Native and non-native speakers' reaction to Grammarly-flagged errors: Implications for L2 teaching of writing

Khaled El Ebyary

(khaled.elebyary@york.ac.uk)
University of York, UK

Abstract

The now ubiquitous use of advanced Web 2.0 tools in writing and the emergence of automated error flagging applications with affordances far beyond Word Processing requires some attention from both L2 researchers and L2 tutors, especially when both native (skilled) writers and non-native (less skilled) writers have, reportedly, started to use various commercial and freemium technological tools that claim to provide automated written corrective feedback. In fact, little is known about tracking writers' editing behaviour when automated error flagging is in place and whether such behaviour would vary between native and non-native writers. Using a pre-post activity interview, an IELTS writing task 2 and screen capture software, the current case study compared the editing behaviours of native and non-native speakers of English when Grammarly was used. Major results revealed that native speakers had overall more flagged errors than non-native speakers did, but the latter group had more grammar errors flagged. However, the two groups followed a similar pattern in reacting to the flagged errors. Both native and non-native writers accepted suggestions from Grammarly. The study also suggests that evidence is needed with regard to teachers' roles in and learners' uptake from error flagging applications.

1 Introduction

As the writing ability forms an important component of any language learning programme in almost any EFL context, improving writing skills is a central objective for second language (L2) teachers, as well as researchers. This is described by Casanave (2004, p. 64) as "*the most consuming of all dilemmas for L2 writing teachers*". The 1980's and 1990's witnessed a plethora of research that focused on understanding the complexities of the composing act (e.g.,Bereiter & Scarmalia, 1987; Cooper & Matsushashi, 1983; Flower & Hayes, 1981). As an adjunct of such research, many studies examined students' errors, which have always been described as inevitable. Unlike mistakes

that can be identified as lapses attributed to the lack of enough attention, errors are consistent features of a learner's production (Ellis, 1997). Research has tried to understand the nature of errors, to identify pedagogical practice that can help teachers, peers or computers to respond to such errors (i.e., provide written corrective feedback) and to improve L2 writers' abilities to eliminate their own errors by doing successful revisions. The importance of identifying errors and their sources has been emphasized in the literature (Delima, 2019). Feedback in its own right has been described as an element of most L2 theories and language pedagogy (Ellis, 2009); and written corrective feedback (CF) has been described by Bitchener and Storch (2016) as follows:

> a written response to a linguistic error that has been made in the writing of a text by an L2 learner. It seeks to either correct the inaccurate usage or provide information about where the error has occurred and/or about the cause of the error and how it may be corrected (p.12).

Indeed, several researchers showed interest in carrying out meta-analysis studies which are generally in support of CF practice (e.g.,Brown, 2014; Li, 2010; Nassaji & Kartchava, 2019; Norris & Ortega, 2000). Equally important is the recognition of the significant contribution of grammar to learners' writing development (Polio, 2012; Spada, 2018).

Over the last few decades, a wide range of technological applications that claim to provide automated corrective feedback (ACF) or automated written corrective feedback (AWCF) on written texts have been introduced and it is not uncommon that the use of such technological tools is described as ubiquitous. Thirty years ago, Hill, Wallace and Haas (1991) anticipated the role of technology in writing instruction to be '*increasingly prevalent*' and described educators as '*hold[ing] high expectations for the ways in which this technology can support writing and the teaching of writing*' (p.83). Twenty years after Hill et al.'s (1991) accounts, Relles and Tierney (2013) stated that the incorporation of technology in the academic culture advocates that the '*writing habits*' of learners will soon be '*navigational across myriad discourse situations that do and will yet exist*' (p.501). This suggests that Hill et al. (1991) were rather optimistic in their prediction of AWCF becoming much more widespread in L2 writing practice. More recently, claims of extensive use of AWCF tools have been maintained by various researchers (Guo et al., 2021b; Weigle & Malone, 2016). This study suggests that whilst L2 writing researchers in the area of written corrective feedback effected to inform EFL classroom instruction, there is a difference in the perspectives of EFL teachers and students. While teachers, particularly those doing face-to-face tutoring, are keen to enable students to write without the support of web tools because this is what is expected, many students are less tolerant of imperfect language output and are reportedly using various commercial and freemium technological tools that claim to provide AWCF. This finds support in Ferris's (2004) statement that students' constant desire for feedback to improve their written output is not dismissible. Furthermore, the ubiquitous technology use has given L2 learners maximum control over which AWCF tools they might use, especially with the affordances found in a wide range of applications. The wide variety includes tools that can log all, some, or no information about how individual writers react to comments. As an AWCF tool, Grammarly (https://www.grammarly.com/) has been described as popular (Zhang et al., 2020), powerful (Barrot, 2020) and extensively used (Guo et al., 2021b). Barrot (2020) defines Grammarly as a tool which "*identifies duplicate content and errors in grammar, vocabulary, mechanics, and language style*" (p.33).

The current study suggests that L2 composition research has shown interest in identifying the similarities between the processes of skilled L1 and less skilled L2 writers so that these processes can be taught in L2 classrooms. More recently, many students are reportedly using Grammarly to obtain automated feedback on their written texts. Equally, Grammarly designers advertise aggressively and seem to put money and effort into advertising across multiple Media formats that target skilled (native speaker) and less skilled (non-native speaker) writers. An early survey conducted by Grammarly designers suggested that 68% of Grammarly users are native speakers compared to 32% non-natives (Grammarly, 2012). However, little is known about how skilled (native speaker) and less skilled (non-native speaker) writers react to the flagged errors and whether L2 students' editing

behaviours are similar/different to native speakers. The current study examines the difference, if any, between the editing behaviours of natives and non-natives when the full version of Grammarly is used. This could help in understanding the effect of using Grammarly as an error flagging tools on writing and would particularly have implications for L2 teaching of writing.

2 Literature review

2.1 *Automated writing evaluation*

With the pervasive use of advanced technology, Li, Dursun and Hegelheimer (2017) identified three major categories of technological applications in L2 writing, one of which is Automated Writing Evaluation (AWE). Some researchers describe the recent widespread use of AWE as based on the belief that such applications allow teacher feedback to focus more on higher-level writing skills while the computer would target lower-level errors (Link et al., 2020; Wilson & Czik, 2016). AWE research has extended over the last few decades; however, the focus has mainly been on the famous commercial tools such as E-rater (e.g.,Attali, 2004; El Ebyary & Windeatt, 2010; Li et al., 2015), MyAccess (e.g.,Hoang & Kunnan, 2016; Rudner et al., 2006) WriteToLearn (e.g.,Liu & Kunnan, 2016) and others, which are normally purchased by educational institutions for instructional and/or research purposes. Many of the studies that dealt with commercial AWE focused on comparing the computerized feedback and scoring with human raters by examining the reliability of each of these systems or comparing the computerized feedback with other forms such as teacher feedback (Attali & Burstein, 2004; Dikli & Bleyle, 2014; Enright & Quinlan, 2010). Other researchers focused on the revision act when these commercial applications are used (e.g.,Link et al., 2020). Portals used in commercial packages often log writing data (e.g., word count, time spent in writing, analytical feedback, holistic scores…etc.) for instructors. However, A clear line can be drawn in relation to the available tools in this area between famous commercial AWE tools and almost free error flagging applications such as Grammarly, which provides real time feedback in which errors are underlined *while* users are producing the text. Although this might be seen by some as restricting the instructional value of such application, Grammarly's widespread use by students demonstrates its worth. In fact, Grammarly has drawn the attention of various researchers (e.g.,Barrot, 2021; Dembsey, 2017; Guo et al., 2021b; Koltovskaia, 2020; O'Neill & Russell, 2019; Qassemzadeh & Soleimani, 2016; Ventayen & Orlanda-Ventayen, 2018). The current study synonymously uses the term automated error flagging applications and automated written corrective feedback (AWCF) which has been used in some studies (Guo et al., 2021a; Ranalli, 2018).

2.2 *Grammarly*

Although Grammarly's official webpage (https://www.grammarly.com/) does not offer a definition as such, it provides an explanation of what the system can do and why it is worth using. According to its designers' webpage, "*Grammarly automatically detects grammar, spelling, punctuation, word choice and style mistakes*". While users can type (or cut and paste) their text directly into the Grammarly portal, the application can work from within other platforms that involve text production (e.g., email, social media, Microsoft Word…etc.). Here Barrot (2020) describes Grammarly availability as a web application, a web browser extension, a Microsoft Word extension or as a desktop tool. She further explains the role of Grammarly in the editing stage as providing "*real-time feedback*" (Barrot, 2020, p. 2). Table 1 explains categories of errors and language focus. Errors flagged are colour-coded as explained in Table 1. Figure 1 provides an example from the data obtained in this study.

Table 1.

Grammarly Colour-Coded Feedback

Error Category	Focus	Colour Coding
Correctness and Writing Mechanics	Spelling, grammar, and punctuation	Red
Clarity, Conciseness, and Readability	Wordy, overly complex sentences	Blue
Engagement (Vocabulary & Variety)	Word choice, repetitive sentences, monotonous passages	Green
Delivery (Formality, Politeness, and Confidence)	appropriate tone and attitude (e.g. friendliness and professionalism, choosing words that convey the right tone)	Purple

Figure 1.

Examples of Colour-coded Feedback in Grammarly

the description that team sports is more beneficial for people, which gains many supports from the majority of the public. to be specific, when most people think of sports, they probably think of some kind of team sports, such as football,basketball and volleyball. these are by far the most popu;ar. so, we often watch them on TV or play with our friends. in fact, they often help us develop essential life skills, especially for children , to participate in team sports is valuable way to develop their interpersonal skills and team work can help them to better connect with others, because they will be able to communicate well and delegate responsibility for these skills will be in the age of teenagers and adults for their professional and personal life to provide good service. expect of training tean spirit, the advantages of promoting more and more people to join in the soprts cannot be ignored.

- team work · Correct your spelling
- them to · Remove the space
- good · Choose a different word
- in fact, they often help us devel... · Rewrite the sentence
- expect · Capitalize the word
- tean · Correct your spelling
- be ignored · Rewrite the sentence

3   Methodology

The present case study involved 6 participants of whom 3 were native speakers from the UK and 3 were Chinese. The participants' age ranged from 22-25 and all non-native participants had a score of 7 overall in their IELTS tests. All participants were students at a UK university at the time this study was conducted and they were all Grammarly users. All non-native participants described their previous language education contexts as examination-focused, which finds support in various studies involving international students (Miaoa et al., 2006). This case study compared native participants (i.e., Participants 1, 2 and 3) and non-native participants (i.e., Participants 4, 5 and 6) who were required to write 400-500 word compositions on a topic. For the purpose of authenticity, a writing task (i.e., a prompt) was randomly selected from the IELTS writing task 2 assigned in January and February 2018. Pre and post activity semi-structured interviews were used un this study. The pre-activity interview collected data about participants' perceptions and use of automated feedback applications in general and how they perceived feedback from Grammarly in specific. Additionally, participants in the pre-activity interview were individually asked about familiarity with the topic involved in the writing task and the level of difficulty, and they all thought the topic was familiar and not difficult. Data elicited from the pre-activity interview was thematically analysed. The post activity interview focused on each participant's revision behaviour while automated feedback was being provided. These interviews were similar to stimulated recall in the sense that they involved segments of recorded videos. However, the focus was more on giving the participants the opportunity to explain their perceptions and revision behaviour. Grammarly was used in their browsers in which errors were flagged on the right-hand side of the page. MS Word grammar and spelling

checkers were disabled. Participants were free to *adopt*, *reject* or even *avoid* suggestions from Grammarly. This simplified coding was based on the scheme used by Chapelle, Cotos and Lee (2015) which included six categories (i.e., no change, remove, add, delete, change and transpose). Although the small number of participants forms a limitation on this study, it is not intended to make generalisable quantitative claims. This small number of participants made for usefully detailed qualitative analysis collected from the writing sessions which were video recorded using the screen capture software Screen-O-Matic https://screencast-o-matic.com/). The recording of the writing sessions with the relevant timestamps enabled analysis of participants' editing behaviour when AWCF was provided. A total of 112 minutes of actual writing time was analyzed. Error flagged by Grammarly and participants' behaviours were the focus of such observational data analysis.

4   Findings

4.1   Participants' experiences with error-flagging tools

As a departure point in this study, it was useful to understand participants' experiences with error flagging tools in general and Grammarly in specific. In terms of participants' experiences with error flagging software, this research aimed to explore whether knowledge and use of such applications varied between native and non-native participants. In other words, although this study attempted to focus on native and non-native participants' editing behaviour when they received AWCF from Grammarly, the study tried to initially find out how they perceive the feedback from Grammarly and what other applications they might be using. Data from the pre-activity interview revealed that Grammarly is the most commonly known/used application among participants. Further analysis of interview data examined whether knowledge, and use, of the application varied from one group to another.

Participants 1 was a native speaker who uses Grammarly regularly when she writes an assignment. This participant stated,

> *I am confident about my writing, but sometimes I use language that seems perfectly natural to me, and I then figure out from Grammarly it is ungrammatical.*

Another comment from Participant 2, who was also a native speaker of English, was that the feedback from Grammarly is better than MS Word checkers. This participant stated,

> *I'm not saying Grammarly is perfect as it sometimes underlines words or phrases that are correct English, but I still feel the app is sort of aware of the context.*

Data from non-native participants also showed more enthusiasm about and trust in Grammarly and they all had it in their electronic devices. Data from all non-native participant showed that Grammarly helps in eliminating language errors that they thought would surely affect how their assignments are marked. Thus, interview data revealed that non-native speakers used Grammarly regularly, but they knew other AWCF applications (e.g., WhiteSmoke, LanguageTool, Writesaver and Virtual Writing Tutor) which seemed understandable as L2 learners would be less confident about their writing quality and would seek help from available sources.

4.2   Participants' errors and editing behaviour

Various L2 studies focused on examining revision behaviours occurring while students are writing (e.g.Lindgren & Sullivan, 2003; New, 1999). This study aimed to examine and compare native and non-native participants' real time editing behaviour when AWCF from Grammarly was in place. There was also a necessity to recognize specific editing behaviours not only in terms of comparing native to non-native speakers (i.e., between groups), but also in terms of discrete behaviour of each

individual (i.e., within group). Analysis of such behaviours was based on the examination of screen capture data as this type of technology offers what Seror (2013) described as "*the unique advantages of being able to unobtrusively gather, store and replay what have traditionally remained hidden sequences of events at the heart of L2 writers' text production*" (p.1). Reporting errors as a proportion based on text length was not possible because the length of the texts produced was not the same among participants.

A corpus of data on nine types of errors (i.e. grammar, spelling, punctuation, passive voice, conciseness, unclear antecedent, formality, vocab and word choice and repetition) was collected from drafts written by all participants in Grammarly in an attempt to understand the nature of errors highlighted by the system. The total number of errors flagged for all participants was 102 in all language areas (see figure 2).

Figure 2.
Aggregated number of flagged errors for all participants
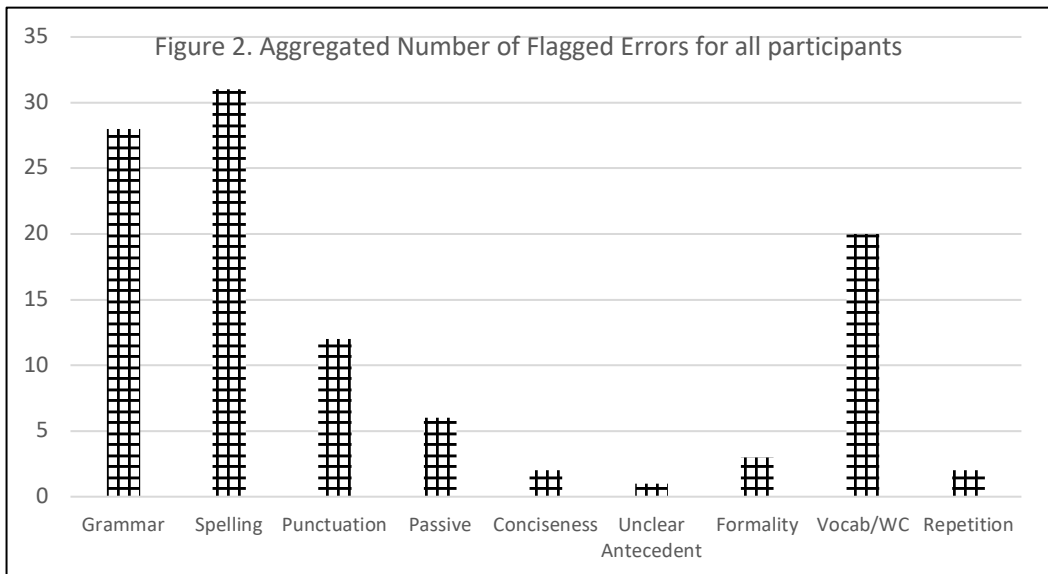
The distribution of the corpus of errors included 31 spelling, 28 grammar and 20 vocabulary and word choice, i.e. these were the most common errors correspondingly. There were also 12 punctuation errors among all participants and 6 passive voice, 2 conciseness and 1 unclear element (see figure 4).

A closer look at which errors were flagged for each group of participants was still needed. This is discussed in terms of the revision behaviour (acceptance/rejection/avoidance) of native and non-native participants. This was undertaken in two steps; the first of which was looking at the aggregated figures of flagged errors and the nature of behaviour (adapt/reject/avoid) for all participants. The second step was comparing the numbers of flagged errors for each group of participants (native vs. non-native students) and then examine the adoption behaviour for each individual in each group. Analysis of the data revealed that participants adapted all (28 grammar and 20 vocabulary and word choice), or almost all (30 out of 31 spelling and 10 out of 12 punctuation) suggested errors. In other words, 94 corrections out of 102 were accepted by participants. It was also noticed that only 4 corrections were rejected (only 1 in spelling) and 4 passive errors were avoided (see Table 2).

Table 2.

Aggregated figures of participants' uptake

Focus	Flagged Errors	Adapted	Rejected	Avoided
Grammar	28	28	-	-
Spelling	31	30	1	-
Punctuation	12	10	2	-
Passive	6	2	-	4
Conciseness	2	2	-	-
Unclear Antecedent	1	1	-	-
Formality	2	2	-	-
Vocab & WC	20	20	-	-
Repetition	-	-	-	-
TOTAL	**102**	**94**	**4**	**4**

Furthermore, it was noticed from the overall data analysis that native speakers had more flagged errors (i.e. 62) than non-native speakers (i.e. 40). Non-native speakers however, had more flagged errors only in grammar, but native speakers had more in relation to all other areas. Nevertheless, the two groups followed a similar distribution. Data on errors flagged for native speaker participants was compared to non-native speakers as an initial step to understanding whether or not revision behaviour was dis/similar. Additionally, there was a need to understand the specific revision behaviour not only in terms of comparing native to non-native speakers (i.e. between groups), but also in terms of individual participants (i.e. within the group). Native speaker participants are described as Participants 1, 2 and 3 and non-native speakers are Participants 4, 5 and 6.

Data analysis also revealed that native speakers had a total of 62 flagged errors, which were distributed among the three participants (see table 3). Analysis of the screen capture recorded data showing the revision behaviour of all native speaker participants revealed a total of 4 rejection and 4 avoidance responses among the three native speakers involved. The rest of suggested corrections on errors flagged by Grammarly (i.e. N=54) were accepted.

Table 3.

Native Speakers' Revision Behaviour

	Language Focus	Participant 1 Adapted	Participant 1 Rejected	Participant 1 Avoided	Participant 1 Tot	Participant 2 Adapted	Participant 2 Rejected	Participant 2 Avoided	Participant 2 Tot	Participant 3 Adapted	Participant 3 Rejected	Participant 3 Avoided	Participant 3 Tot
*Cor.	Grammar	-	-	-	-	5	-	-	5	4	-	-	4
*Cor.	Spelling	3	-	-	3	19	1	-	20	-	-	-	-
*Cor.	Punctuation	4	1	-	5	1	1	-	2	2	-	-	2
*Eng	Vocab.	3	-	-	3	4	-	-	4	-	-	-	-
*Eng	W.choice/variety	-	-	-	-	2	-	-	2	3	1	-	4
*Cla.	Conciseness	2	-	-	2					-	-	-	-
*Cla.	Passive V	-	-	1	1	-	-	3	3	-	-	-	-
*Cla.	Unclear	1	-	-	1	-	-	-	-	-	-	-	-
*Del	Tone	1	-	-	1	-	-	-	-	-	-	-	-
	Total errors	**14**	**1**	**1**	**16**	**31**	**2**	**3**	**36**	**9**	**1**	**-**	**10**

*Cor = correctness / *Eng = engagement / *Cla = clarity / *Del = delivery

Further analysis of individual native participants was needed in order to understand whether this group of participants followed a specific response pattern and whether any possible pattern, if any, is dis/similar to non-native speakers.

Non-native speakers' data analysis showed that they had a total of 40 errors highlighted by Grammarly (see Table 4), which was 22 fewer than the aggregated figure logged for native speakers. However, screen capture data showed no rejection or avoidance in the revision behaviours of any of the participants in this group suggesting, many of them trusted the automated comments provided by the system.

Table 4.

Non-native participants' revision behaviour

Language Focus		P4				P5				P6			
		Adapted	*Rejected*	*Avoided*	*Tot*	*Adapted*	*Rejected*	*Avoided*	*Tot*	*Adapted*	*Rejected*	*Avoided*	*Tot*
*Cor.	Grammar	7	-	-	7	6	-	-	6	6	-	-	6
	Spelling	4	-	-	4	3	-	-	3	1	-	-	1
	Punctuation	3	-	-	3	-	-	-	-	-	-	-	-
*Eng.	Vocab.	-	-	-	-	-	-	-	-	-	-	-	-
	W. Choice/ variety	4	-	-	4	2	-	-	2	1	-	-	1
*Cla.	Conciseness	-	-	-	-	-	-	-	-	-	-	-	-
	Passive V	-	-	-	-	1	-	-	1	1	-	-	1
	Unclear	-	-	-	-	-	-	-	-	-	-	-	-
*Del	Tone	-	-	-	-	-	-	-	-	1	-	-	1
Total errors		**18**	**-**	**-**	**18**	**12**	**-**	**-**	**12**	**10**	**-**		**10**

*Cor = correctness / Eng = engagement / Cla = clarity / Del = delivery

More analysis of individual participants was needed in order to understand whether this group of participants followed a specific revision behaviour pattern. This group of participants had 40 errors highlighted between them. While some revision behaviour suggests a straightforward adoption of recommended corrections, occasionally participants had to choose between two suggestions, as this researcher will explain in the following sections.

4.3 Native speakers' Editing Behaviour

4.3.1 Participant 1

Participant 1 was a female native speaker who was using Grammarly regularly at the time this study was carried out. This participant ignored flagged errors until she was 4 minutes in (see Figure 3) when she examined each error one by one. Having reviewed the section she finished, participant 1 then continued writing and did a second batch of revision in the 8th minute.

Although some sentences seemed to be correct, the AWCF provided by Grammarly made this participant unsure. This agreed with her responses in the pre-activity interviews reported earlier in this study. For instance, Figure 3 shows a good example of the extent to which this native speaker writer took some time contemplating. This suggests that she took comments from Grammarly seriously. The suggestion made in the example shown led the student to decide to make changes to the sentence in the end. It was noticed in the example that the student also voluntarily changed 'strongly' to 'firmly' even though it was not flagged up. This may not have been changed if the other words

had not drawn closer attention to this part of the text. This suggests that it is possible that Grammarly might create self-initiated noticing opportunities for a student who is keen to improve their writing.

Figure 3.
Participant 1's revision behaviour after 4 minutes of writing

Figure 4.
Screenshot from Participant 1 contemplating flagged errors on sentences

As explained in the post-activity data, this participant was also hesitant about the underlined text shown in Figure 4, which were left unresolved until almost the end of the writing session because she was not sure about the right form.

4.3.2 Participant 2

Participant 2 was a male native speaker and was a regular user of Grammarly. Unlike participant 1, it was noticed that this participant revised alongside writing. Having written the first sentence, this participant started editing immediately and the same behaviour was maintained after each sentence until the end of the writing session (see Figures 5 and 6). However, if this observation is linked to data obtained about the type of errors flagged, his writing had the highest spelling errors compared to his fellow native speakers as well as non-native speakers. Although the data suggests he is a poor speller (see Figure 6), data from the post-activity interview demonstrates that these errors were just

keyboard stroke errors. His responses also show that the errors are due to a habitual over reliance on the spell checker.

Figure 5.
First example of Participant 2's instant editing of flagged errors

Figure 6.
*Second example of Participant 2's instant editing of flagged errors*

Although most of the errors marked for this participant were in spelling (i.e., 20 errors), Grammarly seemed to have missed flagging the absence of "is" in the sentence "*Either way, it desirable*

to understand and research a little into…" (see Figure 6). It was also noticed that this participant used passive sentences correctly a few times which were flagged by Grammarly as '*rewrite this sentence*'; and in every occasion this was flagged the student changed the sentences despite being correct (see Figure 7).

Figure 7.
Example of passive form of Participant 2
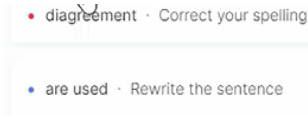
Thus, Participant 2 had mainly spelling or keyboard stroke errors flagged and accepted 19 of these immediately after being flagged. He also seemed to avoid 3 passive errors to side-step such sentences being detected by Grammarly as erroneous.

4.3.3   Participant 3

Participant 3 was a female native speaker and also a regular Grammarly user. To this participants, Grammarly feedback sometimes show some "*embarrassing mistakes*". She said "*it is like a second pair of eyes*". Like participant 2, participant 3 was concurrently writing and editing. This participant looked at the flagged errors after she had finished writing almost every sentence. It was noticed in the screen capture video that participant 3 sometimes read the feedback and adopted the suggestion before finishing the sentence (see Figure 8).

Figure 8.
Participant 3 adoption of comment before finishing the sentence

Out of 10 errors flagged by Grammarly, only one word choice suggestion was rejected by this participant and none was avoided. She accepted all grammar and punctuation comments without consulting any other sources. This participant also indicated the need for contemplating some comments more than others. Thus, reference to the avoidance of full reliance on AWCF from Grammarly was indicated in the interviews. This is simply because Grammarly's intelligent mistakes corrector is incapable of dealing with advanced errors, participant 3 seemed to more trusting of comments on surface errors.

4.4   Non-native speakers' revision behaviour

4.4.1   Participant 4

This participant was a female non-native speaker of English from China. She is accustomed to using Grammarly as well as other AWCF applications such as LanguageTool and Writesaver. This participant had positive comments about Grammarly in the pre-activity interviews. To her, comments on grammar errors is key to successful writing. Participant 4 had a total of 18 flagged errors

of which 7 were in grammar, 3 in punctuation, 4 in spelling and another 4 in word choice. All suggested comments were adopted by her. Participant 4 encountered more than once a flagged error with two suggestions in the automated comment. This needed some contemplation on the part of the participant and a need to resort to her own linguistic resources to decide whether to adopt the suggested comments. An example is given in Figure 9.

Figure 9.
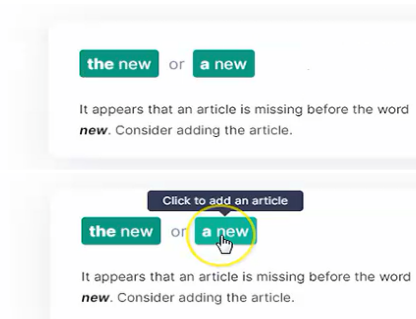*Example of Participant 4 choosing between 2 suggested comments*

Figure 9 shows that participant 4 had to make a choice between '*a new* or *the new*'. The screen capture data initially revealed that she spent time before making any edits. The interview data revealed that this participant is used to reading automated comments several times before editing. The participant adopted 'a new' rather than 'the new'. However, the screen capture later revealed that there was further action as seen in Figure 10.

Figure 10.
Participant 4's manual change of previously adopted comment

Participant 4 re-visited the same sentence again and re-read the sentence reversing her choice to '*the new*' instead of '*a new*'. Interview data with the participant indicated that this participant did not adopt the AWCF unquestioningly and she re-visited the previously adopted suggestion. She stated that feedback from Grammarly is useful, but the comments can sometimes be confusing.

4.4.2 Participant 5

This participant was also a female Chinese student who had been using Grammarly and is familiar with other applications including Chinese examples that were unknown to this researcher. In the pre-activity interview, this participant referred to the use of Google translate as an online strategy for obtaining feedback. Although this was interesting, it is not discussed in this paper as it was viewed to be beyond the scope of this research. She had a total of 12 errors, of which 6 were grammar, 3 were spelling, 2 were word choice and 1 passive voice error. This participant adopted all

flagged comments including the passive voice that other students sometimes avoid by changing their text. Similar to the revision behaviour of participant 4, video evidence showed participant 5 revise a marked error (i.e., *a different* instead of *different*) and adopted it, but decided later to make her own revision and used 'different types' instead of '*a different type*' (see Figure 11).
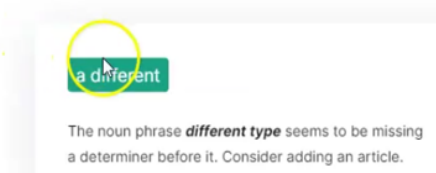
Figure 11.
Participant 5 manual change of previously adopted comment

It was noticed that there was occasional failure on the side of Grammarly to underline/flag punctuation errors in this participant's text. This was seen in the full stop before "illiterate" and capitalization in "therefore" (see Figure 12). Participant 5 manually amended the capitalization, but she did not pick up the full stop.

Figure 12.
Participant 5's revision of unflagged error

Participant 5 showed awareness of the limitations in some comments. Her responses in the post-activity interview suggested that she had built some strategies which would help her decide on which automated comments she would instantly accept. These were mainly comments on erroneous grammar output. She also explained the importance of a final proofread using her linguistic resources.

4.4.2   Participant 6

The last participant was also a Chinese female who had a total of 10 flagged errors with suggested comments, which she adopted. Among these errors, 6 were in grammar and 1 error each in spelling, word choice, passive and tone respectively. This participant was also a regular user of AWCF. Like participant 4, video data of participant 6's writing session revealed that this participant also revised as soon as an error was flagged although the sentence was not finished (see Figure 13).

Figure 13.
Participant 6's example of editing every sentence

To this participant, flagged errors act as distractors and impede her train of thought. She prefers to correct language errors automatically and concentrate on the content. Participant 6 suggested that AWCF were useful to her even in her study in China before coming to the UK. She stated that she has been relying on Grammarly when revising all her assignments. This is reflected in her revision behaviour in relation to the errors flagged by the system, which she immediately adopted. No errors were rejected or avoided in the case of this participant. Participant 6 suggested that AWCF was useful to her even in her study in China before coming to the UK. She stated that she has been relying on Grammarly when revising all her assignments. This is reflected in her editing behaviour in relation to the errors flagged by the system, which she immediately adopted. To this participant, flagged errors act as distractors and impede her train of thought. She prefers to correct grammar errors automatically and concentrate on the content. No errors were rejected or avoided in the case of this participant.

5   Discussion and Conclusion

The results obtained in this study suggested that both native and non-native speakers involved in this study are constantly using Grammarly, as well as other AWCF tools, to improve their written texts. This study suggests that students will often know, and probably use, AWCF applications that are not necessarily known for their teachers. The danger here is that teachers would only see final written output and might lose out on important information about how those texts were created and what feedback was received on earlier drafts. This finds support in Yamashita's (2021) emphasis on the importance of collecting thorough data on real time revisions. Scrutiny of the screen capture recorded data showing the editing behaviour of all native-speaker participants revealed a total of 4 rejection and 4 avoidance responses. All the other suggested corrections on errors flagged by Grammarly (i.e.,N=54) were accepted. Non-native speakers' data analysis showed that they had a total of 40 errors highlighted by Grammarly, which was 22 fewer than the aggregated figure logged for native speakers. However, screen capture data showed no rejection or avoidance in the editing behaviours of any of the non-native participants suggesting that they trusted the automated comments

provided by the system. However, non-native participants received more feedback on a specific language area (i.e., grammar), but both parties followed similar editing patterns in their editing behaviour. This indicates high levels of participants' behavioural engagement with feedback from Grammarly. There is also evidence from the pre-activity interview which suggested positive attitudinal engagement with Grammarly on the side of both native and non-native participants. Behavioural engagement is the extent to which students incorporate the suggested accurate forms in their modified texts and attitudinal engagement refers to attitudes towards feedback (Ellis, 2010). This finding agrees with the finding reported by Koltovskaia (2020). It was evident that the screen capture recording of the writing sessions included no further online look-up strategies of the feedback, which reflected trust in AWCF. The current study, however, puts forward the claim that cognitive engagement with error flagging tools in general and Grammarly in particular is under-researched. The concept of cognitive engagement is defined by Ellis (2010) as 'how learners attend to the CF' (p.342).

Whilst the native speaker participants involved in this study rejected some suggested comments and avoided others (i.e., passive voice), non-native participants seemed to adopt all comments including passive voice. There was evidence of contemplating certain errors on word choice and avoiding unclear comments on passive voice. Post activity interview data on avoidance of automated comments on the passive voice was generally described as "unclear". This finding suggests that compared to teacher feedback, it is expected to know a priori that automated feedback on certain areas like the passive voice will remain hard to understand which eliminates the potential uptake from this type of comments. Although this study did not look at any mental processing, the screen recording data suggests that participants did not adopt the automated feedback unquestioningly and some of them re-visited the adopted suggestions despite the low incidence in the data set. Overall, suggestions that this particular AWCF tool (i.e., Grammarly) is useful in providing immediate feedback on surface errors is supported in this study. This was clear in the post-activity interview in which native and non-native participants dismissed the idea of accepting all comments automatically. Many, including the current researcher, think Grammarly can particularly help L2 learners improve their writing abilities (O'Neill & Russell, 2019). Here, Bailey and Lee (2020) state that "*Grammarly should be added to the language learner's L2 writing strategy repertoire*" (p.22). On the other hand, this study suggests that Grammarly is not without its inaccuracies and the benefit from its comments is connected with the language proficiency of the L2 writers and/or the guidance provided on feedback. Apart from the simple local errors (e.g., spelling), a closer look at the screenshots provided suggests that Grammarly might be seen as a poor resource for L2 writers *except at the most advanced levels*, with some of the "*flagged errors*" clearly perfectly correct English, but stylistically different from the overly prescriptive and restricted views of the writing experts used for the software. This is because less advanced learners might not have the linguistic resources that would help them decide how to react to the flagged errors. Thus, the claim that Grammarly is a useful tool that many L1 and L2 writers know and often use can still be acceptable, provided that teachers become involved, and have enough understanding of what the application can offer learners in their specific learning contexts. This might compensate for the lack of enough guidance on the errors flagged for less advanced L2 learners and create opportunities for cognitive engagement.

The term automated error correction is, to some extent, generic, as more applications seem to be included under this umbrella term. This study suggests that a distinction should be made between applications that can provide information to classroom instructors as well as students about their learning, and applications that might improve a script but are impersonal as they do not record an individual student's progress (e.g., error flagging applications). As mentioned earlier in this study, some applications can log all information about the writing processes, and the resulting product, by creating a platform that can be accessed and controlled by a classroom instructor. This can provide a pathway for guided and informed scaffolding. However, other applications can be classified as 'automated error flagging' as these can underline/highlight errors to the student *while* writing and this student must choose whether to adopt, reject or avoid the suggested feedback. While such behaviour can be informed by the student's linguistic resources or based on a further online look-up strategy that might or might not lead to learning, the student's decision can easily be uninformed.

Knowledge about such revision behaviours is simply unknown to teachers in the case of error flagging applications. Therefore, this researcher claims that automated error flagging tools are still useful, but the use of such tools should involve some teacher intervention in terms of the choice of the technological tool and the way students are using them. This could happen by integrating another technology (e.g., screen recording) or a simple student logbook or writing journal that can enable both teachers and students to reflect on flagged errors as well as the revision moves students might make when responding to the flagged errors. Teacher intervention and the integration of a writer's logbook or a journal can help promote self-directed learning.

References

Attali, Y. (2004). *Exploring the Feedback and Revision Features of Criterion* National Council on Measurement in Education (NCME), San Diego, CA. .

Attali, Y., & Burstein, J. C. (2004). *Automated essay scoring with E-rater 2.0.* Conference of the International Association for Educational Assessment, Philadelphia, PA.

Bailey, D. a., & Lee, A. (2020). An Exploratory Study of Grammarly in the Language Learning Context: An Analysis of Test-Based, Textbook-Based and Facebook Corpora. *TESOL International Journal*, *15*(2), 2094-3938.

Barrot, J. S. (2020). Integrating Technology into ESL/EFL Writing through Grammarly. *RELC Journal*. https://doi.org/ https://doi.org/10.1177/0033688220966632

Barrot, J. S. (2021). Using automated written corrective feedback in the writing classrooms: effects on L2 writing accuracy. *Computer Assisted Language Learning*. https://doi.org/DOI: 10.1080/09588221.2021.1936071

Bereiter, C., & Scarmalia, M. (1987). *The psychology of written composition*. Routledge.

Bitchener, J., & Storch, N. (2016). *Written Corrective Feedback for L2 Development* (1st edition ed.). Multilingual Matters.

Brown, D. (2014). The type and linguistic foci of oral corrective feedback in the L2 classroom: A meta-analysis. *Language Teaching Research*, *20*(4), 436-458.

Casanave, C. (2004). *Controversies in Second Language Writing: Dilemmas and Decisions in Research and Instruction* (Second Edition ed.). The University of Michigan Press.

Chapelle, C. A., Cotos, E., & Lee, J. Y. (2015). Validity arguments for diagnostic assessment using automated writing evaluation. *Language Testing*, *32*(3), 385–405. https://doi.org/doi:10.1177/0265532214565386

Cooper, C. R., & Matsuhashi, A. (1983). A theory of the writing process. In M. Martlew (Ed.), *The psychology of written language* (pp. 3-39). John Wiley and Sons.

Delima, E. (2019). Grammarly as a Tool for Enhancing Students' Essays. *Asian EFL Journal Research Articles*, *25*(5.1), 102-112. http://www.asian-efl-journal.com/

Dembsey, J. (2017). Closing the Grammarly Gaps: A Study of Claims and Feedback from an Online Grammar Program. *The Writing Center Journal*, *36*(1), 63-96. https://www.jstor.org/stable/44252638

Dikli, S., & Bleyle, S. (2014). Automated essay scoring feedback for second language writers: How does it compare to instructor feedback? *Assessing Writing*, *22*, 1-17.

El Ebyary, K., & Windeatt, S. (2010). The impact of computer-based feedback on students' written work. *International Journal of English Studies (IJES)*, *10*(2), 121-142. https://files.eric.ed.gov/fulltext/EJ936915.pdf

Ellis, R. (1997). *Second Language Acquisition*. Oxford University Press.

Ellis, R. (2009). A typology of written corrective feedback types. *ELT Journal*, *63*(2), 97–107.

Ellis, R. (2010). Epilogue: A framework for investigating oral and written corrective feedback. In *Studies in Second Language Acquisition - Special Issue on The Role of Oral and Written Corrective Feedback in SLA* (Vol. 32, pp. 335–349).

Enright, M. K., & Quinlan, T. (2010). Complementing human judgment of essays written by English language learners with e-rater scoring. *Language Testing*, *27*(3), 317–334.

Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication*, *32*, 365-387.

Grammarly. (2012). *Grammarly User Survey Analysis*. Retrieved 11/08/2019 from https://www.grammarly.com/press/research/docs/grammarlystudentsurvey-121018133119-phpapp01.pdf

Grammarly. (2021). Retrieved 22/04 from https://www.grammarly.com/

Guo, Q., Feng, R., & Hua, Y. (2021a). How effectively can EFL students use automated written corrective feedback (AWCF) in research writing? *Computer Assisted Language Learning*. https://doi.org/DOI: 10.1080/09588221.2021.1879161

Guo, Q., Feng, R., & Hua, Y. (2021b). How effectively can EFL students use automated written corrective feedback (AWCF) in research writing? *Computer Assisted Language Learning*, 1-20. https://doi.org/10.1080/09588221.2021.1879161

Hill, C., Wallace, D., & Haas, C. (1991). Revising on-line: Computer technologies and the revising process. *Computers and Composition*, *9*(1), 83-109.

Hoang, G., & Kunnan, A. (2016). Automated Essay Evaluation for English Language Learners:A Case Study of MY Access. *Language Assessment Quarterly*, *13*(4), 359-376. https://doi.org/ DOI: 10.1080/15434303.2016.1230121

Karyuatry, L. (2018). Grammarly as a tool to improve students' writing quality: free online-proofreader across the boundaries. *JSSH (Jurnal Sains Sosial dan Humaniora) 2*(1). https://doi.org/DOI:10.30595/JSSH.V2I1.2297

Koltovskaia, S. (2020). Student engagement with automated written corrective feedback (AWCF) provided by Grammarly: A multiple case study. *Assessing Writing*, *44*, 1004503. https://doi.org/https://doi.org/10.1016/j.asw.2020.100450.

Li, J., Link, S., & Hegelheimer, V. (2015). Rethinking the role of automated writing evaluation (AWE) feedback in ESL writing instruction. *Journal of Second Language Writing,*, *27*, 1-18. https://doi.org/https://doi.org/10.1016/j.jslw.2014.10.004

Li, S. (2010). The Effectiveness of Corrective Feedback in SLA: A Meta-Analysis. *Language Learning*, *60*(2), 309–365.

Li, Z., Dursun, A., & Hegelheimer, V. (2017). Technology and L2 Writing. In C. A. Chapelle & S. Sauro (Eds.), *The Handbook of Technology and Second Language Teaching and Learning* (pp. 77-). John Wiley & Sons, Inc.

Lindgren, E., & Sullivan, K. (2003). Stimulated Recall as a Trigger for Increasing Noticing and Language Awareness in the L2 Writing Classroom: A Case Study of Two Young Female Writers. *Language Awareness*, *12*, 172–186.

Link, S., Mehrzad, M., & Rahimi, M. (2020). Impact of automated writing evaluation on teacher feedback, student revision, and writing improvement. *Computer Assisted Language Learning*, 1-30. https://doi.org/DOI: 10.1080/09588221.2020.1743323

Liu, S., & Kunnan, A. (2016). Investigating the Application of Automated Writing Evaluation to Chinese Undergraduate English Majors: A Case Study of WriteToLearn. *CALICO Journal 33*(1), 71-91. http://www.jstor.com/stable/calicojournal.33.1.71

Miaoa, Y., Badger, R., & Zhenc, Y. (2006). A comparative study of peer and teacher feedback in a Chinese EFL writing class. *Journal of Second Language Writing*, *15*(3), 179-200.

Nassaji, H., and , & Kartchava, E. (2019). A meta-analysis of the effects of instruction and corrective feedback on L2 pragmatics and the role of moderator variables. *Special issue of International Journal of Applied Linguistics 170:2 ()*, *170*(2), 151-153.

New, E. (1999). Computer-Aided Writing in French as a Foreign Language: A Qualitative and Quantitative Look at the Process of Revision. *The Modern Language Journal*, *83*(1), 80-97.

Norris, J., and, & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, *50*, 417–528.

O'Neill, R., & Russell, A. (2019). Stop! Grammar Time: University Students' Perceptions of the Automated Feedback Program Grammarly. *Australasian Journal of Educational Technology*, *35*(1), 42-56.

O'Neill, R., & Russell, A. M. (2019). Grammarly: Help or hindrance? Academic Learning Advisors' perceptions of an online grammar checker. *Journal of Academic Language and Learning*, *13*(1), A88-A107. https://journal.aall.org.au/index.php/jall/article/view/591

Polio, C. (2012). The relevance of second language acquisition theory to the written error correction debate. *Journal of Second Language Writing 21*, 375–389.

Qassemzadeh, A., & Soleimani, H. (2016). The impact of feedback provision by Grammarly® software and teachers on learning passive structures by Iranian EFL learners. *Theory and Practice in Language Studies*, *6*, 1884 - 1894.

Ranalli, J. (2018). Automated written corrective feedback: how well can students make use of it? *Computer Assisted Language Learning*, *31*(7), 653-674.

Ranalli, J., & Yamashita, T. (2020). Automated written corrective feedback: Error-correction performance and timing of delivery. *Language Learning & Technology*, *xx*(x), xx.

Relles, S., & Tierney, W. (2013). Understanding the Writing Habits of Tomorrow's Students: Technology and College Readiness. *Journal of Higher Education*, *84*,(4), 477–505. https://doi.org/10.1080/00221546.2013.11777299

Rudner, L. M., Garcia, V., & Welch, C. (2006). An evaluation of IntelliMetric essay scoring system. *Journal of Technology, Learning, and Assessment*, *4*(4), 3-29. http://escholarship.bc.edu/jtla/vol4/4

Seror, J. (2013). Screen capture technology: A digital window into students' writing processes. *Canadian Journal of Learning and Technology*, *39*(3), 1-16. https://files.eric.ed.gov/fulltext/EJ1029325.pdf

Spada, N. (2018). Isolating or integrating attention to form in communicative instruction: a dilemma? *Babel*, *53*(1), 7–13. https://search.informit.org/doi/10.3316/aeipt.221355

Ventayen, R., & Orlanda-Ventayen, C. (2018). Graduate Students' Perspective on the Usability of Grammarly® in One ASEAN State University *Asian ESP Journal*, *14*(7.2), 1-24. https://ssrn.com/abstract=3310702

Weigle, S. C., & Malone, E. (2016). Assessment of English for academic purposes. In K. H. P. Shaw (Ed.), *The Routledge handbook of English for academic purposes* (pp. 608–620). NY: Routledge.

Wilson, J., & Czik, A. (2016). Automated essay evaluation software in English Language Arts classrooms: Effects on teacher feedback, student motivation, and writing quality. *Computers & Education*, *100*, 94-109.

Yamashita, T. (2021). Corrective feedback in computer-mediated collaborative writing and revision contributions. *Language Learning & Technology*, *25*(2), 75–93. http://hdl.handle.net/10125/73434

Zhang, J., Ozer, Z. a., & Bayazeed, B. (2020). Grammarly VS. Face-to-face Tutoring at the Writing Center. *Praxis: A Writing Center Journal*, *17*(2), 33-47.

Zhang, Z., & Hyland, K. (2018). Student engagement with teacher and automated feedback on L2 writing, . *Assessing Writing*, *36*, 90-102. https://doi.org/https://doi.org/10.1016/j.asw.2018.02.004.

About the author(s)

Khaled El Ebyary (https://orcid.org/0000-0002-1710-2519) is a Lecturer and MA TESOL Programme Leader in the Department of Educa-tion at the University of York and a member of the Centre for Research in Language Learning and Use (CReLLU). His research interests include CALL, washback in language testing, curriculum design, classroom-based assessment, corrective feedback and automated writing evaluation. He pub-lished a number of research articles on eye-tracking and feedback, automated writing evaluation and others.