

The effectiveness of automated writing evaluation: a structural analysis approach

Abdulaziz B. Sanosi¹, Mohammed Omar Musa Mohammed²

¹Department of English Language and Literature, College of Science and Humanities, Prince Sattam bin Abdulaziz University, Hawtat Bani Tamim, Saudi Arabia

²Department of Accounting, College of Business Administration, Prince Sattam bin Abdulaziz University, Hawtat Bani Tamim, Saudi Arabia

Article Info

Article history:

Received Sep 3, 2022

Revised Sep 22, 2023

Accepted Oct 5, 2023

Keywords:

Automated corrective feedback

Feedback

Grammarly

Technology acceptance

Writing correction

ABSTRACT

Modern advancement in learning technologies and tools has presented innovative written corrective feedback (WCF) methods based on artificial intelligence (AI) and existing corpora. Research has shown that these tools are perceived as exciting and useful by students, yet studies on their effectiveness and impact on students' writing are relatively insufficient. To this end, the present study investigated the effectiveness of Grammarly writing assistant as perceived by 98 undergraduates who used the tool for a 14-week semester. The study adopted a questionnaire based on a modified technology acceptance model (TAM). The gathered data was analyzed using SmartPLS 3 software. The results revealed that different factors predict students' perceptions about Grammarly and their intention to use it. Some of these factors were not presupposed. The findings imply using Grammarly as an extra learning tool rather than a basic one. It is suggested that future research on the efficacy of Grammarly should adopt longitudinal and experimental approaches.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Abdulaziz B. Sanosi

Department of English Language and Literature, College of Science and Humanities,

Prince Sattam bin Abdulaziz University

Hawtat Bani Tamim 16628, Saudi Arabia

Email: a.assanosi@psau.edu.sa

1. INTRODUCTION

For several decades, it has been an essential practice to provide written corrective feedback (WCF) to English as foreign language (EFL) learners. Instructors ordinarily ask their students to revise their work according to these remedial suggestions and repeat the process as often as possible [1]. Parallely, researchers have been engaged in a long-running debate regarding the effectiveness of WCF and its impact on EFL learners. Unfortunately, no consensus has been reached yet regarding answers to these questions. For some researchers, WCF promotes EFL learners' writing and "significantly improves accuracy" [2]. Contrariwise, other researchers believe that WCF is not only ineffectual for both logical and functional reasons but also "has a harmful effect" [3].

Parallel to the recent technological advancements, WCF has witnessed significant transformations similar to other teaching and learning strategies. Researchers and teachers have become more interested in automated writing evaluation (AWE), in which learners' writing can be evaluated using artificial intelligence (AI) applications and existing corpora. These applications are basically commended as they can solve problems of heavy workload on teachers that are likely to prevent them from providing sufficient and accurate WCF to their students [1]. It was established that AWE effectively appraises students' spelling,

grammar, word choice, tone, and plagiarism [4]. Yet again, there is no decisive proof regarding their impact on improving students' accuracy, whether in the long or short term.

Up to now, research related to AWE has provided sufficient insights to understand the approach. Previous studies investigated how this technology is used in teaching, and the perceived benefits learners could gain [5]. Furthermore, several studies evaluated the types of feedback presented by AWE applications [6], while other studies explored the effect of AWE applications in improving the student's performance in writing [7], [8]. Notwithstanding, relatively few studies have explored AWE's effectiveness, how it affects the learning process, and the learners' attitudes towards it.

The lack of studies investigating AWE is unfortunate since adopting computer assisted language learning (CALL) methods in language teaching will likely entail various students' prospects and attitudes. Consequently, evaluating the appropriateness of a CALL program is determined by several factors, including learners' expertise, cognitive overhead, the role of the learner, and technological suitability [9]. These factors formulate learners' attitudes towards the program. Since it has been proved that "positive attitudes are associated with a willingness to keep learning" [10], it will be intuitive to investigate EFL students' views on the effectiveness of new CALL applications and tools. Findings of such inquiries can generate implications for using and developing such tools and applications and provide suggestions on their use. To this end, the present study investigates the effectiveness of Grammarly, a well-known AI-based AWE application, from the EFL learners' viewpoints. The researchers adopt a modified model based on the technology acceptance model (TAM) [11] to assess the effectiveness of Grammarly through four factors: perceived usefulness (PU), perceived ease of use (PEOU), perceived self-efficacy (PSE), and perceived enjoyment (PE). These factors are hypothesized to positively affect the learners' attitudes (intention to use) Grammarly and hence prove the effectiveness of the programs.

2. LITERATURE REVIEW

2.1. Corrective feedback

Teachers provide one of two types of feedback to language learners. It can be positive to reinforce correct language production or corrective, including "any reaction of the teacher which transforms, disapprovingly refers to, or demands improvement of the learner utterance" [12]. Both methods are believed to raise learners' motivation and ensure linguistic accuracy [13]. This practice is deeply rooted in applied linguistic literature and can be traced back to the theories of behaviorism and structuralism.

Behaviorists believe that reinforcing learners correct output is achieved through positive feedback provision. Simultaneously, teachers should provide corrective remarks to prevent incorrect output that may result in bad habit formation [14]. This approach resulted in the so-called structure-based approach, where "errors are frequently corrected, and accuracy tends to be given priority over meaningful interaction" [10]. The method had dominated until the advent of the communicative language teaching (CLT) approach.

In the late 1970s, it was held that educators should focus on enabling students to use language in a realistic setting. The approach was based on the comprehensible input hypothesis [15], which postulates that "we acquire by going for meaning first, and as a result, we acquire structure". Therefore, continuous error correction is not encouraged because it interrupts the communicative flow. Consequently, some researchers regarded error correction as a "serious mistake" since it makes students defensive and focuses on structure rather than meaning. Notwithstanding, it was found later that "abundant comprehensible input is not a sufficient condition for developing a near-native level of accuracy" [16]. As accuracy is the ultimate aim of most teaching methods [17], corrective feedback remains the norm followed by most teachers to achieve it.

The controversy around the efficacy of corrective feedback becomes deeper regarding WCF. Through the preceding decades, the effectiveness and practicality of WCF have remained a debatable topic [13], [16]. Primarily, researchers consider WCF as a "means of fostering learner motivation" [13] and believe that it improves accuracy [2]. Nevertheless, others consider it ineffective and even "has harmful effect" [3] as "grammar correction is a bad idea" [18]. Thus far, there is no conclusive statement about the exact effect of WCF in improving EFL learners' accuracy. However, it remains a standard and indispensable practice in the EFL classroom [6]. Moreover, its role in L2 development is "an exciting and dynamic area of investigation and, as such, is likely to continue engaging the energy and insights of established and emerging scholars" [19]. With the development of language teaching methods and techniques, WCF has witnessed new changes and thus entails new domains for inquiry.

2.2. Automated writing evaluation

A new approach that resulted from the massive advance in technology and the widespread of CALL techniques is automated writing evaluation (AWE). The central concept of WCF underlies AWE, as most of the latter applications evaluate students' writing and reflect where students make mistakes. These applications achieved their aims by comparing students' texts to existing writing corpora and measuring them

against specific rubrics that assess lexical, syntactic, and grammatical aspects [20], [21]. The rationale for introducing such applications primarily lies in the heavy workloads on teachers that might prevent them from giving accurate or sufficient feedback. Subsequently, it was believed that utilizing AWE would provide faster, cheaper and more precise scoring [22]. Of the most well-known modern tools in this strand is Grammarly. However, it should be noted that there are slight terminological issues regarding the classification of Grammarly as an AWE application.

The main issues in defining the role of AWE tools and other grammar assistants lie in the application's amount of feedback and how users can control it. Woodworth and Barkaoui [23] determined three features that distinguish AWE applications. They stated that conventional grammar assistants, such as Grammarly, "cannot be moderated by the teacher, do not evaluate writing quality, and do not include any portfolio and class management tools". Notwithstanding, many recent studies consider Grammarly an AWE tool [22], [24]. In contrast, others adopted a more precise term, i.e., automated written corrective feedback (AWCF) [5]. The present study adopted the term AWCF to avoid any possible confusion that may result from using other terms.

2.3. Grammarly for feedback

A considerable body of research investigated the potential of Grammarly in detecting students' writing and providing proper feedback. In this regard, Gavilánez and Sánchez [24] employed a pre-test/post-test experimental research design to investigate the development of 28 university students' writing during a semester of study. The participants used Grammarly and Grammark, another AI writing assistant, for AWCF. The results demonstrated a significant increase in the participant's performance in the post-test in most aspects of writing accuracy, including grammar, punctuation, mechanics, and style. The researchers traced the improvement to the student's motivation to learn independently. However, they asserted the role of the teacher in compensating for the limitations of the tools, which they found to be in content development.

Several researchers also investigated students' perceptions of Grammarly. Findings related to thereof are comparable with slight variations. O'Neill and Russell [25] found that students who employed Grammarly as a writing assistant were more satisfied with the grammar advice, they received than those mentored by human teachers. They reached this finding after implementing a mixed-methods sequential explanatory design to compare the writing of two groups. One received regular teacher instruction, and the other followed Grammarly suggestions. Nevertheless, the participants reported shortcomings related to inaccurate suggestions and skipping many errors.

Previous studies revealed several potentials of Grammarly though they are not comparable to human-based corrective feedback. Moreover, it was found that EFL students generally have favorable views toward Grammarly; however, the psychological factors that contribute to formulating the students' intention to use Grammarly are not apparent. This is an important area to investigate since it can provide significant implications for using Grammarly in teaching practice. Furthermore, it has been supported that students and researchers have reservations about the application's practicality. This conclusion implies further inquiries about the extent of Grammarly's effectiveness. Students might be interested in using Grammarly as a modern and sophisticated tool for learning, just like other new technologies. Accordingly, it is envisaged that the effectiveness of Grammarly should be further investigated to proportionate its current extensive use and to rationalize adopting it as a trustful learning tool. The present study attempts to contribute to tackling these questions.

2.4. Conceptual model

The vast advance in CALL and instructional technology necessitates a proper assessment of their impact and expediency. Intuitively, new technologies have many benefits to language learners; however, EFL learners view some tools differently based on their convenience, practicality and usefulness. Accordingly, measuring learners' acceptance of technology through different theories has become a trend. The TAM "is considered the most influential and commonly employed theory for describing an individual's acceptance of information systems" [26]. This model was proposed by Davis [11] and is used to predict end-users' acceptance of information systems. TAM achieves this by applying "scales for two specific variables, PU and PEOU which are hypothesized to be fundamental determinants of user acceptance" [11]. This principle is widely established and applied in a considerable body of research on the acceptance and adoption of technology, including learners' and teachers' use and acceptance of CALL technologies and tools [6].

Research by Davis [11] defined the two constructs of TAM. He stated that "PU is defined as the degree to which a person believes that using a particular system would enhance his or her job performance ... PEOU, in contrast, refers to the degree to which a person believes that using a particular system would be free of effort." It is envisioned that external factors, e.g., system design, determine PU and PEOU which directly influence users' attitudes towards using the system, and attitude, in turn, determines the actual

system use as shown in Figure 1. TAM was developed over the years according to continuous research and implementation. The final version of TAM [27] is displayed in Figure 2. As shown in Figure 2, the construct of attitude toward using was eliminated after finding that PU and PEO have a direct influence on behavioral intention (BI) [28] and BI, in turn, were found to be better predictors of system usage [27].

During later experiments, other researchers modified TAM, adding different variables and factors according to the nature of their studies and the technologies in question [29], [30]. In a meta-analysis study [31], the researchers reviewed 107 studies that employed TAM in e-learning. They found various external factors used by researchers and generated the modified TAM versions. The most common aspects used by the studies are self-efficacy, subjective norms, enjoyment, computer anxiety and experience. However, these factors are measured through their influence on the primary constructs of the original TAM, i.e., PU and PEOU, as these two factors “have been proven to be antecedent factors that have affected the acceptance of learning with technology” [32]. Based on the vitality of this conceptual model and the high credibility of the findings generated by applying it, the present study adopted a slightly modified TAM to measure students’ acceptance of the AWCF Grammarly application.

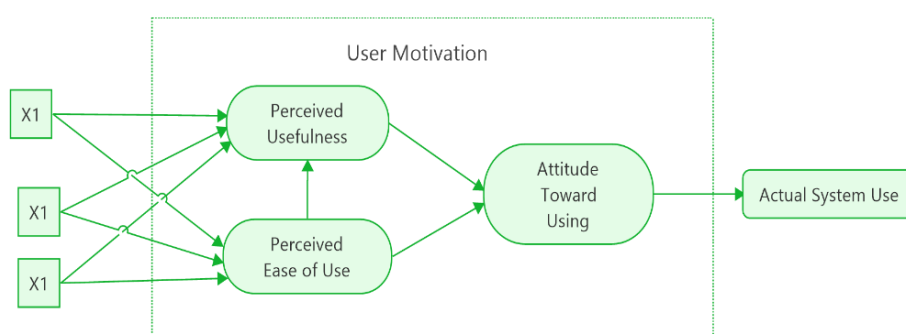


Figure 1. TAM [11]

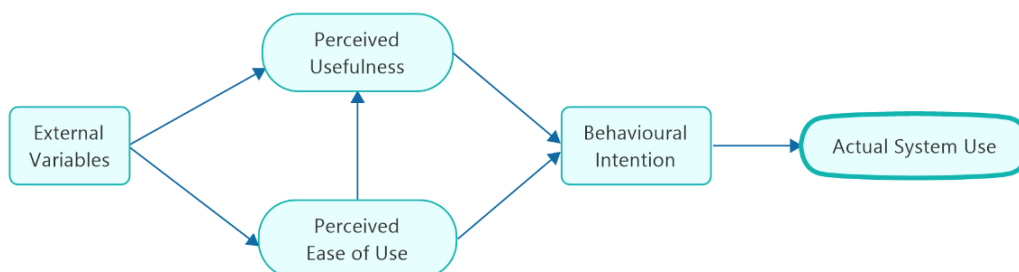


Figure 2. The final version of TAM [27]

3. RESEARCH METHOD

3.1. Design

The study adopted a quantitative analytical approach with a cross-sectional survey. According to Dörnyei [33], a cross-sectional design helps describe variables and patterns of the relationship as they exist at a specific time, especially when multivariate statistical procedures are followed. It is also preferable as we are “less exposed to the detrimental impact of unforeseen events beyond our control.” Since the primary focus of this research is on students writing mistakes and corrections, which are likely to be impacted by several extraneous factors such as learning from other courses and practice effects, it was presupposed that a cross-sectional survey would suit the study. Accordingly, a 24-item questionnaire based on a modified version of TAM was distributed to the participants.

3.2. Participants

The study sample incorporated 98 male and female undergraduates studying two courses: ENG351 (Applied Linguistics) and ENGL4760 (CALL) at the English Language and Literature Department, Prince Sattam bin Abdulaziz University in Saudi Arabia. The participants’ ages range from 20 to 23 years old, and

their L1 is Arabic. Although no available data concerning their exact proficiency levels according to the standard benchmarks, it can be said that their levels range from intermediate to upper-intermediate. By the time of the study, they had studied English as a general course for about ten years at public schools and as a major for four to six semesters, where they had studied several courses in language skills, translation, general linguistics and literature. They are familiar with modern learning technologies, especially after the pandemic lockdown in 2020. They study many online courses through learning management system (LMS), use most office applications, and are familiar with Grammarly writing assistant. Table 1 displays more information on the distribution of the participants. The researchers followed the intact class sampling method, where the whole sections of the courses were assigned to the sample. It is noted that female and level VI students are more than male and level VII ones according to the normal distribution of students in the department.

Table 1. Participants information

Level and course	Male	Female	Total
Level VI–Applied Linguistics	24	33	57
Level VII–CALL	13	28	41
Total	37	61	98

3.3. Structural model and hypotheses

The structural model adopted in this study is based on the TAM [11]. It is nevertheless modified according to specific external factors proved by previous research to be pivotal in determining technology acceptance. Ultimately, besides the PU and PEOU, the adopted structural model includes PSE and PE.

3.3.1. Self-efficacy

The concept of self-efficacy is central in human behavior and psychology. As a general concept, self-efficacy is “concerned with judgments of how well one can execute courses of action required to deal with prospective situations” [34]. It was established that “the higher the level of induced self-efficacy, the higher the performance accomplishments and the lower the emotional arousal” (ibid). In information technology literature, computer self-efficacy is related to “individuals’ control beliefs regarding his or her ability to use a system” [30]. Regarding TAM, self-efficacy is an essential determinant of PEOU as “an individual’s perception of a particular system’s ease of use is anchored to her or his general self-efficacy at all times” [27]. Accordingly, subsequent versions of TAM investigated self-efficacy extensively, making it the most external factor utilized by the modified models of TAM [31]. Based on these findings, PSE is adopted as the third construct of the present study conceptual model that inspires the first research hypothesis formulation: PSE positively affects the PEOU of Grammarly (H1).

3.3.2. Perceived enjoyment

The concept of enjoyment is related to intrinsic motivation. For some researchers, intrinsic motivation is the individual’s desire for something caused by constant enjoyment [35]. Other researchers define enjoyment as “an emotion, attitude, blend of affect and cognition, the satisfaction of intrinsic needs, and some imprecise positive reaction to the media content” [36]. The concept of intrinsic motivation in user-system interaction is associated with the PE of the user. This construct is defined as “the extent to which the activity of using a specific system is perceived to be enjoyable in its own right, aside from any performance consequences resulting from system use” [29]. Stemming from this concept, enjoyment is essential to exploring technology acceptance in different settings.

Considering eLearning settings, it is believed that intrinsically motivated activities may provide inner rewards to students and satisfy their psychological needs [35]. These activities include using online learning and digital video games [36], and implementing new technologies such as virtual reality [35]. Recent research showed that PE impacted both PU and PEOU [31], making it one of the most assessed factors to measure technology acceptance [37], [38]. Accordingly, it is adopted as the fourth construct of the present study’s conceptual model generating the following hypotheses: PE positively affects the PU of Grammarly (H2) and PE positively affects the PEOU of Grammarly (H3). The rest of the study hypotheses are inspired by the previous literature on TAM, and they are formulated: PEOU positively affects the PU of Grammarly (H4); PU positively affects the BI of using Grammarly (H5); PEOU positively affects the BI of using Grammarly (H6); BI positively affect perceived effectiveness (PEF) (H7).

The present study adopted the modified TAM model shown in Figure 3. As shown in Figure 3, four factors are posited to determine learner BI to use Grammarly and assess the tool’s effectiveness. These factors are PU, PEOU, PSE, and PE. The structure in its final representation entails that the sample size is

suitable for the research according to the 10-times rule [37]. According to this rule, the sample size should be equal to “10 times the largest number of formative indicators used to measure a single construct”.

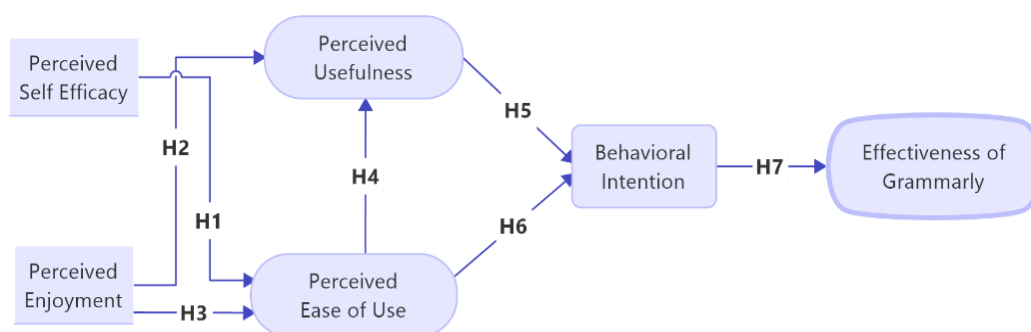


Figure 3. The proposed structural model of the study

3.4. Instrument and validation

The questionnaire was designed according to the previous literature on TAM and extended TAM models [11], [27], [31], [38]. The items' wording was built to account for the application's causal factors and nature. The final version is divided into six constructs, each containing four items designed to measure a distinct perceived factor that is believed to affect students' perception of Grammarly. Three content, eLearning experts, and educational psychology checked the survey items. They suggested minor modifications to the wording and the ordering of the items. Further, the questionnaire was piloted on 24 students and faculty members to determine the items' appropriateness and understandability. Based on the feedback from the experts and the respondents, the final copy of the questionnaire was refined. Furthermore, factor analysis was conducted after the data was gathered to assess the instrument's convergent validity and internal consistency.

3.5. Data collection

Although the nature of the two courses is related to the concept of AWCF, the participants were informed that participating in the research is optional and independent of any course assessment. From the beginning of the semester, they were asked to use Grammarly to check and proofread their writing in all the departmental courses and other writing work. Technical support and follow-up were provided for those students who encountered problems in installing, running or using the app, though those were rare cases. After granting the necessary participants' and administrative consent, the questionnaire was translated into Arabic. Two professors in Arabic language and linguistics checked the translated version for naturalness and correctness who first compared it to the original English version. The final version was then published through the Google Forms tool, and its link was sent to students through blackboard LMS. The respondents were asked to report their level of agreement with the questionnaire items on a 5-point Likert scale ranging from Strongly agree to Strongly disagree.

3.6. Data analysis

The study adopted the structural equation model (SEM) to test the hypotheses since the model can work appropriately with latent indicators and various constructs. Following the study aims, the research sample size, and the nature of the causal relations between the constructs, the partial least square (PLS) is considered suitable for analyzing the research data. Accordingly, the researchers used SmartPLS 3.0 software. The data analysis process incorporated two stages. In the first stage, a confirmatory factor analysis was computed to assess the convergent validity of the structure. Then, bootstrapping was conducted to evaluate the structural model and test the research hypotheses.

4. RESULTS AND DISCUSSION

4.1. Factor analysis

The adopted structural model incorporated six constructs: PSE, PE, PU, PEOU, BI, and PEF. To assess the structure's convergent validity, i.e., the relationship between the structures and their markers, the researcher computed the factor loading, composite reliability (CR), and the average variance extracted (AVE). The results are displayed in Table 2.

Table 2. Convergent validity of the structural model

Construct	Item	Factor loading	CR	AVE
PSE	PSE1	0.871	0.936	0.786
	PSE2	0.923		
	PSE3	0.907		
	PSE4	0.844		
PE	PE1	0.843	0.904	0.702
	PE2	0.791		
	PE3	0.886		
	PE4	0.829		
PU	PU1	0.757	0.891	0.672
	PU2	0.849		
	PU3	0.789		
	PU4	0.879		
PEOU	PEOU1	0.706	0.853	0.592
	PEOU2	0.763		
	PEOU3	0.862		
	PEOU4	0.738		
BI	BI1	0.819	0.871	0.629
	BI2	0.758		
	BI3	0.767		
	BI4	0.826		
PEF	PEF1	0.783	0.876	0.639
	PEF2	0.880		
	PEF3	0.805		
	PEF4	0.722		

Considering that the optimal value for factor loading should be equal to or greater than 0.708 to maintain convergent validity [37], all the structure items are counted as convergently valid. In other words, each construct's latent variables have much in common and explain a substantial part of each indicator's variance. Likewise, the reliability values are satisfactory since they fall within the margin between 0.70 and 0.95. Furthermore, the AVE values are reasonable since they are greater than the acceptable value of 0.50, indicating that each construct can explain more than half of the variance of its indicators [39].

4.2. Structural model analysis

To determine the predictive power of the constructs, the R^2 values are calculated across the generated structure model. Figure 4 depicts the path coefficients and the relative contribution of each indicator to the construct. R^2 values between 0.19–0.33 are considered (Small), between 0.33–0.67 normal, and <0.67 are regarded large [40].

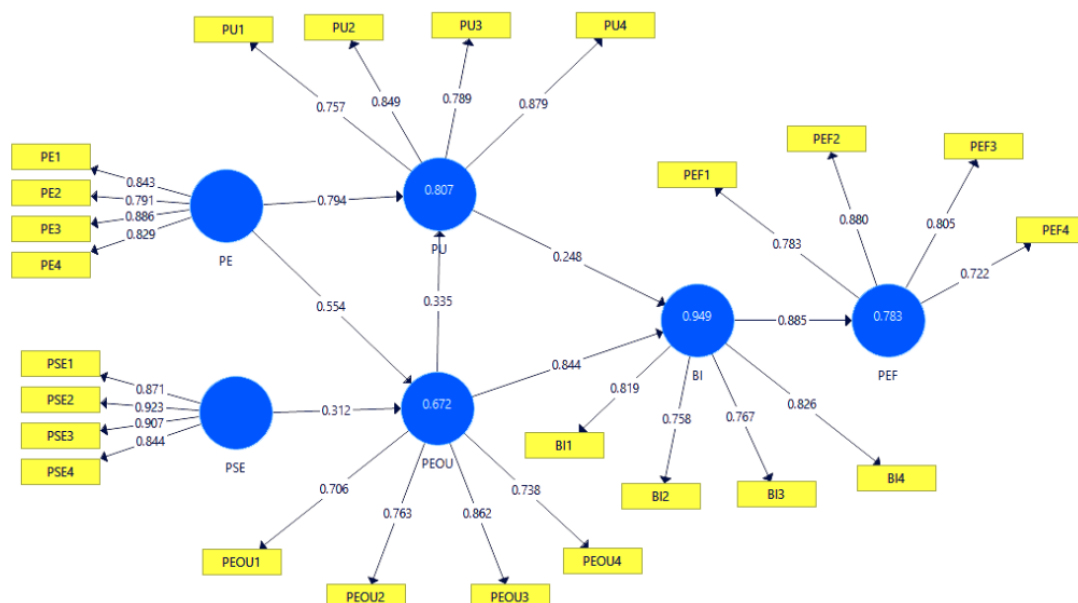


Figure 4. The structural model and path coefficients

There are four exogenous constructs in the structural model which are PEOU ($R^2=0.67$), PU ($R^2=0.81$), BI ($R^2=0.95$), and PEF ($R^2=0.78$). According to the rubric [40], all the values are large, suggesting that all four constructs have significant predictive power. In other words, it can be paraphrased that the two constructs, PE and PSE, explain 67.1% of the variance in PEOU as the least predictive power in the model. PEOU and PE explain 80% of the variance in PU. On the other hand, the two constructs, PEOU and PU, have the strongest predictive power as they can predict 94.9% of the variance in BI, while BI alone can predict 78.3% of the variance in PEF.

4.3. Hypothesis testing

To test the associations between the constructs, a bootstrapping was conducted to generate the T-values and P-values to measure the significance of the association. The results are displayed in Table 3. The results show that all the hypotheses were supported, providing that T-values for all the hypothesized relations are >0.1645 and the P-values are >0.01 [37]. However, the most substantial positive relationship was found between BI and PEF as it has the closest value to +1, while PU to BI has the weakest relation as it is the closest value to zero.

Table 3. Hypothesis testing results

H	Causal effect	Path coefficient (β)	T-value	P-value	Result
H ₁	PSE→PEOU	0.312	3.252	0.001	Supported
H ₂	PE→PEOU	0.554	6.445	0.000	Supported
H ₃	PE→PU	0.794	8.594	0.000	Supported
H ₄	PEOU→PU	0.335	4.466	0.000	Supported
H ₅	PU→BI	0.248	6.053	0.000	Supported
H ₆	PEOU→BI	0.844	19.010	0.000	Supported
H ₇	BI→PEF	0.885	46.690	0.000	Supported

Significant at ($\alpha>0.01$)

4.4. Discussion

The main motive of the present study was to assess the psychological factors that contribute to formulating the students' intention to use the AWCF application Grammarly and evaluate its effectiveness. The research adopted a modified version of TAM [11]. In addition to the essential constructs: PEOU and PU, two proposed factors were postulated, i.e., PSE and PE. A questionnaire based on the model was distributed to 98 students who had used the writing assistant for a semester of study. The results revealed that all the factors strongly predict the variance in the corresponding constructs. Moreover, all the hypothesized factors positively affected their endogenous constructs with varying levels.

Regarding self-efficacy, the results suggest that it slightly affects the ease of use as perceived by students. This result is in line with previous literature that reported a significant correlation between PSE and PEOU [27], PSE and PU [41], and PSE and BI [42]. Contrary to previous literature and other causal relations in this study, the effect of PSE on PEOU is low. This finding can be traced back to the nature of Grammarly and AI writing assistants.

As there are certain deficiencies in the suggestions provided by Grammarly, students are likely to be uncertain regarding their ability to exploit the entire program's potential. For example, most of the student's responses to the third item in this construct which reads: "I can apply Grammarly suggestions to improve my writing quality," were negative. Although their answers are more likely caused by an academic reason (not understanding the grammatical point in question or why Grammarly corrected it this way), it ultimately affected students' views regarding their PSE. The formulation of this item can also be a reason for the relatively low mean score of the effect of PSE on PEOU. This can be considered a limitation of this study that should be avoided in future research. In other words, PSE should be measured as computer PSE. Hence, questions related to the construct should investigate student PSE about technical aspects of the software rather than academic or contextual aspects.

On the other hand, enjoyment is proved by the present study to have a stronger relationship to both ease of use and usefulness. Again, this result is identical to previous literature [35], [36]. Moreover, it is proved that PU is more determined by PE than PEOU, with a path coefficient of 0.794 for the former and 0.335 for the latter. This finding implied that students' view of Grammarly's usefulness stems from their enjoyment while using the application. As far as the PEOU is concerned, it is also determined by PE as a strong predictor of students' view of the application's easiness.

It may be argued that the causal relation between PE and PEOU is likely to be PEOU→PE rather than PE→PEOU as intrinsic motivation is probable to be developed by a program easiness. This argument is intuitive and supported by previous studies that adopted motivational models; however, research that

followed TAM usually stands for the later causal relation, as PEOU is used as a construct that is affected by external factors and affects PU and BI. In other words, students' enjoyment while using Grammarly help them ignore possible hurdles they might face while using the application. This supported the results of similar studies that found enjoyment a robust predictive factor for ease of use by students [38] and teachers [43] who use technology for educational purposes.

The remaining findings of the study correlate with previous studies that utilized TAM. PEOU is a predictor of PU and BI [11]. However, PU is found to predict BI with an unexpected minor path coefficient value, i.e. 0.248, as the lowest value in the whole structure model. Again, the nature of the software and the observed shortcomings in its feedback may cause this view. Students seem conservative about the ultimate usefulness of the tools because they might think that many feedbacks were either wrong or unintelligible. This result implies that students' BI of using Grammarly stems from their view of its easiness more than its usefulness. Nevertheless, the two factors, PU and PEOU, were found to have a strong predictive power of variance in BI that reaches 94.9%, which make the ultimate findings also consistent with previous literature on TAM, especially [27] that suggested the last modification of the model to include BI of using technology rather than attitude to use it.

Likewise, the participants' views about the effectiveness of Grammarly are also determined by their BI to use the tool, as suggested by almost all previous studies on TAM. BI was found capable of explaining 78.3% of the variance of the construct PEF with the highest path coefficient value in the whole model, i.e. 0.885. According to H₇ of this research, the effectiveness of Grammarly is represented by students' decision to use the program based on other perspectives, including their PSE, PE, PEOU, and PU. In other words, after using Grammarly for a semester of study, the participants supposed that it is an enjoyable, easy-to-use, and helpful learning tool. Therefore, they develop an intention to use it. Ultimately, this intention to use formulated their perception of the effectiveness of Grammarly.

Overall, the present study's result supported the TAM theory with minor variations in how each construct relates to the other. This slight discrepancy can be traced back to the nature of the application studied. After all, AWCF is still a new trend in education. It has some shortcomings related to its inaccuracy in detecting some errors, inconsistency of some corrective suggestions, and inability to evaluate all aspects of writing. These factors may make students more conservative regarding their views about the tool's usefulness and arouse doubts regarding their ability to get the most out of the technology, i.e. their PSE. Accordingly, it is implied that Grammarly and similar software should be used as an accompanying learning tool to support teachers' corrective feedback rather than fundamental. Incorporating activities that utilize Grammarly to develop students' writing outside the classroom may be an appropriate suggestion to achieve this.

The generalizability of the results is limited by the relatively small sample size of the research. Further research should adopt larger samples to support the findings. Moreover, since the Grammarly-related activities conducted by the participants were not related to coursework, students may be less motivated to complete them appropriately. To account for this limitation, future research can utilize experimental methods that measure the students' actual performance in fundamental writing coursework and investigate students' perceptions of the tool and whether it affects their language learning process.

5. CONCLUSION

The development of CALL technology and tools will likely affect students' attitudes towards language learning. Moreover, students' acceptance of technology determines their implementation and hence reflects the effectiveness of such a technology. One of the relatively modern applications of this kind is Grammarly, a well know writing assistant that exemplifies AWCF tools. Grammarly has been heavily studied in the previous decade. However, there is a relative paucity of research investigating its effectiveness as perceived by the students. This is an indispensable aspect of research as responding to it can provide practical implications for employing AWCF tools in language learning. Accordingly, the present study investigated students' perceptions of the effectiveness of Grammarly utilizing a modified TAM.

The study revealed that students viewed Grammarly as a practical learning application. This perception is based mainly on their perception that it is an enjoyable, easy-to-use, and helpful learning tool. The results revealed that the participants had developed a considerable intention to use Grammarly for their future writing in both computer and mobile settings. This implies that they believe it is an efficient learning tool that may positively affect their language learning process. The relatively low level of agreement with some constructs of the structural model suggests that students are cautious regarding their views about the tool's usefulness and easiness. These results imply using Grammarly as a learning assistant, not a primary teaching or learning tool. Students also showed enthusiasm regarding using Grammarly, which calls for incorporating it and similar tools in different language learning activities.

Future research should utilize a larger sample to eliminate possible limitations that might be present in the current study. These studies should also be conducted longitudinally with frequent checkpoints to investigate the effect of incorporating AWCF on language learning. Moreover, experimental research is more suitable for measuring the impact of AWCF tools on writing performance, especially when control groups are used. A suggested area for future research would be investigating what features of writing can benefit from the application of such technology and to what extent.

ACKNOWLEDGEMENTS

This study is supported via funding from Prince Sattam bin Abdulaziz University project number (PSAU/2023/R/1444).




REFERENCES

- [1] J. Burstein, M. Chodorow, and C. Leacock, "Automated essay evaluation: the criterion online writing service," *AI Magazine*, vol. 25, no. 3, pp. 27–36, 2004, doi: 10.1609/aimag.v25i3.1774.
- [2] J. Bitchener, "Evidence in support of written corrective feedback," *Journal of Second Language Writing*, vol. 17, no. 2, pp. 102–118, Jun. 2008, doi: 10.1016/j.jslw.2007.11.004.
- [3] J. Truscott, "The case against grammar correction in L2 writing classes," *Language Learning*, vol. 46, no. 2, pp. 327–369, Jun. 1996, doi: 10.1111/j.1467-1770.1996.tb01238.x.
- [4] L. M. Rudner and T. Liang, "Automated essay scoring using Bayes' theorem," *Journal of Technology, Learning, and Assessment*, vol. 1, no. 2, pp. 1–22, 2002.
- [5] J. Ranalli, "Automated written corrective feedback: how well can students make use of it?" *Computer Assisted Language Learning*, vol. 31, no. 7, pp. 653–674, 2018, doi: 10.1080/09588221.2018.1428994.
- [6] Y. Luo and Y. Liu, "Comparison between peer feedback and automated feedback in college English writing: a case study," *Open Journal of Modern Linguistics*, vol. 07, no. 04, pp. 197–215, 2017, doi: 10.4236/ojml.2017.74015.
- [7] G. Dizon and J. Gayed, "Examining the impact of Grammarly on the quality of mobile L2 writing," *The JALT CALL Journal*, vol. 17, no. 2, pp. 74–92, Aug. 2021, doi: 10.29140/jaltcall.v17n2.336.
- [8] A. Qassemzadeh and H. Soleimani, "The impact of feedback provision by Grammarly software and teachers on learning passive structures by Iranian EFL learners," *Theory and Practice in Language Studies*, vol. 6, no. 9, p. 1884, Sep. 2016, doi: 10.17507/tpls.0609.23.
- [9] K. Beaty, "Computers in the language classroom," in *Practical English language teaching*, McGraw-Hill, 2003, pp. 247–267.
- [10] P. Lightbown and N. Spada, *How languages are learned*, 4th ed. Oxford University Press, 2013.
- [11] F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," *MIS Quarterly: Management Information Systems*, vol. 13, no. 3, pp. 319–340, 1989, doi: 10.2307/249008.
- [12] C. Chaudron, "A descriptive model of discourse in the corrective treatment of learners' errors," *Language Learning*, vol. 27, no. 1, pp. 29–46, Jan. 1977, doi: 10.1111/j.1467-1770.1977.tb00290.x.
- [13] R. Ellis, "Corrective feedback and teacher development," *L2 Journal*, vol. 1, no. 1, Apr. 2009, doi: 10.5070/L2.V1I1.9054.
- [14] M. Demirezen, "Behaviorist theory and language learning," *Journal of Hacettepe University Faculty of Education*, vol. 3, no. 3, pp. 135–140, 1988.
- [15] K. J. Krahnke and S. D. Krashen, *Principles and practice in second language acquisition*. Pergamon Press Inc, 1982, doi: 10.2307/3586656.
- [16] C. van Beuningen, "Corrective feedback in L2 writing: theoretical perspectives, empirical insights, and future directions," *International Journal of English Studies*, vol. 10, no. 2, p. 1, Dec. 2010, doi: 10.6018/ijes/2010/2/119171.
- [17] Q. Huang, "Accuracy & fluency: inspiration from error-correction of interlanguage theory," *Asian Social Science*, vol. 5, no. 2, Feb. 2009, doi: 10.5539/ass.v5n2p84.
- [18] J. Truscott, "The case for 'the case against grammar correction in L2 writing classes': a response to Ferris," *Journal of Second Language Writing*, vol. 8, no. 2, pp. 111–122, May 1999, doi: 10.1016/S1060-3743(99)80124-6.
- [19] J. Bitchener and D. R. Ferris, "Written corrective feedback in second language acquisition and writing," *Written Corrective Feedback in Second Language Acquisition and Writing*, pp. 1–232, 2012, doi: 10.4324/9780203832400.
- [20] N. Elliot, A. Gere, G. Gibson, C. Toth, C. W. And, and A. Presswood, "Uses and limitations of automated writing evaluation software," *WPA-CompPile Research Bibliographies*, vol. 73, pp. 1–26, 2013.
- [21] N. Hockly, "Automated writing evaluation," *ELT Journal*, vol. 73, no. 1, pp. 82–88, Jan. 2019, doi: 10.1093/elt/ccy044.
- [22] D. Miranty and U. Widiati, "Automated writing evaluation (AWE) in higher education: Indonesian EFL students' perceptions about Grammarly use across student cohorts," *Pegem Egitim ve Ogretim Dergisi*, vol. 11, no. 4, pp. 126–137, Oct. 2021, doi: 10.47750/pegegog.11.04.12.
- [23] J. Woodworth and K. Barkaoui, "Perspectives on using automated writing evaluation systems to provide written corrective feedback in the ESL classroom," *TESL Canada Journal*, vol. 37, no. 2, pp. 234–247, Dec. 2020, doi: 10.18806/tesl.v37i2.1340.
- [24] L. P. Gavilán and X. C. Sánchez, "Automated writing evaluation tools in the improvement of the writing skill," *International Journal of Instruction*, vol. 12, no. 2, pp. 209–226, 2019, doi: 10.29333/iji.2019.12214a.
- [25] R. O'Neill and A. Russell, "Stop! Grammar time: University students' perceptions of the automated feedback program Grammarly," *Australasian Journal of Educational Technology*, vol. 35, no. 1, Mar. 2019, doi: 10.14742/ajet.3795.
- [26] Y. Lee, K. A. Kozar, and K. R. T. Larsen, "The technology acceptance model: past, present, and future," *Communications of the Association for Information Systems*, vol. 12, no. 1, p. 50, 2003, doi: 10.17705/ICAIS.01250.
- [27] V. Venkatesh and F. D. Davis, "A model of the antecedents of perceived ease of use: development and test," *Decision Sciences*, vol. 27, no. 3, pp. 451–481, Sep. 1996, doi: 10.1111/j.1540-5915.1996.tb01822.x.
- [28] P. Lai, "The literature review of technology adoption models and theories for the novelty technology," *Journal of Information Systems and Technology Management*, vol. 14, no. 1, pp. 21–38, Jun. 2017, doi: 10.4301/S1807-17752017000100002.
- [29] V. Venkatesh, "Determinants of perceived ease of use: integrating control, intrinsic motivation, and emotion into the technology acceptance model," *Information Systems Research*, vol. 11, no. 4, pp. 342–365, Dec. 2000, doi: 10.1287/isre.11.4.342.11872.




- [30] V. Venkatesh and H. Bala, "Technology acceptance model 3 and a research agenda on interventions," *Decision Sciences*, vol. 39, no. 2, pp. 273–315, 2008, doi: 10.1111/j.1540-5915.2008.00192.x.
- [31] F. Abdullah and R. Ward, "Developing a general extended technology acceptance model for e-learning (GETAMEL) by analysing commonly used external factors," *Computers in Human Behavior*, vol. 56, pp. 238–256, Mar. 2016, doi: 10.1016/j.chb.2015.11.036.
- [32] A. Granić and N. Marangunić, "Technology acceptance model in educational context: a systematic literature review," *British Journal of Educational Technology*, vol. 50, no. 5, pp. 2572–2593, Sep. 2019, doi: 10.1111/bjet.12864.
- [33] Z. Dörnyei, *Research design: qualitative, quantitative, and mixed methods approaches*, Oxford University Press, 2007.
- [34] A. Bandura, "Self-efficacy mechanism in human agency," *American Psychologist*, vol. 37, no. 2, pp. 122–147, Feb. 1982, doi: 10.1037/0003-066X.37.2.122.
- [35] Y.-J. Lin and H. Wang, "Using virtual reality to facilitate learners' creative self-efficacy and intrinsic motivation in an EFL classroom," *Education and Information Technologies*, vol. 26, no. 4, pp. 4487–4505, Jul. 2021, doi: 10.1007/s10639-021-10472-9.
- [36] M. Ebrahimzadeh and S. Alavi, "Motivating EFL students: e-learning enjoyment as a predictor of vocabulary learning through digital video games," *Cogent Education*, vol. 3, no. 1, p. 1255400, Dec. 2016, doi: 10.1080/2331186X.2016.1255400.
- [37] T. Teo and J. Noyes, "An assessment of the influence of perceived enjoyment and attitude on the intention to use technology among pre-service teachers: a structural equation modeling approach," *Computers & Education*, vol. 57, no. 2, pp. 1645–1653, Sep. 2011, doi: 10.1016/j.compedu.2011.03.002.
- [38] L. Zhou, S. Xue, and R. Li, "Extending the technology acceptance model to explore students' intention to use an online education platform at a University in China," *SAGE Open*, vol. 12, no. 1, pp. 1–13, Jan. 2022, doi: 10.1177/21582440221085259.
- [39] J. F. Hair, G. T. M. Hult, C. M. Ringle, and M. Sarstedt, *A primer on partial least squares structural equation modeling (PLS-SEM)*, SAGE Publications, 2014.
- [40] K. K.-K. Wang, "Partial least square structural equation modeling (PLS-SEM) techniques using SmartPLS," *Marketing Bulletin*, vol. 24, no. 1, pp. 1–32, 2013.
- [41] W. Chin, "Partial least squares approach to structural equation modeling for tourism research," in *Modern methods for business research*, 1998, pp. 295–336.
- [42] Y.-H. Lee, C. Hsiao, and S. H. Purnomo, "An empirical examination of individual and system characteristics on enhancing e-learning acceptance," *Australasian Journal of Educational Technology*, vol. 30, no. 5, Nov. 2014, doi: 10.14742/ajet.381.
- [43] M. Gong, Y. Xu, and Y. Yu, "An enhanced technology acceptance model for web-based learning," *Journal of Information Systems Education*, vol. 15, no. 4, pp. 365–374, 2004.

BIOGRAPHIES OF AUTHORS



Abdulaziz B. Sanosi    Studied English and linguistics at the University of Khartoum, Sudan. He got his PhD in English applied linguistics from Omdurman Islamic University, Sudan. Currently, he is a lecturer at the department of English language at Prince Sattam bin Abdulaziz University in Saudi Arabia. He teaches applied and theoretical linguistics courses. His research interests include corpus linguistics, discourse analysis, and CALL. He is a compiler and cofounder of several learner and user corpora designed to explore the English language applications in different registers, such as academic writing and social media. He can be contacted at email: a.assanosi@psau.edu.sa.



Mohammed Omar Musa Mohammed    has PhD in Applied Statistics from university of Kwazulu Natal in South Africa. His main research interests are in analyzing complex survey data, the biological health sciences, particularly modelling population and disease dynamics. At the University of Alneelain and now prince Sattam Bin Abdulaziz university, He has taught theoretical and applied courses, including Biostatistics courses at both undergraduate and graduate levels covering key areas in biostatistics, namely general epidemiology principles, cohort studies, case-control studies, survival analysis and clinical trials. He can be contacted at email: moh.mohammed@psau.edu.sa.