SO SÁNH DUOLINGO AI VÀ CHATGPT LÀ CÔNG CỤ TỰ HỌC CHO SINH VIÊN KHÔNG CHUYÊN TIẾNG ANH TẠI THÀNH PHỐ HỒ CHÍ MINH

COMPARING DUOLINGO AI AND CHATGPT AS SELF-STUDY TOOLS FOR NON-ENGLISH MAJOR STUDENTS IN HO CHI MINH CITY

NGUYỄN NGỌC THẢO, *thaonn@hcmute.edu.vn*
Trường Đại học Sư phạm Kỹ thuật Thành phố Hồ Chí Minh.

THÔNG TIN

Ngày nhận: 19/8/2025
Ngày nhận lại: 10/10/2025
Duyệt đăng: 18/10/2025
Mã số: TCKH-S04T10-2025-B13
ISSN: 2354 - 0788

Từ khóa:
ChatGPT, duolingo AI, công cụ tự học, sinh viên không chuyên ngữ.
Keywords:
ChatGPT, duolingo AI, self-study tools, non-English-major students.

TÓM TẮT

Nghiên cứu này so sánh hai công cụ trí tuệ nhân tạo hỗ trợ tự học ngoại ngữ là ChatGPT và Duolingo AI trong bối cảnh sinh viên không chuyên tiếng Anh tại Thành phố Hồ Chí Minh. Dựa trên khung lý thuyết UTAUT2 điều chỉnh, năm nhân tố ảnh hưởng đến Behavioral Intention (BI) gồm Performance Expectancy (PE), Effort Expectancy (EE), Hedonic Motivation (HM), Price Value (PV) và Facilitating Conditions (FC) được kiểm định thông qua khảo sát 80 sinh viên. Phân tích bằng Cronbach's Alpha, EFA, hồi quy tuyến tính đa biến và kiểm định T-test độc lập cho thấy HM, PE và FC có tác động tích cực và có ý nghĩa thống kê đến BI, trong khi EE và PV không đáng kể. Kết quả cũng chỉ ra ChatGPT nổi trội về PE và EE, còn Duolingo AI mạnh ở HM và FC nhờ gamification. Những phát hiện này củng cố giá trị của UTAUT2 và gợi ý việc kết hợp hai công cụ để vừa nâng cao hiệu quả vừa duy trì động lực học tập cho sinh viên không chuyên.

ABSTRACT

This study compares two artificial intelligence tools that support self-learning of foreign languages, namely ChatGPT and Duolingo AI, with non-English-major students in Ho Chi Minh City. Based on the adjusted UTAUT2 theoretical framework, five factors affecting Behavioral intention (BI) - Performance expectancy (PE), Effort expectancy (EE), Hedonic motivation (HM), Price value (PV) and Facilitating conditions (FC) were assessed through a survey of 80 students. Analysis by Cronbach's Alpha, EFA, multivariate linear regression and Independent Samples T-test showed that HM, PE, and FC had a positive and statistically significant impact on BI, while EE and PV were negligible. The results also show that ChatGPT excels in PE and EE, while

Duolingo AI is strong in HM and FC thanks to gamification. These findings reinforce the value of UTAUT2 and suggest a combination of the two tools to both improve efficiency and maintain academic motivation for non-professional students.

1. Introduction

In the last decade, the rapid development of artificial intelligence (AI) has created fundamental changes in the way students learn foreign languages on their own, especially non-English majors. AI tools have opened many opportunities to access knowledge, supporting learners to develop communication skills, critical thinking, and the ability to self-manage the learning process. Among them, ChatGPT and Duolingo AI stand out as two popular but distinct tools: ChatGPT is designed as a generalized language model that can support a variety of fields, while Duolingo AI is an application focused on language teaching and learning with a specialized feature system. It is this difference that has led to different potential impacts on the learning experience of non-professional students, especially in the context of higher education in Vietnam, which is under pressure to both ensure outcome standards and create sustainable self-learning motivation for learners (Huang, 2024; Strzelecki, 2024).

However, there has been a lack of empirical studies to date that directly compare the two tools in the same context. Previous works have often focused individually on ChatGPT with its ability to support language creation or on Duolingo with its gamification features that increase motivation and the ability to maintain learning habits (Poveda-Balbuena et al., 2024; Cabero-Almenara et al., 2025). This research gap poses an urgent need for comparative analyses to show the strengths, limitations and suitability of each tool for Vietnamese students.

From that fact, this study aims to compare the academic performance, acceptance and intention to continue using ChatGPT and Duolingo AI in the English self-study process of non-professional students. To achieve the goal, the study posed two key questions: (1) What factors in the UTAUT2 model - including performance expectations, effort expectations, entertainment motivation, perceived value and support conditions - influence the intention to use ChatGPT and Duolingo AI?; (2) Is there any significant difference in acceptance and intent to use between the two tools? Answering the above two questions not only fills the gap in international research but also has practical significance in orienting the selection of tools to support foreign language learning more effectively at training institutions in Vietnam.

2. Content

2.1. Literature review

2.1.1. AI in language learning

AI has become one of the most important trends in the research and application of foreign language teaching over the past decade. AI-based language learning apps are proven to enhance language acquisition and production skills, create personalized learning environments and enhance learner motivation (Smith et al., 2024). It is worth noting that AI not only replaces the role of traditional documents but also expands the scope of the learning experience by simulating real-life communication situations, providing instant feedback and adjusting the difficulty according to the learner's progress. However, the researchers also warn about the sustainability of motivation when learners rely excessively on technology tools instead of developing self-learning strategies (Shortt et al., 2021).

2.1.2. ChatGPT as a self-study tool

Among modern AI models, ChatGPT has emerged as a versatile tool for self-learning English, especially in its open-language interoperability, providing suggestive and creative responses. Strzelecki's (2024) study shows that factors in UTAUT2 such as PE and HM have a strong impact on the intention to adopt ChatGPT in learning. However, an important limitation is that the reliability of the generated content is still unstable, along with the risk that learners lack verification skills and easily fall into a state of dependency (Moradi, 2025). This suggests that ChatGPT is more suitable when used in an adjunct role, combined with direction from a lecturer or standardized documentation.

2.1.3. Duolingo AI

Unlike ChatGPT, Duolingo is essentially a language learning application designed based on gamification; in which, some recent features have been integrated with AI such as Roleplay, Video Call or Explain My Answer. These features not only help learners practice practical communication contexts but also enhance self-efficacy, i.e. confidence in one's own learning capacity (Poveda-Balbuena et al., 2024). Smith et al. (2024) demonstrate that using Duolingo over a three-month period results in significant improvements in vocabulary and communication skills. However, some studies have also shown that as learners reach a higher level, Duolingo can be less effective due to limitations in developing academic language competencies and critical thinking (Shortt et al., 2021).

2.1.4. UTAUT2 framework

Venkatesh et al.'s (2012) UTAUT2 model has become a popular theoretical framework for explaining the adoption and use of the technology. In the context of education, recent studies indicate that five important variables PE, EE, HM, PV and FC are often the factors with the most pronounced influence on behavior using learning technology (Cabero & Almenara et al., 2025). Other variables in UTAUT2 such as "habits" or "social influences" may be important in broader contexts but for this research goal, focusing on these five variables simplifies the model while retaining significant explanation.

2.1.5. Research gap

Although there are many studies applying UTAUT2 to analyze the adoption of ChatGPT or Duolingo individually, there is currently no work directly comparing these two tools in the context of Vietnamese students not majoring in English. This gap is not only academic but also practical, as universities are facing the question: which AI tools support students to learn more effectively on their own and how to optimize their application in higher education.

2.2. Research Model and Hypotheses

Based on the UTAUT2 theoretical framework and findings from previous studies, the project builds a research model to test the factors that influence students' behavioral intentions when using ChatGPT and Duolingo AI in foreign language self-learning. The model was adapted to focus on five key factors - PE, EE, HM, PV and FC, thereby forming specific research hypotheses.

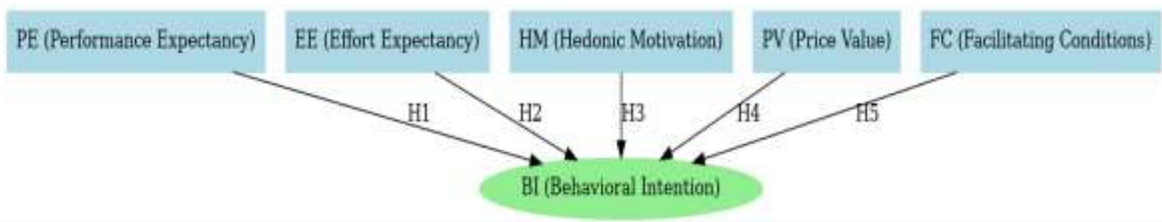

Figure 1. The study model of ChatGPT and Duolingo AI adoption in self-learning of foreign languages by non-specialized students according to UTAUT2

In this research model, five hypotheses were proposed to examine the influence of factors in the UTAUT2 framework that govern students' behavioral intentions (BI) when using foreign language self-learning tools. Specifically, as follows:

H1: Performance Expectancy has a positive effect on BI.

H2: Effort Expectancy has a positive effect on BI.

H3: Hedonic Motivation has a positive effect on BI.

H4: Price Value has a positive effect on BI.

H5: Facilitating Conditions have a positive effect on BI.

2.3. Methodology

In this study, a total of 80 non-English language majors in Ho Chi Minh City participated in the study, divided into two groups: 40 students using ChatGPT and 40 students using Duolingo AI. Participants were selected according to a convenient sample selection method combined with voluntary registration. Most of them are second- and third-year students in socio-economic disciplines of a public university in Ho Chi Minh City in the second semester of the 2024 - 2025 academic year. The selection criteria include: (i) being a non-specialized student, (ii) having a minimum English proficiency level 2 according to Vietnam's 6-level foreign language proficiency framework and (iii) not having experience in regularly using ChatGPT or Duolingo AI before the time of research.

The research process is conducted in four steps. Firstly, the lecturer announces and invites students to register to participate in the class. Second, the students were randomly assigned to two groups corresponding to the two tools. Third, for four weeks, students are instructed to experience the assigned tool for an average of 3-5 hours per week, under the supervision of the instructor in charge. Finally, after the experience period, a 5-level Likert questionnaire based on the UTAUT2 framework was played to collect quantitative data. In addition, 10 students were randomly selected from both groups to participate in semi-structured interviews to supplement the qualitative data for the analysis. This way of organizing ensures scientific, accuracy, and reliability, while limiting the influence of individual differences in previous experience.

Data analysis: Data were processed using JASP 18.0 in the following steps: scale reliability testing using Cronbach's Alpha ($\alpha \geq 0.70$) (Nunnally & Bernstein, 1994), discovery factor analysis (EFA) with eigenvalue > 1 and load factor ≥ 0.50 (Hair et al., 2010), multivariate linear regression analysis to test the H1-H5 hypothesis, and the Independent T-Test to compare the differences between the two groups. Besides, descriptive statistics are used to determine the mean and standard deviation of variables. This approach allows both to evaluate the reliability of the scale and to compare its effectiveness, acceptance and intention to continue using ChatGPT and Duolingo AI in the context of non-English-major students in Ho Chi Minh City.

2.4. Findings

Before presenting the results of the analysis, it is necessary to recall the scale structure used in the study. Based on the UTAUT2 theoretical framework, the factors were measured by 18 observed variables in the questionnaire, distributed as follows: i) Performance Expectancy including Q1 - Q3; ii) Effort Expectancy including Q4 - Q6; iii) Hedonic Motivation including Q7 - Q9; iv) Price Value including Q10 - Q12; v) Facilitating Conditions including Q13 - Q15 and vi) Behavioral Intention including Q16 - Q18. This division is the basis for conducting reliability tests, factor analysis, linear regression and comparison between the two groups in the next steps.

2.4.1. Reliability test

Before conducting the factor analysis, the study used Cronbach's Alpha to assess the intrinsic reliability of the scales. The results showed that the reliability between the scales fluctuated significantly, reflecting different degrees of uniformity in how students rated each aspect when experiencing the two self-learning tools.

For the Performance Expectancy (PE) scale, the Cronbach's Alpha coefficient is 0.495 (Table 1). This value shows that the internal consistency is not high, the Q1 - Q3 observation variables do not uniformly reflect students' perception of the effective benefits of using the tool. However, as suggested by Nunnally and Bernstein (1994), scales with Cronbach's Alpha above 0.45 may still be acceptable during the exploratory research phase.

Table 1. Cronbach's Alpha results for the PE scale

Estimate	**Cronbach's α**
Point estimate	0.495
95% CI lower bound	0.272
95% CI upper bound	0.659
Frequentist Individual Item Reliability Statistics	
Item	**Cronbach's α**
Q1	0.327
Q2	0.387
Q3	0.454

For the Effort Expectancy scale, Cronbach's Alpha coefficient is 0.521 (Table 2). This value is higher than PE but still below the 0.7 threshold, indicating limited internal consistency.

The Q4-Q6 variables ranged from 0.328 to 0.499, with Q5 being the lowest, reflecting the dispersion in the evaluation of ease of use. However, according to Nunnally and Bernstein (1994), an Alpha level above 0.5 is still acceptable in exploratory research, so the EE scale is retained for the next step of factor analysis.

Table 2. Cronbach's Alpha results for the EE scale

Estimate	**Cronbach's α**
Point estimate	0.521
95% CI lower bound	0.304
95% CI upper bound	0.679
`Frequentist Individual Item Reliability Statistics	
Item	**Cronbach's α**
Q4	0.422
Q5	0.328
Q6	0.499

Meanwhile, the Hedonic Motivation scale only reaches Cronbach's Alpha = 0.379 (Table 3), which is significantly lower than the acceptance threshold. The observation variables Q7 - Q9 showed a poor degree of uniformity, reflecting that students rated the interesting and inspiring aspect of the tool as highly differentiated. These results suggest that the removal or adjustment of this scale should be considered in future studies, to increase the reliability of the measurement.

Table 3. Cronbach's Alpha results for the HM scale

Estimate	**Cronbach's α**
Point estimate	0.379
95% CI lower bound	0.100
95% CI upper bound	0.583

Frequentist Individual Item Reliability Statistics	
Item	**Cronbach's α**
Q7	0.264
Q8	0.362
Q9	0.232

The Price Value scale has Cronbach's Alpha at 0.249 (Table 4), which is below the acceptance threshold. This suggests a lack of stability between the Q10-Q12 observed variables, reflecting large differences in how students perceive the correlation between costs and benefits. As a result, the PV scale has very low reliability and is difficult to ensure its value at the factor analysis stage.

Table 4. Cronbach's Alpha results for the PV scale

Estimate	**Cronbach's α**
Point estimate	0.249
95% CI lower bound	-0.091
95% CI upper bound	0.496
Frequentist Individual Item Reliability Statistics	
Item	**Cronbach's α**
Q10	0.274
Q11	0.090
Q12	0.180

In contrast, the Facilitating Conditions scale showed quite positive results with Cronbach's Alpha = 0.664 (Table 5). The Q13 - Q15 observation variables showed relatively high consistency, reflecting student consensus in assessing infrastructure, technical support and encouragement from the learning environment. This result shows that FC is one of the most stable scales in the study.

Table 5. Cronbach's Alpha results for the FC scale

Estimate	**Cronbach's α**
Point estimate	0.664
95% CI lower bound	0.512
95% CI upper bound	0.775
Frequentist Individual Item Reliability Statistics	
Item	**Cronbach's α**
Q13	0.595
Q14	0.608
Q15	0.500

Finally, the Behavioral Intention scale has Cronbach's Alpha = 0.373 (Table 6), indicating low reliability. The Q16 - Q18 observation variables do not produce strong enough uniformity to reflect the intention to continue using the tool. This partly explains why in later analysis steps, BI may be affected by the removal of the observed variable or the restructuring of the scale.

Table 6. Cronbach's Alpha results for the BI scale

Estimate	**Cronbach's α**
Point estimate	0.373
95% CI lower bound	0.093
95% CI upper bound	0.578
Frequentist Individual Item Reliability Statistics	
Item	**Cronbach's α**
Q16	0.249
Q17	0.240
Q18	0.350

The results of the reliability test show that the scales have varying degrees of consistency, with some scales only meeting the minimum acceptance level, even below the recommended threshold. This requires further factor analysis to consider the underlying structure of the toolkit, and to identify inappropriate observational variables. Therefore, the next step in the study is to conduct an exploratory factor analysis to reinforce the structural value of the scales.

2.4.2. Discovery Factor Analysis (EFA)

To examine the underlying structure of the toolkit, discovery factor analysis is first conducted with 18 observed variables. The discovery factor analysis was initially conducted with 18 observational variables to validate the structural value of the toolkit. The results showed that the data were suitable for analysis, however some variables were unsatisfactory. Specifically, the Q7 and Q8 have a low load factor (<0.45) and high uniqueness above 0.75, the Q11 has a "uniqueness" that exceeds 0.8, while the Q17 even has a negative load factor.

Table 7. EFA test results and factor load coefficient of 18 observed variables

	Value	**df**	**p**
Model	46.262	60	0.904

Factor Loadings							
	Factor 1	**Factor 2**	**Factor 3**	**Factor 4**	**Factor 5**	**Factor 6**	**Uniqueness**
Q15	0.709						0.533
Q13	0.663						0.611
Q14	0.631						0.414
Q10		0.697					0.430
Q18		0.689					0.454
Q12			0.662				0.466
Q9			0.504				0.670
Q7			0.442				0.750
Q17			-0.430				0.530
Q5				0.635			0.535
Q4				0.521			0.721
Q6				0.494			0.571
Q1					0.557		0.648

Factor Loadings							
	Factor 1	**Factor 2**	**Factor 3**	**Factor 4**	**Factor 5**	**Factor 6**	**Uniqueness**
Q2					0.555		0.677
Q3					0.435		0.611
Q16						0.677	0.590
Q8							0.952
Q11							0.811

According to the recommendation of Hair et al. (2010), variables with a loading factor of less than 0.4 or "uniqueness" above 0.7 should be removed to ensure convergence values and avoid interference with the model. Therefore, these four variables were excluded before re-analysis. When conducting EFA with the remaining 14 variables, the results showed that the measurement model achieved a high level of relevance (Chi-square = 6.600, df = 15, p = 0.968). The factor load coefficients all reached above 0.48, many variables exceeded 0.6, proving that the observed variables converged well into six factors. The six extracted factors correspond to the original theoretical structure of UTAUT2, including Effort Expectancy, Hedonic Motivation, Price Value, Facilitating Conditions and Behavioral Intention. This confirms that after fine-tuning, the measurement toolkit becomes stable and has a solid basis for use in subsequent analysis.

Table 8. EFA test results and factor load factor 2 of 14 observed variables

Exploratory Factor Analysis			
	Value	**df**	**p**
Model	6.600	15	0.968

Factor Loadings							
	Factor 1	**Factor 2**	**Factor 3**	**Factor 4**	**Factor 5**	**Factor 6**	**Uniqueness**
Q15	0.763						0.373
Q13	0.588						0.629
Q14	0.576						0.577
Q10		0.972					-0.004
Q5			0.712				0.440
Q6			0.485				0.589
Q4			0.484				0.759
Q1				0.698			0.488
Q2				0.514			0.657
Q16					0.995		0.004
Q18						0.475	0.625
Q3							0.691
Q12							0.845

In addition, the analysis results showed that the six factors extracted after removing the inappropriate variables explained 48.7% of the total variance. Although this figure does not exceed the 50% threshold, it is considered acceptable in social science research, where learner behavior is often governed by many random factors (Peterson, 2000). According to Costello and Osborne (2005), sometimes an interpretation rate of 50-60% can be considered sufficient in practical EFA application.

Table 9. Eigenvalues and cumulative variance of factors following EFA

	Unrotated solution				**Rotated solution**		
	Eigenvalues	**SumSq. Loadings**	**Proportion var.**	**Cumulative**	**SumSq. Loadings**	**Proportion var.**	**Cumulative**
Factor 1	2.055	1.541	0.119	0.119	1.342	0.103	0.103
Factor 2	1.920	1.512	0.116	0.235	1.205	0.093	0.196
Factor 3	1.541	1.152	0.089	0.324	1.066	0.082	0.278
Factor 4	1.399	0.968	0.074	0.398	1.018	0.078	0.356
Factor 5	1.202	0.834	0.064	0.462	1.015	0.078	0.434
Factor 6	0.887	0.326	0.025	0.487	0.679	0.052	0.487

In conclusion, the EFA results have contributed to affirming the structural value of the toolkit. After removing the four unsatisfactory observed variables, the remaining 14 variables converged into six stable factors, reflecting the original theoretical basis. This is the foundation for continuing the regression analysis to test research hypotheses.

2.4.3. Research Hypothesis testing (Linear Regression)

After the analysis of the discovery factor confirms the structure of the 6 stable factors, the next step is to test the research hypotheses through multivariate linear regression. The goal of this step is to determine the extent to which each independent factor (PE, EE, HM, PV, FC) affects the dependent variable Behavioral Intention (BI). The results of the analysis showed that the regression model achieved values of $R = 0.626$ and $R^2 = 0.392$, i.e., five independent factors explained about 39.2% of BI's variability. This is a pretty good explanation in social science research, which is influenced by many complex factors. The ANOVA test value ($F = 9.555$, $p < .001$) confirms that the model is statistically significant overall, i.e. that at least one factor in the model has a significant impact on BI (Table 9).

Table 10. Results of testing of multivariate linear regression model

Model	**R**	**R²**	**Adjusted R²**	**RMSE**
H_0	0.000	0.000	0.000	0.440
H_1	0.626	0.392	0.351	0.355

ANOVA						
Model		**Sum of Squares**	**df**	**Mean Square**	**F**	**p**
H_1	Regression	6.005	5	1.201	9.555	< .001
	Residual	9.301	74	0.126		
	Total	15.306	79			

Coefficients						
Model		**Unstandardized**	**Standard Error**	**Standardized**	**t**	**p**
H_0	(Intercept)	3.463	0.049		70.371	< .001
H_1	(Intercept)	0.294	0.591		0.498	0.620
	PE	0.327	0.088	0.344	3.711	< .001
	EE	0.021	0.090	0.022	0.233	0.817

Coefficients						
Model		**Unstandardized**	**Standard Error**	**Standardized**	**t**	**p**
	HM	0.368	0.074	0.454	4.941	< .001
	PV	-0.030	0.077	-0.036	-0.388	0.699
	FC	0.232	0.082	0.262	2.835	0.006

In addition, the analysis of the normalized regression coefficient (β) shows that three factors have a positive and statistically significant influence on BI. Specifically, Hedonic Motivation has a coefficient of $\beta = 0.454$, $p < .001$, which is the most powerful factor in Behavioral Intention, indicating that the interest and joy of using the tool play a central role in motivating students to continue learning. Performance Expectancy also had a significant impact ($\beta = 0.344$, $p < .001$), reflecting that when students believe in the learning benefits that the tool provides, they tend to be more engaged. In addition, Facilitating Conditions reached $\beta = 0.262$, $p = 0.006$, proving that the infrastructure environment and convenience in the use process also contribute significantly to the formation of intentions. In contrast, Effort Expectancy ($\beta = 0.022$, $p = 0.817$) and Price Value ($\beta = -0.036$, $p = 0.699$) were not statistically significant, suggesting that ease of use or cost considerations were not decisive factors for the group of students surveyed. This result is consistent with the research context, as both tools have user-friendly interfaces and most students reach out through free plans, so the cost factor does not make a significant difference.

On that basis, linear regression confirms that the three research hypotheses H1, H3 and H5 are supported, while H2 and H4 are refuted. These results show that intention to use English self-learning tools is most affected by Hedonic Motivation, followed by Performance Expectancy and Facilitating Conditions. This is an important basis for a deeper discussion of the role of each element in the next section.

2.4.4. Comparison between ChatGPT and Duolingo AI (Independent Samples T-test)

The regression results identified three main factors affecting behavioral intention, but did not show whether there were significant differences between the two surveyed tools. To answer this question, the study conducted an independent T-test to compare the average value of factors between two groups of students using ChatGPT and Duolingo AI.

The results showed that most of the factors did not differ statistically significantly between the two groups, including Performance Expectancy ($p = 0.937$), Effort Expectancy ($p = 0.806$), Price Value ($p = 1.000$), Facilitating Conditions ($p = 0.824$) and Behavioral Intention ($p = 0.138$). This reflects that in general, students rated the two tools quite similar in terms of learning effectiveness, ease of use, cost and facilitating conditions, as well as the level of willingness to continue using them.

Table 11. Independent Samples T-test results and descriptive statistics between the two teams of ChatGPT and Duolingo AI

Independent Samples T-Test					
	t	**df**	**p**	**Cohen's d**	**SE Cohen's d**
PE	0.080	78	0.937	0.018	0.224
EE	0.246	78	0.806	0.055	0.224
HM	-1.992	78	0.050	-0.445	0.229
PV	0.000	78	1.000	0.000	0.224

Independent Samples T-Test					
	t	df	p	Cohen's d	SE Cohen's d
FC	0.224	78	0.824	0.050	0.224
BI	-1.498	78	0.138	-0.335	0.227

Group Descriptives						
	Group	N	Mean	SD	SE	Coefficient of variation
PE	ChatGPT	40	3.567	0.513	0.081	0.144
	Duolingo AI	40	3.558	0.416	0.066	0.117
EE	ChatGPT	40	3.217	0.450	0.071	0.140
	Duolingo AI	40	3.192	0.458	0.072	0.144
HM	ChatGPT	40	3.225	0.493	0.078	0.153
	Duolingo AI	40	3.462	0.571	0.090	0.165
PV	ChatGPT	40	3.313	0.463	0.073	0.140
	Duolingo AI	40	3.313	0.596	0.094	0.180
FC	ChatGPT	40	3.492	0.420	0.066	0.120
	Duolingo AI	40	3.467	0.569	0.090	0.164
BI	ChatGPT	40	3.390	0.441	0.070	0.130
	Duolingo AI	40	3.536	0.432	0.068	0.122

The notable point lies in the Hedonic Motivation, with p = 0.050 and Cohen's coefficient d = -0.445, approaching the meaning threshold of 0.05 and reaching the average effect. The descriptive results showed that the group of students using Duolingo AI (Mean = 3,462) had a higher HM score than the ChatGPT group (Mean = 3,225). This suggests that Duolingo AI has the advantage of being able to excite and motivate users to learn, thanks to gamification elements and close interaction with the language learning environment.

In general, the T-test analysis indicates that except for HM which has a slight difference in favor of Duolingo AI, other factors are evaluated similarly between the two tools. This result combined with previous regression suggests that although HM is a factor with a strong impact on Behavioral Intention, the differences between the two tools are mainly concentrated in this aspect, while the remaining factors do not create a significant gap.

2.5. Discussion

2.5.1. Interpretation of results under the UTAUT2 theoretical framework

The results of the study show that three factors have a significant impact on students' behavioral intentions in using language self-learning tools: leisure motivation, expectations of effectiveness and supportive conditions. This finding not only reaffirms the UTAUT2 theoretical model but also suggests how students make decisions in the context of self-learning using emerging technology. According to UTAUT2, in addition to the perceived factor of usefulness, the behavior of using technology is also governed by emotional motivations and favorable environmental conditions (Nikolopoulou & Gialamas, 2021). The current experimental results thus both reinforce the core thesis of the theory and provide concrete evidence in the context of non-specialized foreign language learning. Notably, HM emerged as the strongest factor, reflecting the decisive role of positive emotions in maintaining behavior. This makes perfect sense when placed in a self-learning environment, where a lack of external motivation makes it easy for learners to give up if the tool does not provide interest. HM's superiority is reminiscent of the notion that in online learning systems, the emotional factor is often as important as or even greater than the cognitive factor (Zheng, 2025). As such, learners' continued use is not only based on objective benefits but also depends on how much they find the tool interesting and worth experiencing.

Meanwhile, PE and FC also show a role that cannot be ignored. PE reflects learners' belief that using the tool will help them improve their skills, and this belief has a positive impact on behavioral intentions, which is consistent with many previous studies in the field of educational technology (Grassini, 2024). FC emphasizes the aspect of the supporting environment, from technological infrastructure to convenience of access, which if guaranteed will create conditions for learners to stick with the tool. The importance of FC once again confirms that behavioral intention cannot be explained only in terms of individual psychology but also closely tied to organizational conditions and the learning environment. Overall, when interpreted through the lens of UTAUT2, the behavior of non-specialized students using self-learning tools is simultaneously affected by joy and excitement, belief in learning effectiveness and conditional support. This convergence not only reflects the correctness of the theoretical model but also provides an empirical basis for asserting that, in the context of self-directed education, emotional motivation and a favorable environment are key factors for sustaining long-term learning behavior (Venkatesh et al., 2012; Nikolopoulou & Gialamas, 2021; Zheng, 2025).

2.5.2. Explaining the Difference Between ChatGPT and Duolingo AI

Upon further analysis, the behavior of using two self-learning tools, ChatGPT and Duolingo AI, reflects a fundamental difference in the way users are designed and experienced. ChatGPT stands out for its high personalization and flexibility, helping students receive feedback almost instantly and tailored to their specific needs. This characteristic enhances the perception of efficiency expectations and effort expectations, as learners believe that they easily achieve their learning goals without significant hindrance. Recent research by Wang and Fan (2025) also shows that ChatGPT has a marked influence on academic performance as well as students' positive perceptions of the tool's usefulness.

Meanwhile, Duolingo AI offers a differentiated experience by focusing on gamification design with a system of rewards, badges and levels and providing a clear and structured learning path. This helps to increase leisure motivation and reinforce supportive conditions, as students are both encouraged to maintain interest and feel companionship in the learning process. Shortt et al. (2023) said that Duolingo is one of the typical examples of language learning with the application of gamification, thereby improving motivation and engagement. Complementing this argument, Díaz et al. (2024) through meta-analysis demonstrated that gamification contributes to a significant improvement in motivation levels in online learning, while Tajik (2025) points out that AI tools have a gamification element that helps learners maintain active engagement over the long term.

From the above analysis, ChatGPT is strong in on-demand language processing and instant responses, while Duolingo AI creates a cohesive, systematic and entertaining learning environment. Thereby, these two tools do not exclude each other but complement each other, and the correct identification of the strengths of each platform opens the direction of integration to build a more optimal self-study program for non-professional students.

2.5.3. Practical implications

The results showed that Hedonic motivation, Performance expectancy and Facilitating conditions were the three factors that were significant in the formation of Behavioral Intention while Effort expectancy and Price value were not statistically significant. The first implication to be confirmed is the nature of the tool in the study: ChatGPT is a generative AI tool and Duolingo is a language learning application that integrates several AI features. Therefore, practical solutions must be based on the differences in design principles and user experience analyzed in section 6.2. When placed within the UTAUT2 framework, the coordination between the two tools is not substitute but complementary, to amplify HM, strengthen PE and create sustainable FC at both the classroom and program levels (Venkatesh et al., 2012; Xue, 2024; Zheng, 2025).

From a learner's perspective, the combination of ChatGPT and Duolingo allows for cognitive learning and motivational performance at the same time. ChatGPT helps students receive instant and personalized feedback, while Duolingo encourages regular practice through gamification. Meta-analyses have demonstrated gamification has a positive impact on motivation and academic engagement (Díaz, 2024; Li et al., 2024) and many studies indicate that Duolingo can improve self-efficacy and language proficiency at an early stage (Poveda-Balbuena et al., 2024; Smith et al., 2024). However, it is also important to note that gamification can lead to an emphasis on task completion rather than the development of an academic mindset, as reflected in an overview by Shortt et al. (2021). This reinforces the argument that ChatGPT is suitable for tasks that require academic discourse, while Duolingo is useful for forming long-lasting habits and interests. At the instructional and program level, this coordination allows for the design of logical structured sequences of activities: boot up with Duolingo to activate motivation, continue with ChatGPT for language processing and feedback and return to Duolingo to sustain learning behavior. This result is compatible with many studies on the central role of PE and HM in the adoption of educational technology (Grassini, 2024; Habibi et al., 2024; Moradi, 2025; Cabero-Almenara et al., 2025) as well as affirming that FC does not only stop at technical infrastructure but also includes pedagogical standards, support resources and feedback mechanisms (Duan, 2024).

Finally, to avoid misunderstandings about the nature of the tool, it is necessary to clearly communicate that Duolingo is not a purely AI tool, but a language learning application with integrated AI in some functions. This delineation is consistent with the UTAUT2 framework, which emphasizes contextual and task-based differences, and is in harmony with studies on gamification in language learning and recent efficacy assessments of the adoption of AI technologies (Shortt et al., 2021; Díaz, 2024; Li et al., 2024; Grassini, 2024). When operated on the right understanding, the combination of ChatGPT and Duolingo will take advantage of each tool to both improve the efficiency of instant learning and maintain long-term motivation, thereby realizing the core implications of UTAUT2 in the context of AI-powered language learning in Vietnam (Venkatesh et al., 2012; Xue, 2024; Zheng, 2025).

3. Conclusion

Research has shown that during the use of language self-learning tools, the three factors Hedonic Motivation, Performance Expectancy and Facilitating Conditions have a positive and statistically significant effect on students' Behavioral Intention, while Effort Expectancy and Price Value have no significant impact. The results of the comparison also prove that ChatGPT excels in instant and personalized support, while Duolingo, which is a language learning app with several AI-integrated features, stands out for its gamification factor that helps maintain interest and persistence in learning.

Scientifically, the study has expanded the scope of application of the UTAUT2 theoretical framework in the context of foreign language education. The results add to the empirical evidence that not all variables in the model have a uniform impact but that some factors such as HM, PE and FC play a central role in shaping technology-enabled learning behavior. The comparison of two tools of different nature, in which ChatGPT is a generic AI tool and Duolingo is a language learning application with integrated AI, has clarified the differences and complementarity, thereby contributing a new point compared to previous studies that have focused on a single tool.

In practical terms, the results of the study offer many useful suggestions. Students can use both tools simultaneously to optimize the learning process, with ChatGPT responding to the need for in-depth explanations and personalized feedback and Duolingo helping to maintain interest and long-term engagement. Instructors can integrate ChatGPT into lectures to provide timely guidance and feedback, while assigning learning tasks through Duolingo to increase engagement. For program designers, the combination of these two tools opens a new direction in self-learning activity design, balancing immediate efficiency and perseverance, in line with Xue's (2024) suggestion.

Besides the contributions, there are still some limitations to the research. The short conduct time, small sample size and scope of the survey are confined to one university, making the results unrepresentative of the entire Vietnamese student context. These are factors to consider when paraphrasing. In the future, research may expand in many directions, such as assessing the impact of AI tools on other skills such as writing or listening, expanding the survey audience to specialized students or conducting comparisons with many other AI platforms to have a more comprehensive view of the role of artificial intelligence in foreign learning languages.

REFERENCES

Cabero-Almenara, J., Palacios-Rodríguez, A., Rojas Guzmán, H. d. l. Á., & Fernández-Scagliusi, V. (2025). *Prediction of the use of generative artificial intelligence through ChatGPT among Costa Rican university students: A PLS model based on UTAUT2*. Applied Sciences, *15(6)*, 3363. https://doi.org/10.3390/app15063363

Costello, A. B., & Osborne, J. W. (2005). *Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. Practical Assessment, Research & Evaluation, 10(7).* https://doi.org/10.7275/jyj1-4868

Díaz, A. F. (2024). A meta-analysis on the effectiveness of gamification in student learning achievement. *International Journal of Education in Mathematics, Science and Technology*. DOI:10.46328/ijemst.4185.

Duan, C. (2024). Examining UTAUT2 factors in adoption and implementation of open education. *Asian Association of Open Universities Journal, 19*(3), 313-330. 10.1108/AAOUJ-09-2024-0119

Grassini, S. (2024). *Understanding university students' acceptance of ChatGPT*. Interactive Learning Environments. https://doi.org/10.1080/08839514.2024.2371168

Habibi, A., et al. (2024). *Understanding university students' acceptance of ChatGPT: Insights from the UTAUT2 model*. Applied Artificial Intelligence, *38*(1). https://doi.org/10.1016/j.appai.2024.102331.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Pearson Prentice Hall.

Li, L., et al. (2024). *Gamification enhances student intrinsic motivation in technology-enhanced learning: A meta-analysis*. Educational Technology Research and Development, *72*(2), 345-369. https://doi.org/10.1007/s11423-023-10337-7

Merz, A. (2025). *Performance expectancy and social influence drive the acceptance of immersive technologies: A modified UTAUT2 model*. Virtual Reality. https://doi.org/10.1007/s10055-025-01182-5

Moradi, H. (2025). Integrating AI in higher education: Factors influencing ChatGPT acceptance among Chinese university EFL students. *International Journal of Educational Technology in Higher Education, 22*, Article 30. https://doi.org/10.1186/s41239-025-00530-4.

Nikolopoulou, K., & Gialamas, V. (2021). *Habit, hedonic motivation, performance expectancy and acceptance: An extension of UTAUT2*. Computers in Human Behavior Reports, *3*, 100041. https://doi.org/10.1016/j.chbr.2021.100041.

Peterson, R. A. (2000). *A meta-analysis of variance accounted for and factor loadings in exploratory factor analysis*. Marketing Letters, *12*(3), 157-169. DOI:10.1023/A:1008191211004.

Poveda-Balbuena, M. (2024). *The digital language learning platform Duolingo as an effective and motivational tool at initial stages of face-to-face instruction*. In Language, technology, humanities in society *5.0* (pp. 219-232). V&R Unipress. https://doi.org/10.14220/9783737017633.219

Shortt, M., Tilak, S., Kuznetcova, I., Martens, B., & Akinkuolie, B. (2021*). Gamification in mobile-assisted language learning: a systematic review of Duolingo literature from public release of 2012 to early 2020*. Computer Assisted Language Learning, *36*(3), 517-554. https://doi.org/10.1080/09588221.2021.1933540.

Smith, B., Jiang, X., & Peters, R. (2024). *The effectiveness of Duolingo in developing receptive and productive language knowledge and proficiency*. Language Learning & Technology, *28*(1), 1-26. https://hdl.handle.net/10125/73595.

Strzelecki, A. (2024). *Students' acceptance of ChatGPT in higher education: An extended UTAUT2 model*. Innovative Higher Education, *49*(2), 223-245. https://doi.org/10.1007/s10755-023-09686-1.

Tajik, A. (2025). *Gamified and non-gamified AI tools in enhancing EFL listening comprehension: An analysis of Duolingo and Replika's impact on engagement, motivation, and learning outcomes* (Version 1) [Preprint]. Research Square. https://doi.org/10.21203/rs.3.rs-6032009/v1

Urbaite, G. (2025). Adaptive learning with AI: How bots personalize foreign language education. *Euro-Global Journal of Linguistics and Language Education, 2*(1). https://doi.org/10.69760/lumin.20250001002.

Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). *Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology*. MIS Quarterly*, 36(1), 157-178*. https://doi.org/10.2307/41410412.

Vo, T. K. A. (2025). *Transforming language learning with AI: Adaptive learning environments and personalized support*. Applied Sciences, *15*(7), 5678. https://doi.org/10.3390/app15075678.

Wang, J., & Fan, W. (2025). *The effect of ChatGPT on students' learning performance and perception: A meta-analysis*. Humanities and Social Sciences Communications. https://doi.org/10.1057/s41599-025-04787-y.

Xue, L. (2024). *The Unified Theory of Acceptance and Use of Technology (UTAUT) in higher education: A review*. SAGE Open, *14*(1). https://doi.org/10.1177/21582440241229570

Zheng, H. (2025). *A meta-analytic review based on the UTAUT2 model*. Education and Information Technologies.https://doi.org/10.1007/s10639-024-13299-2.