

Error Flagging Tools and Online Revision Behaviour: Evidence from Native and Non-native Speakers

Khaled El Ebyary
Damanhour University, Egypt

Abstract

The assumption that L2 writing is a daunting task for L2 learners and that native-speakers are privileged in the complex act of writing has often been present in second language research. 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 corrective feedback. In fact, little is known about tracking writers' revision behaviour when error flagging is in place, whether such behaviour would vary between native and non-native writers and how L2 writing instruction can benefit from such evidence. Using a pre-activity questionnaire, an IELTS writing task 2 and a screen capture software, the study compared the revision behaviours of native and non-native speakers of English when an error flagging application (i.e., 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.

Keywords: automated feedback; error flagging; Grammarly; revision behavior

تطبيقات رصد أخطاء الكتابة وسلوك المراجعة: مقارنة بين متعلمي اللغة الانجليزية كلغة أجنبية ومتحدثي اللغة الأصليين

المستخلص

ما تزل الفرضية بأن عملية الكتابة باللغة الثانية شاقة بالنسبة لمتعلمي اللغة الأجنبية، وسهلة بالنسبة لمتحدثي اللغة الأصليين حاضرة في أبحاث اللغة الإنجليزية كلغة أجنبية. بالتالي؛ ظهرت حاجة الباحثين ومعلمي اللغة الإنجليزية – بوصفها لغة ثانية – إلى Web 0.2 الاهتمام باستخدام أدوات الشبكة المتقدمة في عملية الكتابة ومثيلاتها من تطبيقات الرصد الإلكتروني لأخطاء الكتابة خاصة مع وجود نتائج كثير من الدراسات التي تؤكد تتمتع تطبيقات الرصد الإلكتروني لأخطاء الكتابة بإمكانات متطورة من شأنها تقديم تغذية راجعة دقيقة مما قد يساعد في مستوى الكتابة. وقد أظهرت الأدبيات ندرة الأبحاث المعنية بتتبع سلوكيات المراجعة التي يقوم بها الطلاب) سواء متحدثو اللغة الأصليين أو متعلموها كلغة أجنبية(، والي اي مدي قد تختلف تلك السلوكيات فيما بينهم؛ وكيفية الاستفادة من هذه السلوكيات في تدريس اللغة الإنجليزية كلغة ثانية. وعليه؛ حاولت الدراسة الحالية سد هذه الثغرة البحثية بمقارنة سلوكيات المراجعة التي يتبناها عدد من متعلمي اللغة الإنجليزية كلغة ثانية وعدد من Grammarly الناطقين باللغة الانجليزية كلغة أولى حال استخدامهم لبرنامج التصحيح الإلكتروني المهمة الثانية ELTS. -للإجابة عن أسئلة البحث استخدم الباحث استبانة مبدئية، و مهمة كتابة من اختبار

وأحد التطبيقات الإلكترونية المعنية بتصوير شاشات الكمبيوتر التي تظهر التغذية الراجعة وتتبع سلوكيات العينة بعد حصولهم عليها وقد أظهرت النتائج أن عدد الأخطاء الكتابية كانت أكثر لدى ناطقي اللغة الإنجليزية بوصفها لغة أولى منها لدى متعلميها بوصفها لغة ثانية؛ غير أن هذه المجموعة كانت أخطائها النحوية أكثر من المجموعة الأولى كما أوضحت النتائج تشابه سلوكيات المراجعة بين مجموعتي الدراسة وعليه، اقترح الباحث ضرورة دراسة دور المعلم المعني بتقديم تغذية راجعة في ضوء وجود تطبيقات إلكترونية تقوم بهذا الدور وسلوك المتعلمين الناتج عن التغذية الراجعة المقدمة من هذه التطبيقات ودوره في تطور مهارة الكتابة لدى متعلمي اللغة الإنجليزية

Error Flagging Tools and Online Revision Behaviour: Evidence from Native and Non-native Speakers

Khaled El Ebyary

Damanhour University, Egypt

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) as '*the most consuming of all dilemmas for L2 writing teachers*' (p.64). The 1980's and 1990's witnessed a plethora of research that focused on understanding the complexities of the composing act (e.g., Bereiter & Scardamalia, 1987; Cooper & Matsuhashi, 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. 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). This study claims that whilst L2 writing researchers in the different research dimensions effected to inform EFL classroom instruction, there is a difference in the perspectives of EFL teachers and students. While teachers are keen to enable students to write without the support of web tools because that 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 automated corrective feedback (ACF). This finds support in the feedback literature and the literature that focused on the role of technology in L2 writing. Ferris (2004) states that the strong desires for feedback on the side of L2 learners is undismissible. Thirty years ago, Hill, Wallace and Haas (1991, p. 83) described the role of technology as ‘increasingly prevalent in writing classrooms’ and described educators as ‘hold[ing] high expectations for the ways in which this technology can support writing and the teaching of writing’. Twenty years after Hill et al.’s (1991) accounts, Relles and Tierney (2013, p. 501) state 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’. This suggests that Hill et al. (1991) were rather optimistic in their prediction of ACF becoming much more widespread in L2 writing practice. More recently, claims of extensive use of ACF tools have been maintained by various researchers (Guo et al., 2021; Weigle & Malone, 2016).

2. The Study Focus

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 classrooms. Here, Silva (1993) states “ESL practitioners have frequently been advised to adopt practices from L1 writing” (p657). The current study lent itself to the assumption that particularly L2 writers use error flagging technological tools to improve their written output with or without real learning occurring and they produce language output that may or may not authentically reflect their abilities. The ubiquitous technology use has given L2 learners maximum control over which automated feedback tool(s) 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. Grammarly seems to put money and effort into advertising across multiple Media formats that target skilled (native speaker) and less skilled (non-native speaker) writers. Whilst an early survey conducted by Grammarly designers suggested that 68% of Grammarly users are native speakers compared to 32% non-natives (Grammarly, 2012), various second language researchers examined L2 Grammarly users’ perception or its impact on their writing abilities. In fact, little is known about L2 learners’ revision processes *while* writing (reaction to error flagged by the full version of Grammarly) and whether

these L2 revision behaviours are similar/different to native speakers who are reportedly using the same tool. Therefore, the study aimed to answer two questions:

1. What feedback comments from Grammarly do non-native L2 learners receive during the composing stage of their written output as compared to feedback comments received by native speakers of English at the same stage of writing?
2. What revision behaviours do non-native L2 learners adopt, and to what extent are these behaviours similar to/different from the behaviours adopted by native speakers of English?

Such evidence is thought to be beneficial to L2 writing instruction.

3. Revision Behaviour

Four decades ago when writing was only pen-and-paper and before the emergence of what is now L2 online writing environments, Faigley and Witte (1981) described revision as a *'tidying-up activity aimed at eliminating surface errors in grammar, punctuation, spelling, and diction'* (p.400), which suggested that such a process is deferred to the final writing stages. This account of the earlier stages of revision behaviour models seems to suggest an initial divide between what Hayes (2004) described as models that saw revision as a process done to a *previously produced script*, and models that pictured revision as behaviour occurring *during* the creation of a given script. Developments in researchers' understanding of the revision process over the years seem to disagree with views suggesting that revision is only undertaken with previously created texts (Chenoweth & Hayes, 2001; Chenoweth & Hayes, 2003; Fitzgerald, 1987; Kaufer et al., 1986). In fact, Hayes (2004) describes revision as behaviour undertaken by L2 writers at any stage of writing. This view can be supported by a simple reflection on the nature of revision for an L2 writer, which is described as detecting a formal error or a discrepancy between the actual L2 text in the making and the writer's plans for this text (Chenoweth & Hayes, 2001). However, almost all researchers agree with Faigley & Witte (1981) in their description of revision as complex (Barkaoui, 2016; De Larios et al., 2006; Hayes, 2012).

In their account for analysing revision, Lindgren, and Sullivan (2006) cited earlier researchers (e.g., Chanquoy, 2001) in describing the extensive purpose of revision as *'improvement or verification of the external text, or as improvement or verification of the internal*

representation of the text'(p.84). In earlier taxonomies of revision, external improvements were referred to as surface corrections (e.g., Chanquoy, 2001; Faigley & Witte, 1981) and meaning related corrections were described as internalised text representation (e.g., Lindgren & Sullivan, 2006), text-based changes (e.g., Faigley & Witte, 1981) or semantic revisions (e.g., Chanquoy, 2001).

4. AWE and Error Flagging Tools

With the pervasive use of advanced technology and the advent of Web 2.0, 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 Write&Improve (<https://writeandimprove.com/>) Grammarly (<https://www.grammarly.com/>), PaperRater (<https://www.paperrater.com/>), Online Correction (<https://www.onlinecorrection.com/>) among others. More recently, the pendulum has swung enthusiastically in the direction of the use of these tools by skilled and less skilled writers. Compared to the commercial packages, writing data is generally not logged in most of the error flagging packages (apart from word count in Grammarly) simply because

most of these are add-ons or in-browser applications which writers use while producing texts, and the errors are generally ‘flagged’ on the side and/or highlighted in the text. Although this limits the instructional value of such applications, their widespread use by students demonstrates their worth. The current study, therefore, claims that examining the revision behaviour in error flagging tools and particularly in Grammarly has generally been overlooked. The study synonymously uses the term automated error flagging applications and automated written corrective feedback (AWCF) which has been used in some studies (e.g., Ranalli, 2018). More recently, Grammarly has drawn the attention of various researchers (e.g., Dembsey, 2017; Guo et al., 2021; Koltovskaia, 2020; O’Neill & Russell, 2019; Qassemzadeh & Soleimani, 2016; Ventayen & Orlanda-Ventayen, 2018).

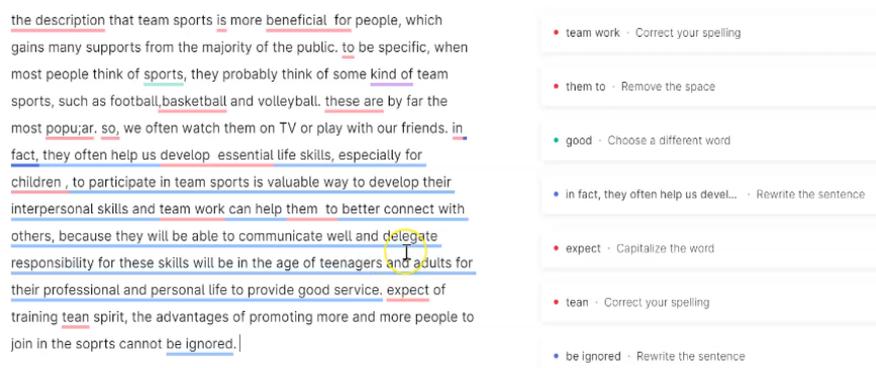
5. Grammarly

Grammarly seems to be increasingly used by both native and non-native writers. However, the insufficiency of research evidence has been acknowledged by some researchers (e.g., O’Neill & Russell, 2019). 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.). Table 1 below 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



It is worth mentioning that Grammarly website suggests that all users have access to extensive suggestions about correctness and clarity; suggestions related to engagement and delivery are available only to Grammarly Premium subscribers.

Various studies suggest that Grammarly can develop writing quality and enhance L2 writers' confidence and engagement (e.g., D. Bailey & R. Lee, 2020; Karyuatry, 2018; Koltovskaia, 2020). In their attempt to study AWCF, Ranalli and Yamashita (2020, p. 2) describe these applications, including Grammarly, as having the "potential to find and correct more common L2 error types than simpler spelling and grammar checkers such as the one included in Microsoft Word (p.2)". At the level of students' perception, O'Neill and Russell (2019) compared the perception of two groups of learners where the first group of 54 received feedback from Grammarly and the second, which involved 42 learners, received conventional teacher feedback. Their study reported positive responses of the learners who received automated error flagging from Grammarly as opposed to less positive from the conventional feedback group. However, inaccuracy of some automated comments was acknowledged and a recommendation of combining both modes of feedback was advised. Bailey and Lee (2020) studied how levels of syntactic and lexical complexity affect ACF from Grammarly. They compared error frequency, error types and writing complexity for what they described as "university admission test essays, textbook-based descriptive essays, social network site (SNS) posts, and SNS comments" (p.2). Findings referred to punctuation errors as the most frequent and the study concluded that Grammarly can be adequate for flagging local

surface-level language errors such as articles, preposition, and verb-noun agreement.

Claims about the need for research on students' engagement with ACF is maintained by many researchers (e.g., Zhang & Hyland, 2018). Koltovskaia (2020) examined the nature of engagement with the feedback provided by Grammarly to ESL college students. Using a case study approach, Koltovskaia (2020) examined the engagement patterns of two participants who received feedback from Grammarly on their final drafts. Based on the behavioural, cognitive, and affective dimensions of engagement, findings indicated that lower cognitive engagement led to immediate acceptance of feedback and higher cognitive engagement led to questioning the automated feedback obtained. However, both patterns resulted in moderate text change.

The current researcher suggests that little is known about the students' revision acts when they use in-browser error flagging tools. In addition, there is a need to compare how skilled (native speaker) and less skilled (non-native speaker) writers react to flagged errors. The study therefore, focused on providing answers to the research questions a) *what error flagging applications, if any, do native and non-native speakers seem to use?*, and b) *what is the impact of using an automated error flagging application (i.e., Grammarly) on native and non-native participants' real-time online revision behaviour?*. The following section explains the study design and instruments.

6. Methodology

Participants

The present study involved 6 participants of whom 3 were native speakers from the UK and 3 were Chinese. Participants' age ranged from 22-25 and all non-native participants had 7 overall in their IELTS tests. All participants were students at a UK university at the time this study was conducted. All non-native participants described their language education contexts as examination-focused, which finds support in various studies involving international students (Miaoa et al., 2006).

Study Design and Questions

This study compared native and non-native participants who were required to write 400-500 word compositions on a topic. Grammarly was used in their browsers in which errors were flagged on the right-hand side of the page. 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/>). This enabled analysis of participants' revision behaviour when error flagging was provided.

The Study Instruments

The current study involved a pre-activity questionnaire that aimed to collect information about participants' experiences with error detection/flagging tools in addition to the perceived usefulness of such tools to their learning. 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. Grammarly was integrated in the browser (MS Word) when participants wrote their texts and MS Word grammar and spelling checkers were disabled. Participants were individually asked about topic familiarity and level of difficulty and they all thought the topic was familiar and not difficult. The screen capture software was used to record participants' writing and revision behaviour.

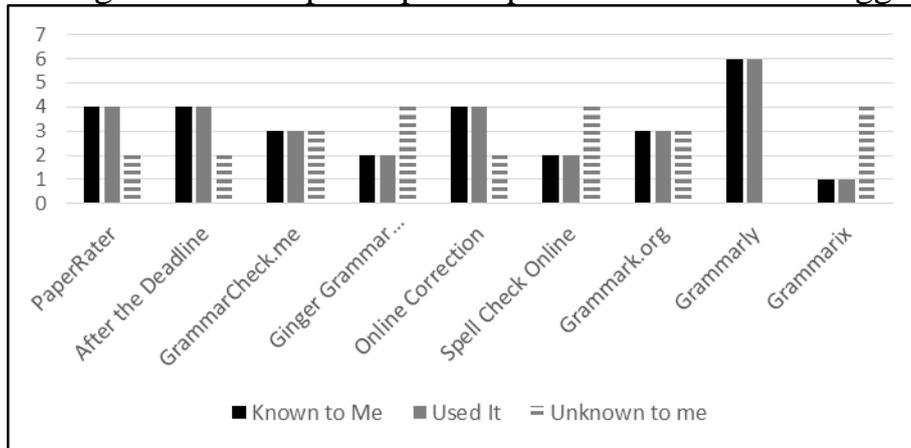
7. The Study Findings

The study findings are arranged in light of the research questions. In doing so, two major themes were used which included a) participants' use of error flagging applications, and b) participants' revision behaviour.

Participants' use of error flagging applications

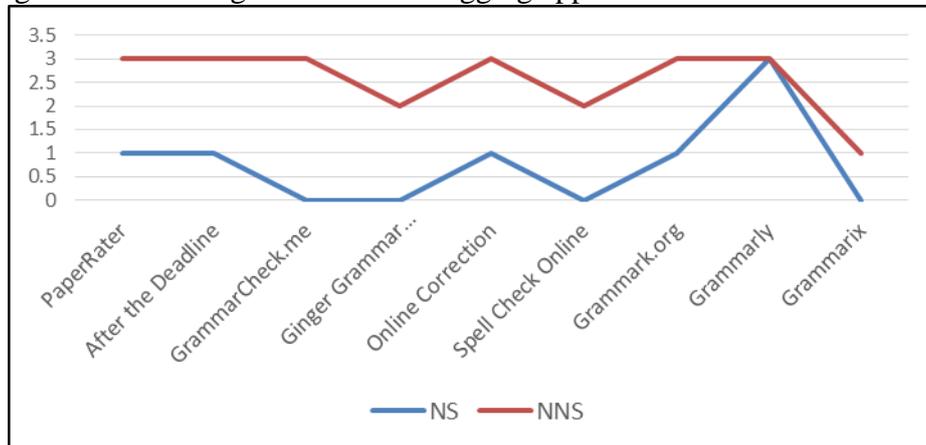
In terms of participants' experiences with error flagging software, this research aimed to explore whether knowledge and use of error flagging applications vary between native and non-native participants. Here, a list of the most common in/off browser commercial and freemium software packages was given to participants in the pre-activity questionnaire. They were asked to indicate whether they knew/used any of the tools on the list (see Figure 2).

Figure 2. Participants' past experiences with error flagging



While some applications were reported to be known/used by some participants rather than others, none of these applications were completely unknown to all participants. Grammarly emerged as the most commonly known/used application. However, it was not clear at this point in the data whether a specific group of participants (e.g., native speakers) tended to know (or use) more/less about such applications. So, further analysis was needed to see whether knowledge, and use, of such applications varied from one group to another (see Figure 3).

Figure 3. Knowledge/use of error flagging applications across NS and NNS



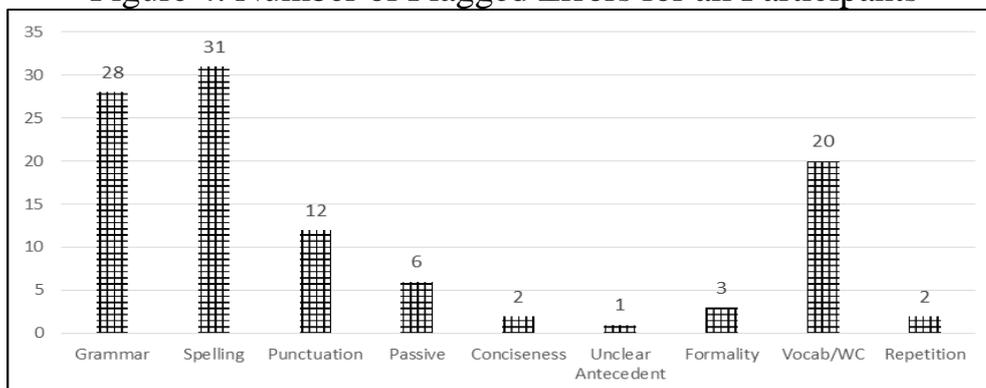
As seen in Figure 3, non-native speakers seemed to know more applications than native speakers. This is understandable as L2 learners would be less confident about their writing quality and would seek help from any available source. Interestingly, in the pre-activity questionnaire some non-native participants provided names of applications that were not on the list provided in the questionnaire (e.g., *WhiteSmoke*, *LanguageTool*, *Writesaver* and *Virtual Writing Tutor*). Equally important

was data in which all participants thought that such technological tools were useful and easy to use, which finds support in various studies (e.g., Ranalli & Yamashita, 2020). Similarly, all non-native participants believed that grammar is the most important element of good L2 writing and that correct grammar meant good scores. Such a link is understandable as these participants were trained in test taking strategies at their previous schooling.

Error flagging and Revision Behaviour

Various L2 research studies focused on examining revision behaviour occurring while students are writing (e.g., Lindgren & Sullivan, 2003; New, 1999). The second research question in this study aimed to examine and compare native and non-native participants' revision behaviour when automated error flagging from Grammarly was in place. In addition, there was also a necessity to recognize specific revision behaviour 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). In fact, the answer to this question was based on analysis of screen capture data and 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”*. 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, vocabulary and word choice, and repetition) was collected from drafts written by all participants in Grammarly in an attempt to understand the nature of flagged errors. The total number of errors flagged for all participants was 102 across all language areas (see Figure 4).

Figure 4. 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). However, a closer look at which errors were flagged for each group of participants was still needed. This is discussed in more detail in the following section in terms of the revision behaviour (acceptance/rejection/avoidance) of native and non-native participants. Data on participants' reaction to the suggestions made by Grammarly was analysed to examine which errors were *adopted*, *rejected* or *avoided*. This was undertaken in two steps, the first of which was looking at the aggregated figures of flagged errors and the nature of revision behaviour (adopt/reject/avoid) for all participants. The second step was comparing the numbers of flagged errors for each group of participants (native vs. non-native) and then examine the revision behaviour for each individual in each group. Analysis of the data revealed that participants adopted 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).

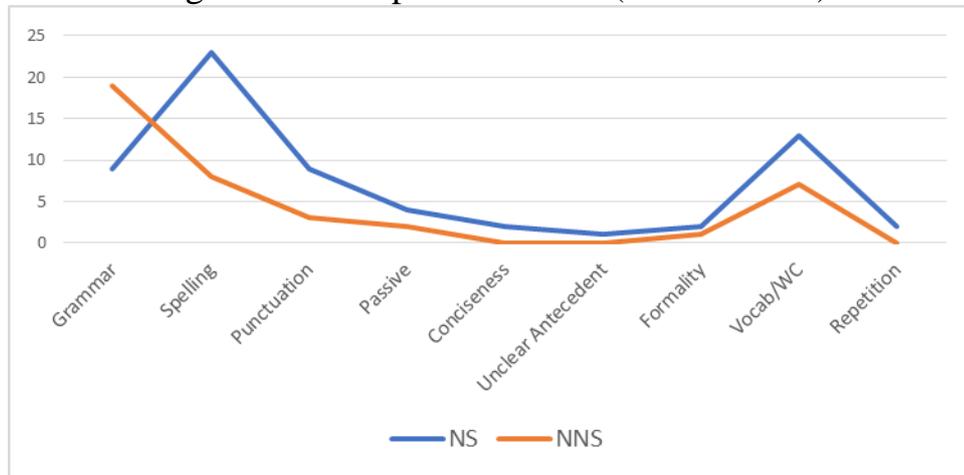
Table 2. Aggregated Figures of Participants' Revision Behaviour

Focus		Flagged Errors	Adopted	Rejected	Avoided
*Cor.	Grammar	28	28	-	-
	Spelling	31	30	1	-
	Punctuation	12	10	2	-
*Eng.	Vocab & WC	20	20	-	-
*Cla.	Passive	6	2	-	4
	Conciseness	2	2	-	-
	Unclear Antecedent	1	1	-	-
*Del	Formality	2	2	-	-
	Repetition	-	-	-	-
TOTAL		102	94	4	4
*Cor =correctness / *Eng =engagement / *Cla =clarity / *Del =delivery					

However, 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 (see Figure 5).

Error Flagging Tools and Online Revision Behaviour: Evidence from Native and Non-native Speakers

Figure 5. A Corpora of Errors (NS Vs. NNS)



Data on errors flagged for native-speaker participants was compared to non-native speakers as an initial step to understanding whether revision behaviour (adopting/rejecting/avoiding) was dis/similar. Furthermore, there was also 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.

(a) Native speakers’ Revision Behaviour

Data analysis revealed that native speakers had a total of 62 flagged errors, which were distributed among the three participants (see Table 3). Scrutiny 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. All the other suggested corrections on errors flagged by Grammarly (i.e.,N=54) were accepted (see Table 3).

Table 3. Native Speakers’ Revision Behaviour

		Participant 1				Participant 2				Participant 3			
		A	R	A	T	A	R	A	T	A	R	A	T
Language Focus		d	e	v	o	d	e	v	o	d	e	v	o
		o	j	o	t	o	j	o	t	o	j	o	t
		pt	e	i	a	pt	e	i	a	pt	e	i	a
		ed	c	d	l	ed	c	d	l	ed	c	d	l
			t	e		t	e		t	e		t	e
		e	d		e	d		e	d	e	d		
		d			d			d		d			
*Cor.	Grammar	-	-	-	-	5	-	-	5	4	-	-	4
	Spelling	3	-	-	3	19	1	-	20	-	-	-	-

Khaled El Ebyary

	Punctuati on	4	1	-	5	1	1	-	2	2	-	-	2
*Eng.	Vocab.	3	-	-	3	4	-	-	4	-	-	-	-
	W. Choice/ variety	-	-	-	-	2	-	-	2	3	1	-	4
*Cla.	Concisen ess	2	-	-	2					-	-	-	-
	Passive V	-	-	1	1	-	-	3	3	-	-	-	-
	Unclear	1	-	-	1	-	-	-	-	-	-	-	-
*Del	Tone	1	-	-	1	-	-	-	-	-	-	-	-
Total Errors		14	1	1	1	31	2	3	36	9	1	-	10

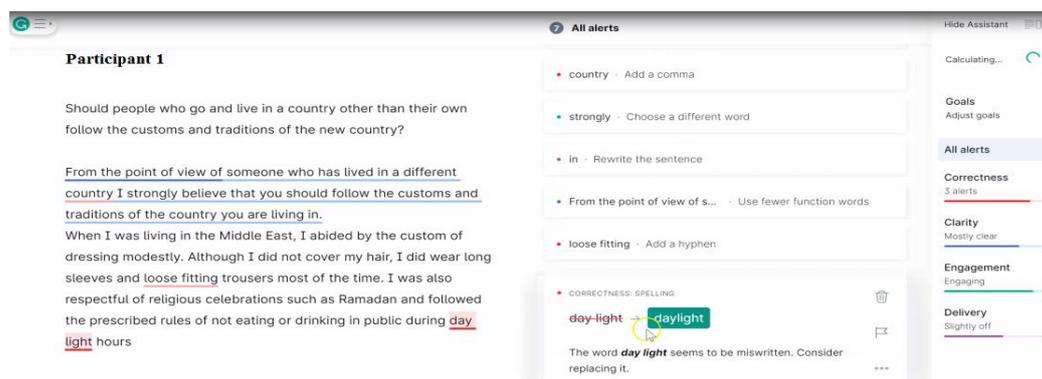
*Cor =correctness / *Eng =engagement / *Cla =clarity / *Del =delivery

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

(i) 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 6) 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.

Figure 6. Participant 1's revision behaviour after 4 min of writing

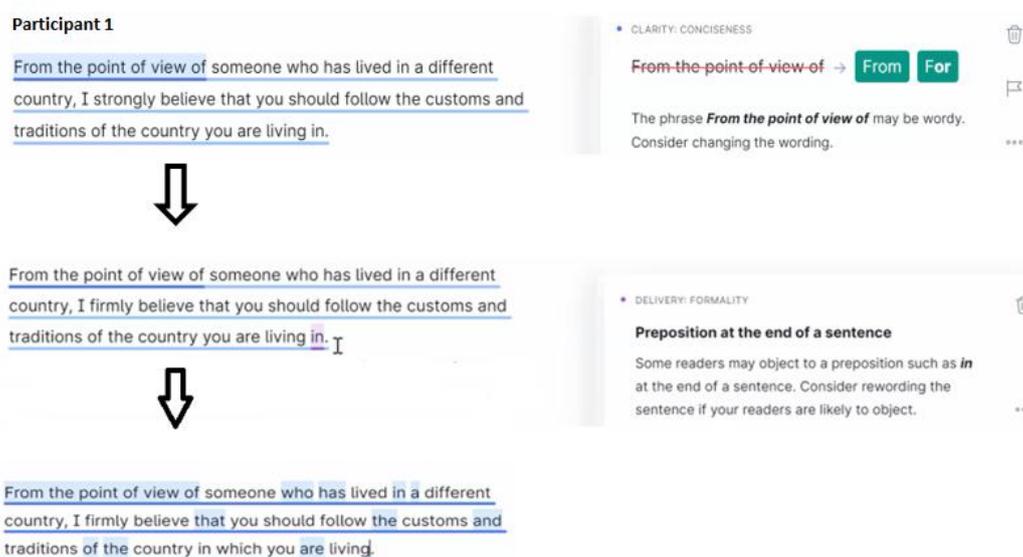


Although some sentences seemed to be correct, the ACF provided by Grammarly made this participant unsure. For example, sentences like those in Figure 7 below are good examples of the extent to which this native speaker participant took some time contemplating suggesting that she took comments seriously. This participant was hesitant about various

Error Flagging Tools and Online Revision Behaviour: Evidence from Native and Non-native Speakers

flagged errors as shown in Figure 7 below, which were left unresolved until almost the end of the writing session because she was not sure about the right form. Similarly, the suggestion made in the example shown below made the student decide to change the sentence in the end.

Figure 7. Screenshot from P1 contemplating flagged errors on sentences



It was also noticed in the example above 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 draws attention to other instances for a student who is keen to improve their writing.

(ii) 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 revising immediately and the same revision behaviour was maintained after each sentence until the end of the writing session (see Figures 8 and 9 below). However, if this observation is linked to data obtained about the type of errors flagged for each participant (see Table 3 above), 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, the examples here in Figure 8 might simply

demonstrate that these errors were just keyboard stroke errors. It is also possible that the errors are due to an habitual over reliance on the spell checker.

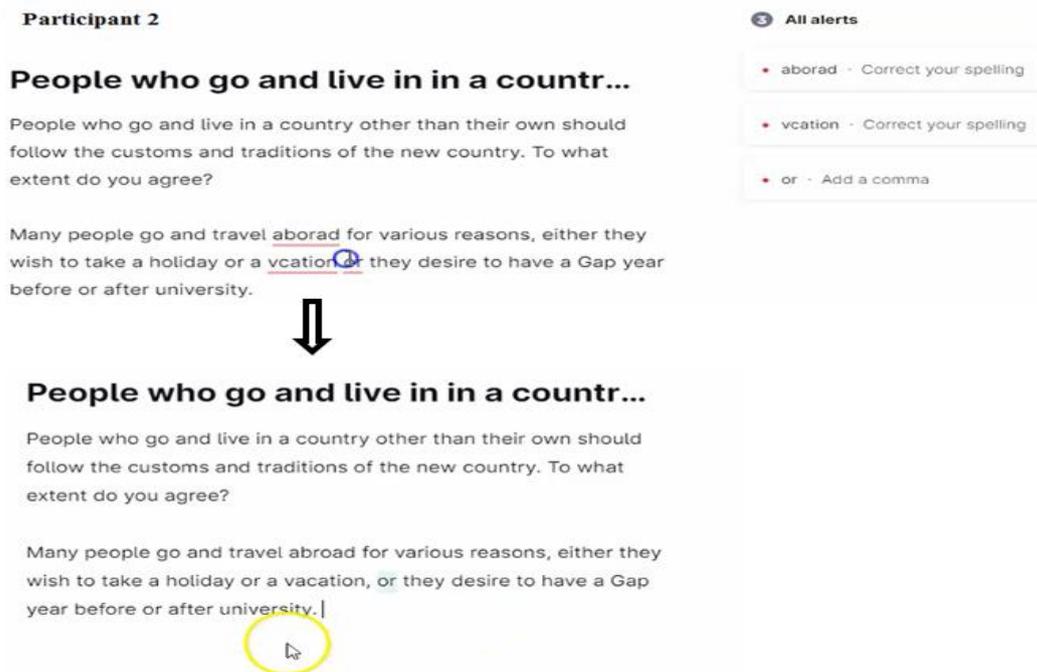
Figure 8. Example 1 of Participant 2’s instant revision of flagged errors

Participant 2

People who go and live in in a countr...

People who go and live in a country other than their own should follow the customs and traditions of the new country. To what extent do you agree?

Many people go and travel aborad for various reasons, either they wish to take a holiday or a vcation or they desire to have a Gap year before or after university.



People who go and live in in a countr...

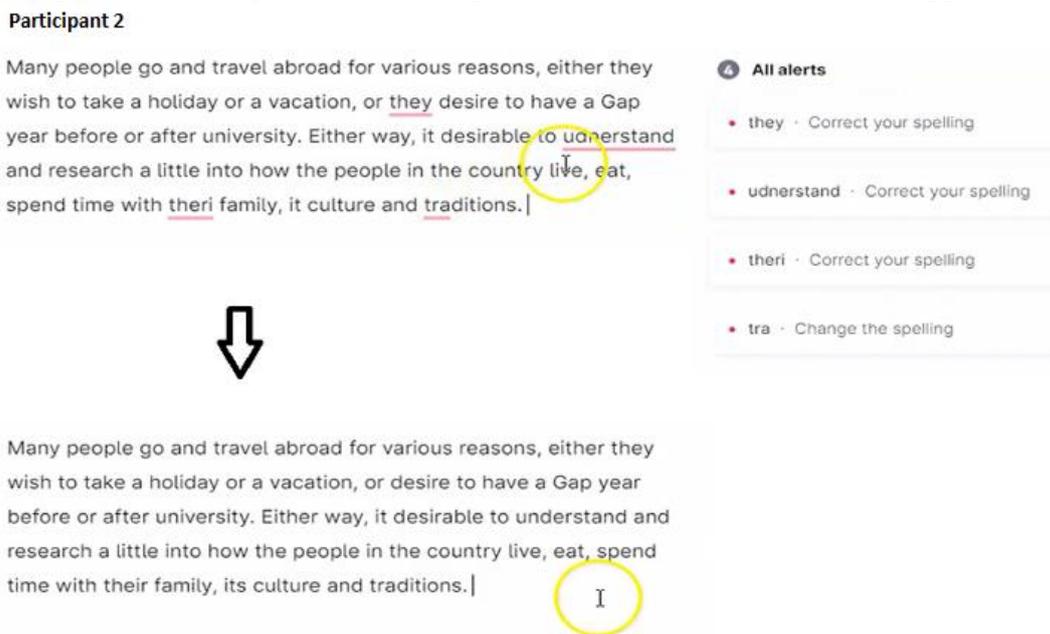
People who go and live in a country other than their own should follow the customs and traditions of the new country. To what extent do you agree?

Many people go and travel abroad for various reasons, either they wish to take a holiday or a vacation, or they desire to have a Gap year before or after university. |

Figure 9. Example 1 of Participant 2’s instant revision of flagged errors

Participant 2

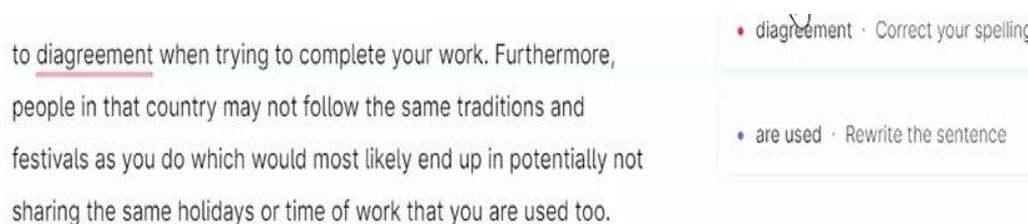
Many people go and travel abroad for various reasons, either they wish to take a holiday or a vacation, or they desire to have a Gap year before or after university. Either way, it desirable to udnerstand and research a little into how the people in the country live, eat, spend time with theri family, it culture and traditions. |



Many people go and travel abroad for various reasons, either they wish to take a holiday or a vacation, or desire to have a Gap year before or after university. Either way, it desirable to understand and research a little into how the people in the country live, eat, spend time with their family, its culture and traditions. |

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 9 above). 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 example in Figure 10).

Figure 10. Example of passive from participant 2

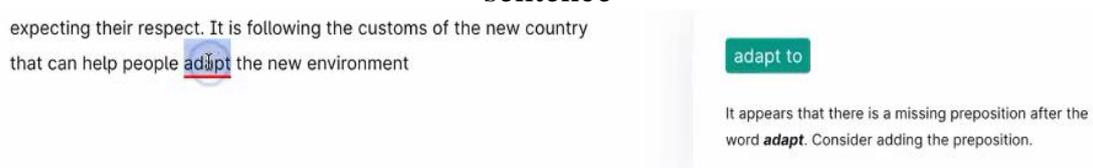


Thus, participant 2 had mainly spelling or keyboard stroke errors flagged and accepted 19 of these immediately after being flagged. He also seemed to avoid 3 passive errors to side-step such sentences being detected by Grammarly as erroneous.

(iii) Participant 3

Participant 3 was a female native speaker and also a regular Grammarly user. Like participant 2, participant 3 was concurrently writing and revising. 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 11).

Figure 11. Participant’s 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.

(b) Non-native Speakers' Revision Behaviour

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 that they trusted the automated comments provided by the system.

Table 4. Non-native Participants' Revision Behaviour

		P4				P5				P6			
		A	R	A	T	A	R	A	T	A	R	A	T
Language Focus		<i>d</i>	<i>e</i>	<i>v</i>	<i>o</i>	<i>d</i>	<i>e</i>	<i>v</i>	<i>o</i>	<i>d</i>	<i>e</i>	<i>v</i>	<i>o</i>
		<i>o</i>	<i>j</i>	<i>o</i>	<i>t</i>	<i>o</i>	<i>j</i>	<i>o</i>	<i>t</i>	<i>o</i>	<i>j</i>	<i>o</i>	<i>t</i>
		<i>pt</i>	<i>e</i>	<i>i</i>	<i>a</i>	<i>pt</i>	<i>e</i>	<i>i</i>	<i>a</i>	<i>pt</i>	<i>e</i>	<i>i</i>	<i>a</i>
		<i>e</i>	<i>c</i>	<i>d</i>	<i>l</i>	<i>e</i>	<i>c</i>	<i>d</i>	<i>l</i>	<i>e</i>	<i>c</i>	<i>d</i>	<i>l</i>
		<i>d</i>	<i>t</i>	<i>e</i>		<i>d</i>	<i>t</i>	<i>e</i>		<i>d</i>	<i>t</i>	<i>e</i>	
		<i>e</i>	<i>d</i>		<i>e</i>	<i>d</i>		<i>e</i>	<i>d</i>	<i>e</i>	<i>d</i>		
*Cor.	Gramm ar	7	-	-	7	6	-	-	6	6	-	-	6
	Spelling	4	-	-	4	3	-	-	3	1	-	-	1
	Punctua tion	3	-	-	3	-	-	-	-	-	-	-	-
*Eng.	Vocab.	-	-	-	-	-	-	-	-	-	-	-	-
	W. Choice/ variety	4	-	-	4	2	-	-	2	1	-	-	1
*Cla.	Concise ness	-	-	-	-	-	-	-	-	-	-	-	-
	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

Further analysis of individual participants was needed to understand whether this group of participants followed a specific revision behaviour pattern. As mentioned above, this group of participants had 40 errors marked 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.

(i) Participant 4

This participant was a Chinese female who is accustomed to using Grammarly as well as similar ACF applications. She 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 below in Figure 12.

Figure 12. Example of Participant 4 choosing between 2 suggested comments

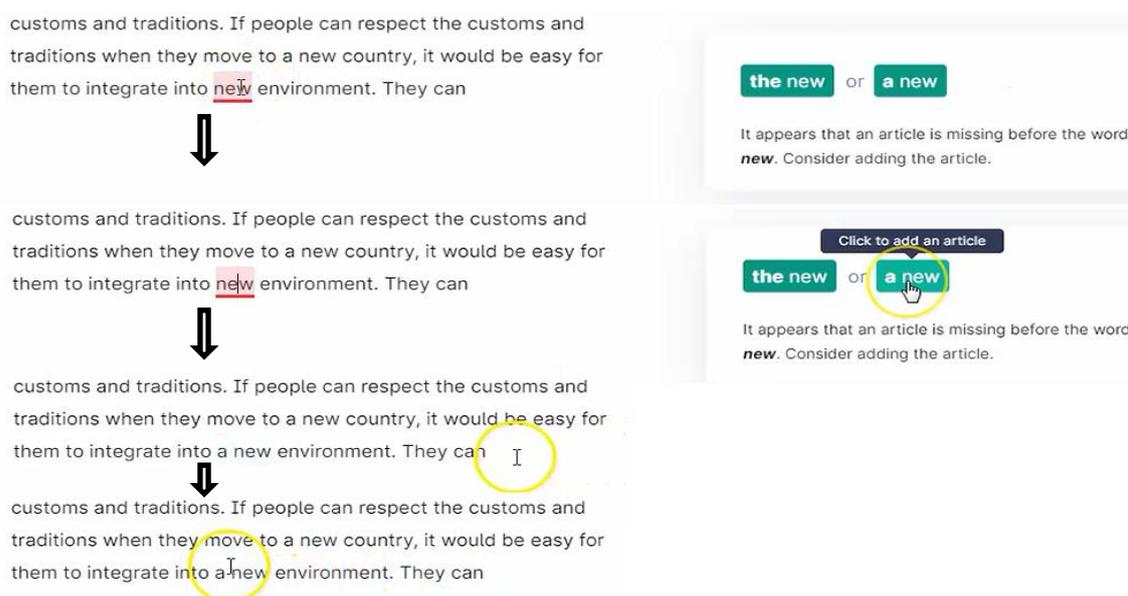


Figure 12 shows that participant 4 had to make a choice between ‘a new or the new’. The screen capture data initially revealed that she spent time reading the comments and then adopted ‘a new’ rather than ‘the new’. However, the screen capture later revealed that there was further action as seen in Figure 13.

Figure 13. Participant’s 4 manual change of previously adopted comment

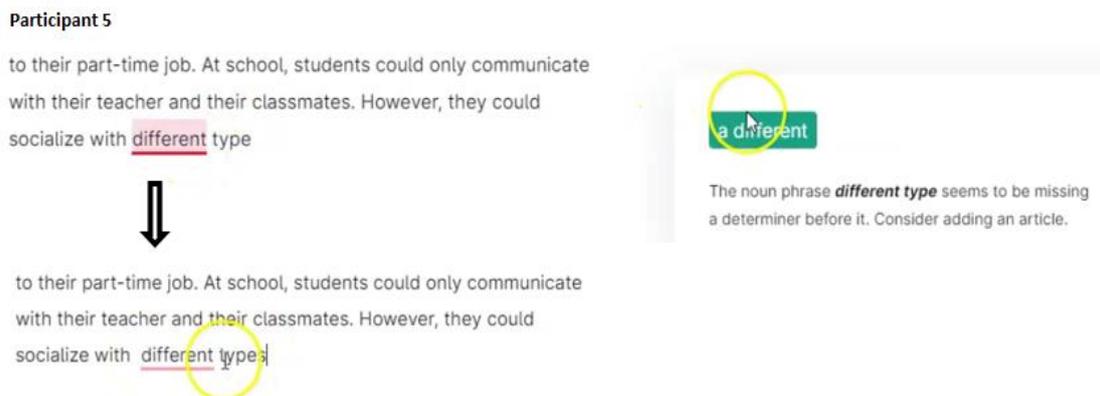
customs and traditions. If people can respect the customs and traditions when they move to a new country, it would be easy for them to integrate into the new environment. They can

Participant 4 re-visited the same sentence again and re-read the sentence reversing her choice to ‘*the new*’ instead of ‘*a new*’. This suggests that this participant did not adopt the ACF unquestioningly and she re-visited the previously adopted suggestion.

(ii) 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. 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 14 above).

Figure 14. 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 15). Participant 5 manually amended the capitalization, but she did not pick up the full stop.

Error Flagging Tools and Online Revision Behaviour: Evidence from Native and Non-native Speakers

Figure 15. Participant 5's revision of unflagged error

Participant 5

to their part-time job. At school, students could only communicate with their teacher and their classmates. However, they could socialize with different types of people. illiterate and literate people, young or old. therefore,| I



socialize with different types of people. illiterate and literate people, young or old. Therefore, working for a part-time job could be a big opportunity for them to learn how to treat and adapt themselves with others. | I

(iii) 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 ACF. 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 16).

Figure 16. Participant 6's Example Revision of Every Sentence

I agree with the view that people need to understand and adhere to local culture and traditions before entreing a



I agree with the view that people need to understand and adhere to local culture and traditions before entering a country or region. when in Roma, do as Romans do.



I agree with the view that people need to understand and adhere to local culture and traditions before entering a country or region. When in Roma, do as Romans do.

first, it is a performance that respects the cultureal

The screenshot displays three error flags from the ACF tool. The first flag, 'CORRECTNESS: SPELLING', shows the word 'entreing' being corrected to 'entouring'. The second flag, 'CORRECTNESS: CONVENTIONS', shows the word 'when' being corrected to 'When'. The third flag, 'CORRECTNESS: SPELLING', shows the word 'cultureal' being corrected to 'cultural'. Each flag includes a brief explanation and an option to add the word to a personal dictionary.

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 ACF 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.

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 ACF. 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 16).

Figure 16. Participant 6’s Example Revision of Every Sentence

I agree with the view that people need to understand and adhere to local culture and traditions before entreing a

I agree with the view that people need to understand and adhere to local culture and traditions before entering a country or region. when in Roma, do as Romans do.

I agree with the view that people need to understand and adhere to local culture and traditions before entering a country or region. When in Roma, do as Romans do.

first, it is a performance that respects the cultureal

CORRECTNESS: SPELLING
entreing → entering
The word **entreing** is not in our dictionary. If you're sure this spelling is correct, you can add it to your personal dictionary to prevent future alerts.

CORRECTNESS: CONVENTIONS
when → When
As the first word of the sentence, **when** should be capitalized.

CORRECTNESS: SPELLING
cultureal → cultural
The word **cultureal** is not in our dictionary. If you're sure this spelling is correct, you can add it to your personal dictionary to prevent future alerts.

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 ACF 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 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.

8. Discussion and Conclusion

The results obtained in this study suggested that both native speakers and non-native speakers are constantly using Grammarly, as

well as other ACF tools, to improve their written texts. The non-native participants involved in this study received more feedback on a specific language area (i.e., grammar), but both parties followed similar reaction patterns in their revision behaviour. This indicates high levels of participants' behavioural engagement with feedback from Grammarly. There is also evidence from the pre-activity questionnaire which suggested positive attitudinal engagement with Grammarly. 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 ACF. 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. 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 ACF tool (i.e., Grammarly) is useful in providing immediate feedback on surface errors is supported in this study. Many, including the current researcher, think Grammarly can 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 above (e.g., Figure 7) 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.

9. 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., & 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.
- Barkaoui, K. (2016). What and When Second-Language Learners Revise When Responding to Timed Writing Tasks on the Computer: The Roles of Task Type, Second Language Proficiency, and Keyboarding Skills. *Modern Language Journal*, 100(1), 320-340.
- Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Lawrence Erlbaum.
- 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.
- Chanquoy, L. (2001). How to make it easier for children to revise their writing: a study of text revision from 3rd to 5th grades. *Br J Educ Psychol*, 15-41. <https://doi.org/doi:10.1348/000709901158370>. PMID: 11307706.
- 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>
- Chenoweth, N. A., & Hayes, J. R. (2001). Fluency in writing: Generating text in L1 and L2. *Written Communication*, 18(1), 80-98.
- Chenoweth, N. A., & Hayes, J. R. (2003). The inner voice in writing. *Written Communication*, 20(99-118).
- 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.
- De Larios, J., Manchón, R. M., & Murphy, L. (2006). Generating text in native and foreign language writing: A temporal analysis of problemsolving formulation processes. *The Modern Language Journal*, 90(1), 100-114.
- 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., & Bleyl, 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.
- Faigley, L. a., & Witte, S. (1981). Analyzing revision. *College Composition and Communication*, 32, 400-412.
- Fitzgerald, J. (1987). Research on Revision in Writing. *Review of Educational Research*, 57(4). <https://doi.org/10.3102/00346543057004481>
- 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. (2021). 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>
- Hayes, J. R. (2004). What triggers revision? . In Linda Allal, L. Chanquoy, & P. Largy (Eds.), *Studies in writing: Volume 13. Revision: Cognitive and Instructional Processes* (pp. 189–207). Kluwer Academic Publishers.
- Hayes, J. R. (2012). Modeling and Remodeling Writing. *Written Communication*, 29(3), 369-388.
- 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>
- Kaufert, D. S., Hayes, J. R., & Flower, L. S. (1986). Composing written sentences. *Research in the Teaching of English*, 20(2), 121-140.
- 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., and , & Sullivan, K. (2006). *Analysing online revision. In Computer Keystroke Logging and Writing: Methods and Applications*. Elsevier.
- 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.
- OnlineCorrection. <https://www.onlinecorrection.com/>
- PaperRater. (2021). Retrieved 21/04/2020 from <https://www.paperrater.com/>
- 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.
- 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>
- 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>
- 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.
- Write&Improve. (2021). Provided in association with Cambridge English. Retrieved 09/11/2020 from <https://writeandimprove.com/>
- 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>.