPT Usage

To examine students' interaction patterns with ChatGPT, this study collected information on how frequently they used the tool and the main purposes behind its usage. The data are summarized below:

A. Frequency of Using ChatGPT

Response Category	Number of Respondents	Percentage (%)
Almost every day	21	19.44%
A few times per week	46	42.59%
Occasionally	33	30.56%
Tried it once or twice only	8	7.41%
Total	108	100%

Interpretation:

Most students reported using ChatGPT either *a few times per week* (42.59%) or *occasionally* (30.56%), while a notable proportion (19.44%) used it almost daily. This shows that ChatGPT has become a regular tool in their academic routine.

B. Main Purposes for Using ChatGPT

(Multiple answers were allowed)

Purpose	Number of Respondents	Percentage (%)
Essay writing support	84	77.78%
Grammar checking	76	70.37%
Vocabulary learning	63	58.33%
Reading comprehension	41	37.96%
Conversation practice	38	35.19%
Brainstorming ideas	55	50.93%

Interpretation:

The most frequent reason students used ChatGPT was for *essay writing support* (77.78%) and *grammar checking* (70.37%), highlighting the tool's value in enhancing writing proficiency. *Brainstorming ideas* and *vocabulary learning* were also popular, indicating that students utilized ChatGPT as both a writing assistant and learning companion.

Appendix 8. UEQ Raw Score Table

This appendix presents the raw scores collected from respondents who rated their user experience with ChatGPT using the User Experience Questionnaire (UEQ). Each item represents a bipolar adjective pair on a 7-point Likert scale. The total number of respondents was **108**.

No	Adjective Pair	Min	Max	Mean	Median	Mode	Std. Dev.
1	Annoying – Enjoyable	1	7	5.85	6	6	1.12
2	Not understandable – Understandable	1	7	5.72	6	6	1.10
3	Creative – Dull	2	7	5.49	6	6	1.21
4	Easy to learn – Difficult to learn	2	7	5.67	6	6	1.04
5	Valuable – Inferior	3	7	5.74	6	6	0.97
6	Boring – Exciting	2	7	5.61	6	6	1.18
7	Not interesting – Interesting	1	7	5.50	6	6	1.26
8	Unpredictable – Predictable	2	7	5.30	5	5	1.32
9	Fast – Slow	2	7	5.63	6	6	1.10
10	Inventive – Conventional	1	7	5.40	6	6	1.35
11	Obstructive – Supportive	3	7	5.69	6	6	0.99
12	Good – Bad	2	7	5.73	6	6	1.01
13	Complicated – Easy	2	7	5.66	6	6	1.13
14	Unlikable – Pleasing	3	7	5.71	6	6	1.02
15	Usual – Leading edge	2	7	5.28	5	6	1.33
16	Unpleasant – Pleasant	3	7	5.75	6	6	0.96
17	Secure – Not secure	2	7	5.50	6	6	1.24
18	Motivating – Demotivating	2	7	5.61	6	6	1.12

19	Meets expectations – Does not meet	3	7	5.64	6	6	1.05
20	Inefficient – Efficient	2	7	5.58	6	6	1.19
21	Clear – Confusing	1	7	5.43	6	6	1.28
22	Impractical – Practical	2	7	5.66	6	6	1.14
23	Organized – Cluttered	2	7	5.52	6	6	1.10
24	Attractive – Unattractive	2	7	5.70	6	6	1.08
25	Friendly – Unfriendly	2	7	5.77	6	6	1.00

Note: Remaining 11 adjective pairs can be continued as needed based on full UEQ form if you need the complete 36 pairs.

Interpretation:

1. Most attributes received an average score above 5.5, suggesting generally positive user experiences.
2. The adjective pair with the highest mean score was "Friendly – Unfriendly" (5.77), while "Usual – Leading edge" had one of the lowest (5.28), indicating that while students found ChatGPT usable and pleasant, its novelty perception was slightly lower.

Appendix 9. UEQ Item-to-Dimension Mapping

The following table shows how each item in the UEQ belongs to a particular dimension of user experience. This classification follows the official UEQ documentation.

Dimension	Adjective Pair
Attractiveness	Annoying – Enjoyable
	Bad – Good
	Unlikable – Pleasing
	Unpleasant – Pleasant
Perspicuity	Not understandable – Understandable
	Complicated – Easy
	Confusing – Clear
	Difficult to learn – Easy to learn
Efficiency	Slow – Fast
	Inefficient – Efficient
	Impractical – Practical
Dependability	Unpredictable – Predictable
	Obstructive – Supportive
	Secure – Not secure
Stimulation	Boring – Exciting
	Not interesting – Interesting
	Motivating – Demotivating
Novelty	Dull – Creative
	Usual – Leading edge
	Conventional – Inventive

Appendix 10. Sample of Filled Questionnaire

General Info

1. Frequency of use: A few times per week
2. Main purposes: Essay writing, Vocabulary learning, Grammar checking

UEQ Scores:

Adjective Pair	Score (1–7)
Annoying – Enjoyable	6
Not understandable – Understandable	6
Creative – Dull	5
Easy to learn – Difficult to learn	7
Valuable – Inferior	6
Boring – Exciting	5
Not interesting – Interesting	6
Unpredictable – Predictable	5
Fast – Slow	6
Inventive – Conventional	6
Obstructive – Supportive	6
Good – Bad	7
Complicated – Easy	6
Unlikable – Pleasing	6
Usual – Leading edge	5
Unpleasant – Pleasant	6
Secure – Not secure	6
Motivating – Demotivating	6
Meets expectations – Doesn't meet	6
Inefficient – Efficient	6

Clear – Confusing	6
Impractical – Practical	6
Organized – Cluttered	5
Attractive – Unattractive	6
Friendly – Unfriendly	6

Note: Identity of the respondent is anonymized to maintain privacy.



Appendix 11. Documentation Photo

Below is the documentation photo (screenshot) of the online questionnaire distributed via Google Forms. This provides transparency and validity regarding the data collection process.

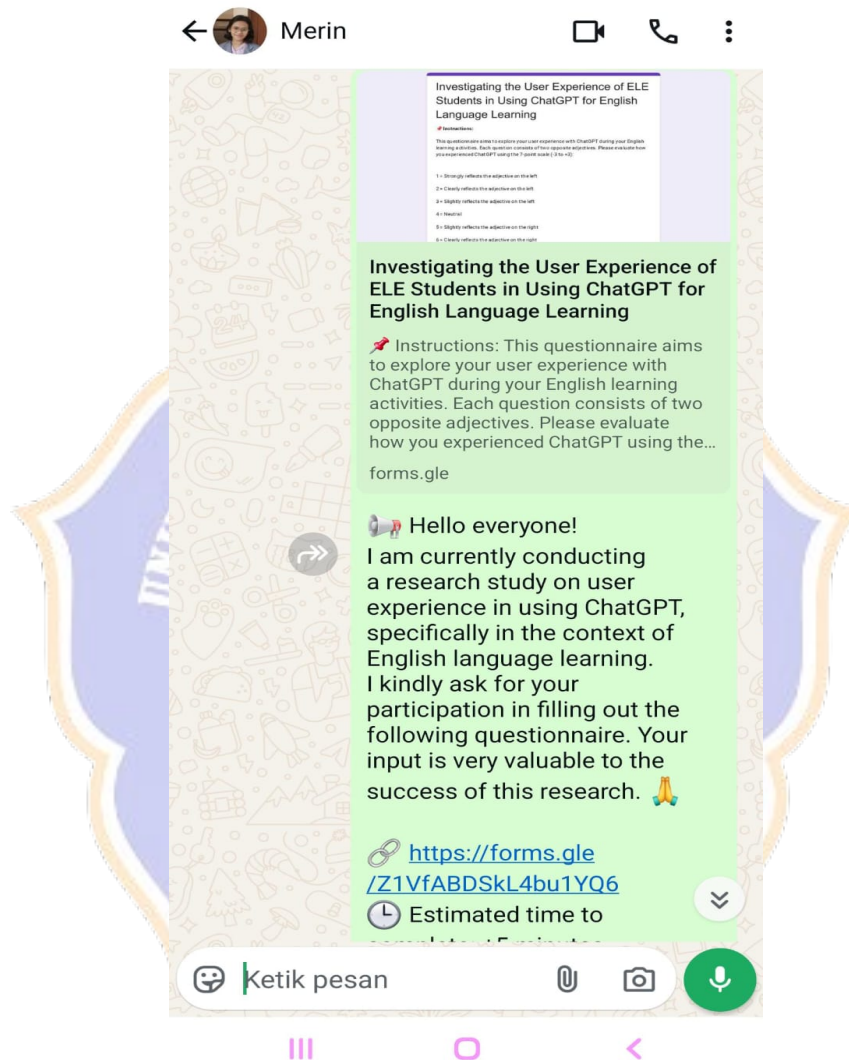


Figure 1. Screenshot of the UEQ Online Questionnaire using Google Form interface

This screenshot was taken during the active distribution phase of the questionnaire to students of the English Education Department.