ARTICLE

A mixed-methods study of feedback modes in EFL writing

Arif Bakla, Ankara Yıldırım Beyazıt University

Abstract

As digital technologies have become ubiquitous thanks to the Internet, new modes of feedback in L2 writing have emerged, yet what remains unclear is how feedback given through alternative modes helps improve writing quality and how new feedback tools fit in the overall context of writing instruction. Therefore, the purpose of this embedded mixed-methods study is to assess how three online feedback modes help improve student writing. Thirty-three intermediate Turkish-L1 learners of English received written, audio and screencast feedback in Google Drive to improve their writing in a multi-draft essay-writing task and an essay-revision task with three parallel essays. The results indicated that it was the audio group that made the highest number of correct revisions in the essay-writing task, while there was not a significant difference among the three feedback modes in the essay-revision task. Semi-structured interviews and screen recordings provided qualitative data about their preferences and how they worked with each mode to address both microlevel and macrolevel problems. The participants did not uniformly prefer a particular feedback mode but highlighted the potential benefits and downsides of each mode.

Introduction

Researchers and teachers have always considered feedback as an essential part of writing instruction in SLA, as it is usually intended for informing learners about the difference between their current and desired level of knowledge. According to a broader conceptualization, it informs learners not only about their learning goals but also about their performance in reaching them and what they should do next (Hattie & Timperley, 2007). However, providing feedback is time consuming, and there are many issues regarding how to provide it and even whether to provide it or not. From a theoretical perspective, several frameworks have offered support for feedback in SLA and accounted for how it could facilitate learning. For example, according to Schmidt's (1990) Noticing Hypothesis, students internalize language by consciously attending to forms in the input, and feedback could facilitate such conscious attention. In other words, it could enable them to recognize the gap between their interlanguage and the target language. Similarly, both Swain's (1995) Output Hypothesis and Long's (1996) revised Interaction Hypothesis posit that feedback assumes a critical role in L2 acquisition. The former suggests that along with comprehensible input, output is also significant as it could facilitate noticing and helps students realize gaps (Swain, 1995), and the latter highlights the importance of communication. Tracing a similar path, sociocultural theories emphasize the shared construction of knowledge through interaction, so feedback is considered as a dialogic process (Lantolf, 2006). Moreover, feedback provides learners with essential scaffolding through interactions between an expert and a novice and allows for interaction within learners' zone of proximal development. In short, all of these theoretical outlooks consider corrective feedback as an essential component of language acquisition.

Nourished by such theoretical frameworks, a large body of research has been carried out so far to study

various aspects of written feedback (WF), such as how direct it should be (direct or indirect), the component of writing for which feedback is provided (form or content and macrolevel or microlevel issues), its source (teacher, student, or computer), audience (various proficiency levels, ESL, or EFL learners), and direction (corrective or positive) (Ferris, 2003). Under the influence of recent technological developments and insights from interactionist and sociocultural theories, which underscore the dialogic nature of feedback, writing feedback research has taken more of an interest in technology-enhanced modes of delivery. This movement has recently enabled researchers to experiment with online feedback modes that could offer such benefits as increased interaction, multimodality, user-friendliness, speed, and so forth. Among alternative modes are audio feedback (AF), which dates back to the 1990s, when it was mostly provided as a part of research studies (Boswood & Dwyer, 1996; Kirschner, van den Brink, & Meester, 1991), and video recordings called screencasts, in which audiovisual information is presented.

However, although it seems quite easy to provide audio feedback (AF) or screencast feedback (SCF), neither has been a regular part of learning environments, and research on these modes is limited. In fact, the idea of providing audiovisual feedback is so new that AF or SCF was not included in a meta-analysis on feedback, carried out earlier this decade (Biber, Nekrasova, & Horn, 2011). We have many more studies today, yet they predominantly investigate student and faculty perceptions (Alvira, 2016; Ghosn-Chelala, & Al-Chibani, 2018; Orlando, 2016). Although such studies are invaluable in understanding the nature of digital modes and how students engage with them, they fall short in testing student performance. Moreover, some studies suffer from limitations or methodological problems. For instance, the conclusions of some studies on SCF (Hynson, 2012; Séror, 2012) seem to rest upon personal observations rather than results of rigorous statistical analyses. Another problem is that most of these studies fail to provide an interrater reliability for their ratings, which, according to Ferris (2003), might undermine the quality of the findings. Finally, to date, very few studies (Orlando, 2016) have integrated all three of these feedback modes in a single study.

In brief, due to insufficient research with limitations and conflicting findings, there is still need for research on digital feedback modes. Calls for further research (Chang, Cunningham, Satar, & Strobl, 2017; Elola & Oskoz, 2016; Harper, Green, & Fernandez-Toro, 2015) seem to confirm this. In line with this need for research, the present study adopts a mixed-methods embedded research design to test the impact of digital feedback given in three different modes (written, digital audio and screencast) on learners' writing quality and to understand their experience of receiving feedback in these digital modes.

Literature Review

In the quest for effective and time-efficient feedback since the 1990s, a particular area of interest has been the mode in which feedback is provided (paper-based, oral, audiotaped, digitally written, digital audio or SCF). While the most widely used mode has always been paper-based written comments, an emergent mode involved the use of audiocassettes (Boswood & Dwyer, 1996; Huang, 2000), and such studies mostly produced promising results. Later, electronic feedback was incorporated into the discussion (Tuzi, 2004). In most studies, this new written mode involved comments provided and received in an electronic environment, such as email and online or offline software. Multimodal feedback and the introduction of audiovisual elements into feedback were also a significant milestone in feedback research. For instance, a combination of audiotaped and written comments was found to be better than written-only comments (Huang, 2000), and WF followed by five-minute conferences produced better results with respect to specific grammar corrections (Bitchener, Young, & Cameron, 2005).

Using a more modern method, Li and Akahori (2008) provided WF to 84 Chinese students, who were learning Japanese, through tablet software by using a pen tool in one group, and additional simultaneous AF accompanied by WF in another. They compared these two feedback conditions in lower and upper proficiency levels. The results indicated that multimodal feedback proved effective in lower groups but not in the upper ones because the latter found the notes redundant. The researchers concluded that drawings and notes on students' writing partially facilitated memory and the AF increased the social presence of the teacher. This study was one of the earliest examples for the incorporation of audiovisual

feedback in L2 research.

A more developed version of such feedback is provided using a screencast, which is defined as a "combination of audio and video recording that could be played in a browser" (Edwards, Dujardin, & Williams, 2012, p. 95). Screencasting involves the recording of the instructor's voice, along with the screen, where the written work is seen and commented on using screen recording software, such as Jing or Screencast-O-Matic (see Bakla, 2017 for an overview). While recording the voiceover, the instructor is able to comment on or highlight certain portions of student writing. Therefore, Lee (2017) considers SCF an improvement over AF as it is something between WF and face-to-face writing conferences (p. 131). The recorded video is usually shared using a web-based tool, allowing the learner view the feedback as many times as desired. The perceived benefits of more recent modes are not limited those mentioned here. The next section gives an overview of potential affordances of these modes, particularly those of screencasts.

Affordances of Audiovisual Feedback

Multimodal feedback is deeply rooted in the idea that it could cater not only to learning styles but also to students' preferences (Mayer, 2003), while too much WF is considered to be biased towards a single learning style (Stannard, 2008). Multimodal corrective feedback could be especially influential for those with auditory/visual learning styles, increasing intake of feedback, or as Cranny (2016) notes, it could affect learners' preference of feedback modes. For example, in a recent study (Ghosn-Chelala & Al-Chibani, 2018), Arabic-L1 learners of English reported that receiving multimodal feedback accommodated their preferences and therefore boosted their engagement; they found the feedback clearer and more valuable.

Another major aspect of audio or audiovisual feedback is related to the term social presence, which is defined as the extent to which participants in a community could present themselves "socially and emotionally, as 'real' people, through the medium of communication being used" (Garrison & Anderson, 2003, p. 50). Current research suggests that SCF increases the social presence of the teacher and is engaging (Harper et al., 2015). This is true of audio feedback as well (Ice, Reagan, Perry, & Wells, 2007), but not of WF. As Chang et al. (2017) note, audiovisual feedback, promotes proximity between the teacher and student, so it could help improve the quality of student writing. For instance, according to the participants in Ducate and Arnold's (2012) study, the teacher was considered more caring in SCF (cited in Elola & Oskoz, 2016). Moreover, while screencasts help increase social presence, they decrease possible pressure caused by face-to-face conferences (Lee, 2017). Such feelings are critical, as they could help establish rapport. As with SCF, students who received AF in several studies also reported that the teacher cared for them and attached importance to their work (Cavanaugh & Song 2014; Ice et al., 2007; Sipple, 2007). In a more recent study, Ali (2016) worked with 63 mixed-level students to compare WF and SCF and obtained findings in favor of the latter. The participants thought that SCF was "personal, specific, supportive, multimodal, constructive, and engaging" (p. 131).

In SCF, learners cannot directly communicate with the teacher unless they respond to the feedback through another screencast or through written or audio comments. However, due to perceived elaboration and proximity, some studies reported improvement in the quality of interaction (the extent to which the instructor and students could freely communicate to address problems in writing; Harper, et al., 2015; Silva, 2012). Furthermore, in combination with written comments, screencasts could better lend themselves to communication and help create a learning community (Mathieson, 2012) by helping teachers establish a dialogic learning environment (Harper et al., 2015). This was supported by Orlando (2016), who investigated the perceptions of six instructors and 30 students about feedback modes. The instructors preferred SCF to audio or written comments, thinking that AF and SCF helped them deliver more information (particularly the latter) and were more personal. In fact, as found in Silva's (2012) study, screencasts look like a simulation of face-to-face meetings. Finally, both AF and SCF could help improve listening skills. Moreover, when students respond to the feedback orally or provide AF or SCF for peer review, they could also develop their speaking skills.

In addition, the feedback tool could affect the quantity and quality of the instructor's comments. In a case study with four advanced English-L1 learners of Spanish, Elola and Oskoz (2016) provided digitally written feedback and SCF. The results indicated that the instructor provided more comments

on global issues in the SCF and fewer in the WF. Screencasts seemed to offer in-depth explanations, particularly on global issues. Data from some other studies also supported this (Moore & Filling, 2012; Orlando, 2016; Vincelette & Bostic, 2013; Silva, 2012). This is because global issues necessitate further elaboration (particularly on how to address them), which is easy to do in SCF. This is also true for AF, as instructors in Cavanaugh and Song's (2014) study showed. Feedback in that study tended to focus on more global issues. Such comprehensive feedback, in turn, was found to be better in establishing rapport and creating a sense of support for students than WF (Thompson & Lee, 2012).

SCF seems to offer benefits with respect to comprehensibility due to in-depth explanations, multimodality and ability to watch it at the desired pace. Due to greater elaboration, students in various studies thought that SCF was more comprehensible (Ducate & Arnold, 2012 cited in Elola & Oskoz, 2016; Harper et al., 2015). Some students even found it better than writing conferences as they could watch screencasts as many times as they wished (Harper et al., 2015). Still another reported benefit of SCF was deeper learner engagement and more autonomy (Hynson, 2012). It is generally reported that SCF leads to positive attitudes in students and teachers (Alvira, 2016; Ali, 2016; Ducate & Arnold, 2012 cited in Elola & Oskoz, 2016; Edwards et al., 2012; Harper et al., 2015; Mathieson, 2012; Moore & Filling, 2012; Mitchell, 2012.; Orlando, 2016). Researchers usually attributed such positive attitudes to SCF's being more engaging, motivating, interactive, and individualized than WF.

In short, research suggests that SCF could guide learners better in drafting their essays, although such claims, in most cases, do not rest upon rigorous empirical findings. However, there were also some contradictory findings regarding learners' preferences. For example, in Moore and Filling's (2012) study, the participants preferred SCF to WF but not to individual writing conferences. In Orlando's (2016) study, although the students mentioned certain benefits of audio and SCF feedback, their preferences did not match this ranking. That is, nearly two-thirds of the students ranked WF as their first preference and the rest preferred SCF; audio did not rank first for any of the students. Orlando attributed this mismatch to the older age of the students, who were accustomed to receiving written comments.

Studies on digital feedback modes have also reported some problems the participants experienced, as successful incorporation of innovative tools in daily learning activities depends on learners' quitting longstanding habits (Elola & Oskoz, 2016; Lin & Yang, 2011), such as being accustomed to single-draft writing, failing to write collaboratively, and unwillingness to provide and receive feedback. Furthermore, previous experience in using digital tools is also essential for successful integration into feedback practices. Moreover, students might not be motivated enough to work with new software because, as found in Orlando's (2016) study, it seems cognitively less demanding to make revisions based on paper-based comments.

Effectiveness of Feedback Modes

Studies on the effectiveness of feedback modes usually discussed it with respect to microlevel (linguistic) versus macrolevel (discourse) issues. However, despite a number of benefits of multimodal feedback reported above, studies have produced partially inconclusive results as to the effectiveness of each mode. Though not empirically based, some studies suggested that electronic feedback was more effective in macrolevel corrections than in microlevel ones (Morra & Asis, 2009; Tuzi, 2004). Others, usually based on perceptions, extended this by stating that screencasts lent themselves better to macrolevel issues, such as organization and content (Ali, 2016; Elola & Oskoz, 2016; Ice et al., 2007; Silva, 2012). This was largely because it proved easier to provide students with lengthier comments on macro issues in screencasts. Contrary to such findings, the results of an experimental study by Mitchell (2012) indicated that SCF helped the participants fix microlevel issues at statistically significant levels, compared to traditional feedback. On the other hand, feedback mode had no impact on the quality of students' revisions in other studies (Ducate & Arnold, 2012 cited in Elola & Oskoz, 2016; Elola & Oskoz, 2016). In short, while the novelty introduced by newer feedback modes has predominantly produced positive attitudes in learners and instructors, this has not always led to improvements in students' writing performance.

Research Questions

Building upon the literature review presented above, this study seeks to answer the following research

questions:

> *RQ1.* Which digital feedback modes (electronically written, audio, screencast) could help the participants perform a higher rate of successful revisions at the microlevel, macrolevel, and global level in a multi-draft essay-writing task?

> *RQ2.* Which digital feedback modes (electronically written, audio and screencast) could help the participants perform a higher rate of successful revisions at the microlevel, macrolevel, and global level in revising the essays supplied to them?

> *RQ3.* What are the participants' preferences of the feedback modes (electronically written, audio, screencast) and what factors could account for these preferences?

> *RQ4.* How did the participants engage with each feedback mode and interact with the researcher?

The next section elaborates on how the quantitative and qualitative phases of the research were carried out, how the data was analyzed and how the findings from the two strands were mixed.

Method

The researcher used a mixed-methods embedded design to test the effectiveness of three digital feedback modes and explore the participants' views of these modes and understand the processes involved while receiving digital feedback. The feedback modes compared in this study were as follows: (1) SCF was provided using screencasting software; (2) AF was provided using a free audio add-on (Kaizena) in Google Drive and (3) WF was provided as marginal comments in Google Drive. After the experiment, a purposefully selected group of participants (*N*=6) were interviewed to "obtain more detailed, specific information than can be gained from the results of statistical tests" (Creswell, 2012, p. 535). Moreover, the participants were requested to record their screen while revising the essays through a think-aloud procedure to triangulate the interview data (Figure 1).
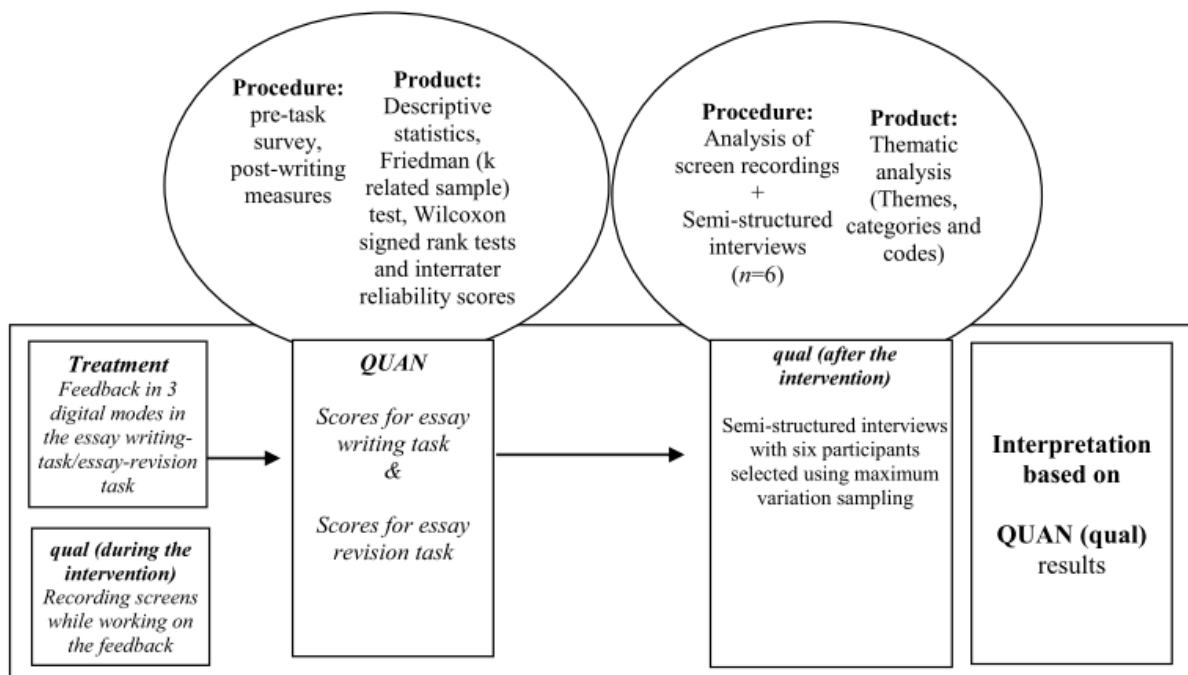

Figure 1. Visual model of this mixed-methods research design.

The participants completed two sets of tasks: (1) writing a multi-draft cause-effect essay and (2) revising three parallel essays previously written by other learners. The participants did both tasks at home, and they were given one week for each draft and revision. For the essay-writing task, upon receiving written peer feedback for the first draft (in Google Drive), 33 learners were divided into three groups and they received feedback through one of the three modes for their second draft, and then received feedback through another in the third. For the essay-revision task, they were randomly divided into three groups,

and they were asked to revise three parallel essays, which were highly similar in (a) their level and length (300, 302 and 306 words respectively), (b) the microlevel and macrolevel problems that the learners were supposed to work on in each of them and (c) the prompts given to the students (Appendix A, Table 4). For each essay, each group received feedback through a different mode (Figure 2).

The essay-revision task was particularly selected due to several reasons. First, using a crossover design, the researcher attempted to minimize, if not fully eliminate, the impact of potential differences not only between the parallel essays but also between the groups of participants or possible variation in the time they spent for revising the essays. Secondly, it helped increase the number of cases to be compared in each group using inferential statistics, thanks to the crossover design. Thirdly, this task enabled the participants in each group to work with the mode that they did not experiment with in the first task. In this sense, the two tasks complemented each other.
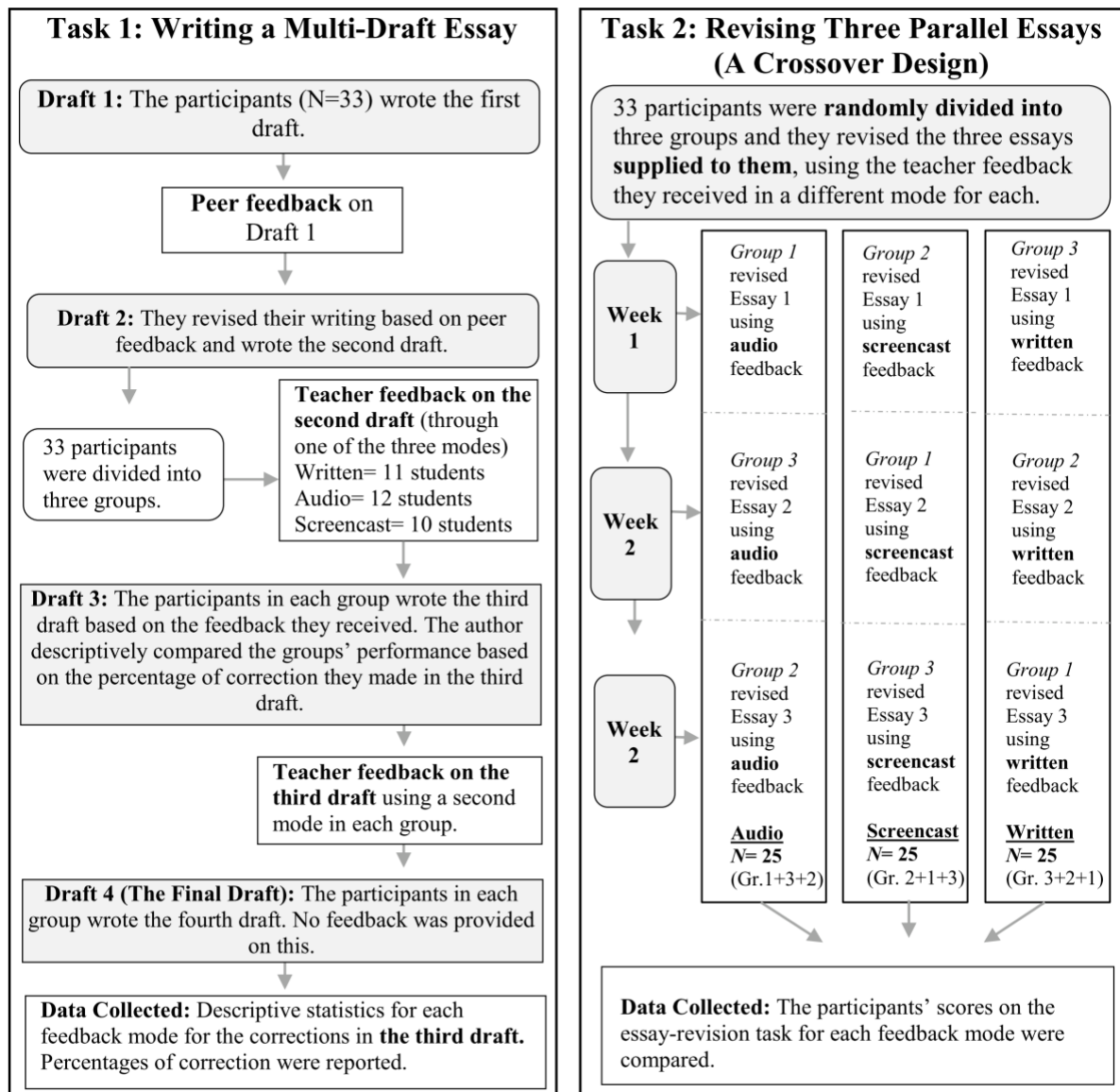
Notes. (1) In the essay-writing task, the participants received feedback on the third draft in a mode other than the one in which they received feedback in the second draft. (2) In the essay-revision task, a crossover design was used, so that all the participants could work with all three modes of feedback and the possible differences between the students or essays could be controlled. 8 people failed to complete at least one of the revisions. Therefore, only those who revised all three essays were included in the analysis, which amounted to 25 participants in each group. (3) Globally considered, the participants received feedback in two of the three modes in the essay writing tasks, while they experimented with all the three modes in Task 2.

Figure 2. The procedures for feedback, drafting and revising in the two tasks.

Participants

This study was carried out in a writing class with 39 English-major freshmen (eight males and 25 females) with an age range between 18 and 20. Thirty-three of these agreed to participate in the study. The participants were intermediate learners of English, studying in the first year of an English language teaching (ELT) department in Turkey, where they began to study English after passing an intermediate proficiency test at the end of a preparatory class. They had studied academic writing for six hours a week during the previous academic year (in the preparatory English class), during which they had been predominantly instructed through a product-oriented approach. They had mostly received paper-based WF or sometimes attended very short writing conferences. The data from the pre-task survey revealed that the majority of the participants (88.2%) had access to a mobile phone with internet connection. In addition, 85.3% of them also had Internet connection at home. All the participants were provided video training on how to use Google Drive and the AF tool. All the feedback in this study was provided by the researcher. The regular class teacher helped the researcher create a schedule, contact the students and notify them of what they had to do.

As participants with dissimilar essay scores could provide richer data for qualitative analysis, the researcher recruited interview participants with heterogeneous (lower, average, higher) essay scores and of different genders. Based on these criteria, the researcher recruited six interview participants using criterion sampling, which refers to the practice of identifying some criteria and selecting participants that differ markedly from each other on these criteria (Creswell, 2013, pp. 156-157). For a participant to be able to participate in the interview, he or she must have written all the drafts and must be willing to participate.

Data Collection

The researcher collected data using both quantitative and qualitative means, with priority given to the former. The qualitative data were collected during the experiment to explore how the participants worked with the feedback modes and immediately after it to understand their perceptions of these modes: QUAN(qual). The data collection tools included the following: a pre-task survey, essay drafts/revisions, semi-structured interviews (Appendix B), and screen recordings. The pre-task survey was intended to collect data about age, gender, digital device ownership, Internet access and so on.

Feedback Procedures

The feedback was provided to address both linguistic problems and issues related to content and organization. A two-step method was followed to select the issues to focus on. First, a list of issues for feedback (Ferris, 2003) was used as a point of reference. Second, some of these issues were identified for use in this study based on (a) the teaching experience of the researcher, (b) the availability of the issues in all three essays and (c) a balance between macrolevel and microlevel issues. The researcher selected the most appropriate three from a group of essays based on teaching experience and the availability of the issues identified by Ferris (Table 1).

Table 1. *The Issues Addressed in the Essay-Revision Task*

Macrolevel Problems	E1	E2	E3	Microlevel Problems	E1	E2	E3
Lack of or Inappropriate title	1	1	1	Verb-tense error	1	1	-
Lack of an appropriate hook	1	1	1	Subject-verb agreement error	2	1	1
Inability to connect the hook and the thesis statement	1	1	-	Active or Passive error	1	1	1
Lack of or inappropriate thesis statement	1	1	1	Word order error	1	1	-
Need for clarification or additional information	2	1	1	Word form error	2	1	1
Logical fallacy and arguments too strong to be true	1	1	3	Word choice error	1	2	2

Poorly developed paragraph(s)	1	1	1	Plurality error	1	1	1
Lack of support	1	-	-	Preposition error	-	1	-
Lack of unity	1	1	1	Fragment	1	1	1
Lack of cohesion or coherence	3	5	2	Comma splice	1	1	-
Lack of or inappropriate topic sentences	1	-	1	Other punctuation problems	1	1	1
Repetition	-	-	1	Articles	1	1	2
Total	14	13	13	Total	13	13	11

Note: E indicates essay

In the essay-writing task, the feedback mechanism involved indirect feedback because research suggests that direct feedback could be more appropriate for beginners, who could hardly benefit from self-correction (Ellis, 2009; Kang & Han, 2015). The feedback itself was considered indirect as it provided no correct answers, but involved the provision of "metalinguistic information" (Godwin-Jones, 2018, p. 7). The indirect feedback involved talking about errors in the AF, talking about it in addition to highlighting the problems by using comment and markup tools in the SCF, and writing short explanations or coded feedback in the WF. A peer provided feedback for the first draft by using a peer-review form (Click here to see the feedback form). The researcher provided feedback on the second and third drafts (Figure 2).

For the essay-revision task, the participants were provided feedback to help them revise the three parallel essays that encompassed issues related to both content and form. For each essay, they received feedback through a different mode (See Figure 3 for the screenshot of the AF. Refer to Appendix A for the other modes). All the feedback in both tasks was provided by the researcher, rather than the regular class teacher, and the learners were assigned grades for the revisions they made in both tasks.
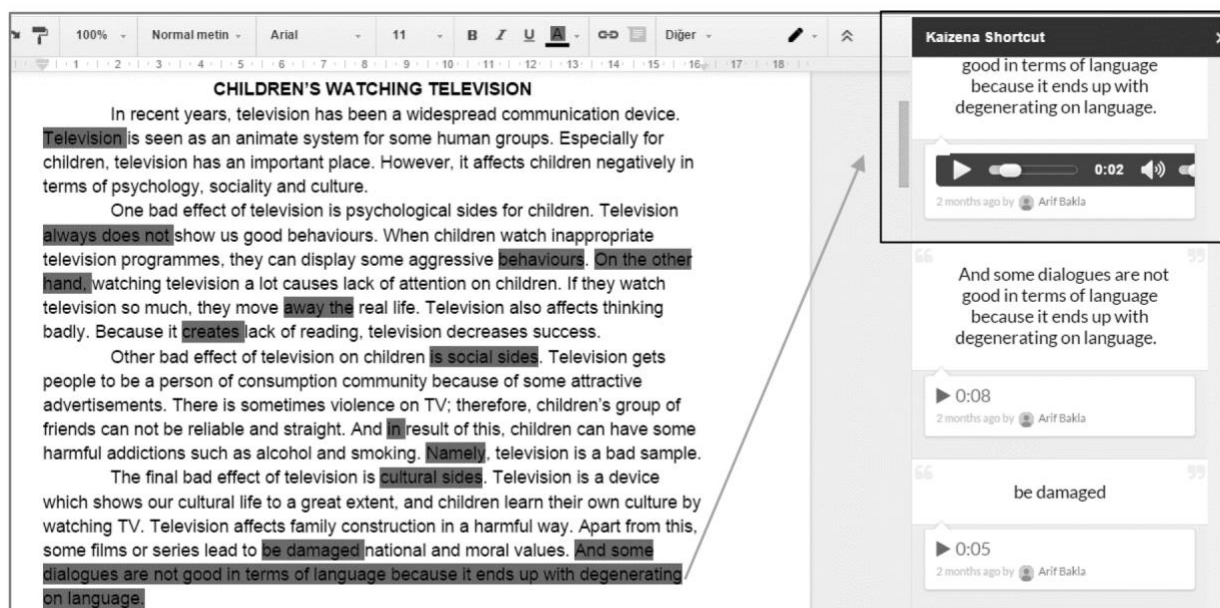
Note. The learner could click on the highlighted area in the text to see the related audio comment in the panel on the right (the previous interface of the tool).

Figure 3. A screenshot of the AF provided in the essay-writing task.

Data Analysis

Quantitative Data Analysis

An analytical scoring procedure was used to calculate descriptive statistics for the essay-writing task. That is, the macrolevel and microlevel issues provided in Table 1 were checked and the number corrections were calculated (Appendix D). Similarly, to compare the participants' performance in the

essay-revision task, a revision score for each feedback mode was calculated for the parallel essays for each student by assigning "1" for every correct revision and "0" for every incorrect one or each lack of revision. A correct revision referred to a participant appropriately addressing the feedback given. The items which did not appear in all three essays were disregarded. For the items with more than one instance in the same essay, only the first appearance was taken into consideration, except for "lack of cohesion/coherence" which occurred at least twice in each essay. The resulting calculation provided a score out of 17. The revision scores were compared using Friedman's *k* (related samples) followed by Wilcoxon signed-ranks tests for post-hoc comparisons.

To validate the scoring procedure, another rater scored each revision in the essay-revision task to calculate the interrater reliability using Cohen's *kappa*. The researcher made sure that all assumptions of this statistic were met. The mean interrater reliability was found to be $K = 0.87$ (p <.0.001), 95% CI (Items=.75, .71, .78, .67, .95, .70, .92, .86, .93, .91, 1.00, 1.00, .96, .93, .94, .83, 1.00), indicating that the raters were in good agreement.

Qualitative Data Analysis

The interview recordings were transcribed and analyzed in Nvivo v10. The analysis approach was inductive content analysis as no predetermined codes were used to guide it. Both frequency and prevalence were taken into account to create the themes. The researcher identified the relationships between the nodes based on preliminary analyses and tested these against the data. Some codes and relationships decayed on the way. For instance, most of the coding initially went around the feedback modes. It seemed that each mode would make a category or a theme, but later some apparently important qualities of these modes came forward (e.g., comprehensibility, multimodality, and interactivity). During the coding, the researcher continuously created annotations and memos to keep track of ideas and facilitate the analysis (See Appendix C for notes on how the researcher used them). In line with Patton's (2002) evaluation criteria for checking codes and categories, the researcher examined homogeneity within each code and category and heterogeneity across them to ensure that everything in a category fits together without overlapping and that content is different across the categories. The exemplar quotes were selected based on their representativeness of the themes.

Results

Quantitative Results

This section gives an overview of the results obtained from the essay-writing and essay-revision tasks. As noted earlier, microlevel issues include grammar and vocabulary, punctuation and so forth, while macrolevel ones are related with the main elements of an essay (e.g., the title, hook, and thesis statement), essay organization (e.g., paragraph development, writing topic sentences, unity, coherence or cohesion) and clarity of expression. *Global level* is used to refer to a combination of both.

Results from the Essay-Writing Task

For the essay-writing task, a correction score was calculated for each level, as mentioned above. The scores were the highest in the audio group, and there was less variability in this group, while the screencast group received the lowest scores both at the macrolevel and global mean scores. The groups performed more similarly on the microlevel issues, while there was a larger difference between the highest and lowest score on the macrolevel scores (Table 2).

Table 2. *Descriptive statistics for the essay-writing task across the three modes*

Group	N	Macrolevel Correction Score (%)	Microlevel Correction Score (%)	Global Correction Score (%)
Audio	12	80.00	80.00	80.00
Screencast	10	56.00	71.00	65.68
Written	11	67.00	73.00	71.33

For a few items (inappropriate title, lack of support or lack of cohesion), the lowest scores were in the

SCF group. Items in the categories unclear ideas; poorly developed paragraphs; and lack of an appropriate title, hook, or thesis statement had the lowest correction scores. For microlevel issues, the participants were able to correct most of the punctuation problems and errors related to subject-verb agreement, pronoun agreement, and plurality. Wrong word choice and preposition errors were among the items with lower correction levels (Appendix D, see Table 5 & 6).

Results from the Essay-Revision Task

In the essay-revision task, compared with macrolevel problems, microlevel problems were more frequently corrected (Table 3), regardless of the feedback mode. However, the audio group corrected fewer microlevel errors than the screencast and written groups. The problem of "fragments" was at the lowest end of the correction scale among the microlevel issues for the three groups, while the lowest score belonged to "poorly developed paragraphs" among the macrolevel ones, and "word form errors" received the lowest correction score in the audio group (52%). That was also the lowest score among the three groups for all of the microlevel issues.

Table 3. *Percentages of correction across the feedback modes in the essay-revision task*

	Issue Addressed	Audio	Screencast	Written
	Inappropriate title	84.00	87.50	78.13
	Lack of an appropriate hook	40.00	50.00	50.00
	Lack of or inappropriate thesis statement	56.00	50.00	50.00
Macrolevel	Need for clarification	48.00	78.13	56.25
	Logical fallacy and strong arguments	64.00	50.00	65.63
	Poorly developed paragraphs	20.00	15.63	15.63
	Lack of unity	80.00	87.50	81.25
	Lack of cohesion or coherence	48.00	45.31	54.69
	Overall correction rate	*54.22*	*56.60*	*56.25*
	Subject-verb agreement error	80.00	90.63	84.38
	Active or Passive error	76.00	96.88	96.88
	Word form error	52.00	84.38	90.63
	Wrong word choice error	80.00	78.13	84.38
Microlevel	Plurality error	76.00	81.25	84.38
	Fragments	72.00	78.13	71.88
	Punctuation problems	80.00	87.50	68.75
	Article error	88.00	90.63	84.38
	Overall correction rate	*75.50*	*85.94*	*83.20*

To compare the participants' global scores for the three modes in the essay-revision task, a non-parametric Friedman test of differences among repeated measures was conducted and rendered a Chi-square value of 4.20, which was not significant, $\chi_2(2) = 4.20$, $p = .12$. This finding indicated that none of the modes was more effective than the others in improving the participants' global revision scores. Two more Friedman tests of differences among repeated measures were conducted to compare the macrolevel and microlevel essay-revision scores. The test was not significant for macrolevel revisions, $\chi_2(2) = 1.14$, $p = .57$, while it was significant for microlevel ones, $\chi_2(2) = 6.11$, $p = .047$. To find out which feedback mode was more effective in improving microlevel corrections, post-hoc analysis with Wilcoxon signed-rank tests was conducted with a Bonferroni correction applied, resulting in a significance level of $p < .017$. Median (IQR) for the audio, written and SCF were 6(2), 7(2) and 7(2), respectively. There were no significant differences between audio and written mode ($Z = -1.75$, $p = .08$), between the SCF and WF ($Z = .20$, $p = .84$) and between the AF and SCF ($Z = -2.19$, $p = .03$).

In short, the audio group performed a higher rate of successful revisions for all the three levels in the essay-writing task (RQ1). However, the inferential statistics calculated for the revision scores from the essay-revision task globally suggested that no feedback mode was more effective than the others in helping the learners to revise a given piece of writing (RQ2).

Qualitative Results

The Participants' Preferences of the Feedback Modes

The analysis of the qualitative data produced three themes, revealing valuable information about the nature of the feedback modes, particularly the participants' preferences and the procedures they followed while working with the feedback.

Theme 1. The respondents were positive about digital feedback despite their initial anxiety.

Globally considered, the participants had a highly positive attitude towards receiving feedback in an online environment. Five of them made 20 positive references about it (count = 5, reference = 20). They expressed their willingness to receive digital feedback as they thought it helped them to see the problems with their writing. However, some reported that they initially felt anxious, and although not particularly requested, they usually compared the level of their anxiety across time using comparative expressions like "initially,… but later…" or "We had been anxious before…". What caused this change in anxiety level was investigated by looking for traces of variables that might initially have caused anxiety. A code was created for the expressions that might signal anxiety or alleviation of it. Later, a list of words that signaled change was made, and the larger contexts of these words were examined. This analysis indicated that their anxiety alleviated as they got used to working with the digital modes and recognized their value. For example, IP01, who was representative of three other respondents, was initially concerned with task completion only. But later, having to deal with lots of tasks and seeing the tools' user-friendliness increased his/her motivation because he or she felt able to improve gradually: "... I initially perceived it as an assignment only, a task to be accomplished. Frankly, at first, I didn't see it as a means for learning" (IP01). Qualitative data revealed that using a different mode to receive feedback led to short-term anxiety (Figure 4).
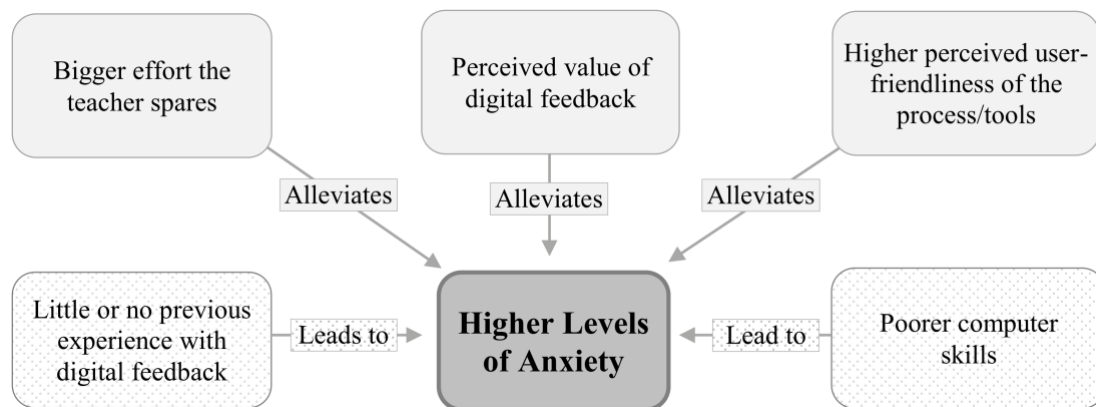

Figure 4. Intervening variables that caused and helped alleviate anxiety.
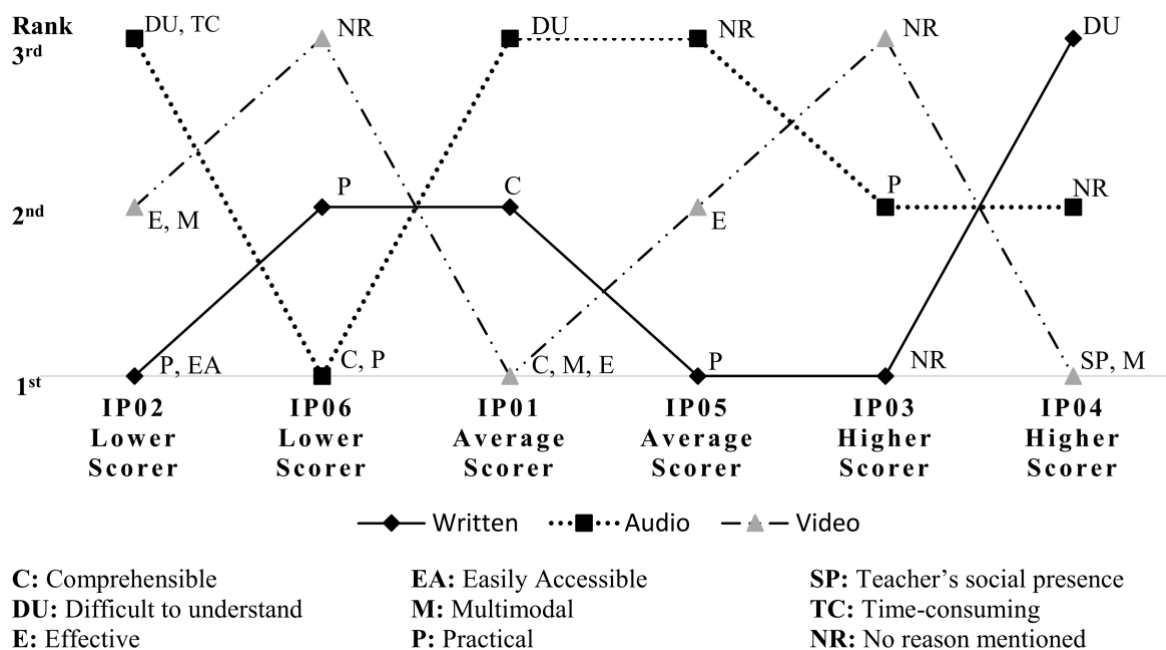
Still another factor was that the researcher's provision of digital feedback gave the impression that a lot of effort was taken to provide feedback to the students (Figure 4). During the interview, the students were also asked to compare digital feedback modes with their previous experience with paper-based feedback. Students mostly found digital feedback (without referring to any particular mode) not only more practical (easy-to-use) but also more effective. Having received WF and attended writing conferences during the previous academic year, they considered face-to-face meetings with the teacher they experienced last year as time-consuming and "less comfortable."

Factors Accounting for the Learners' Preferences

Theme 2. Several interrelated factors accounted for the respondents' preferences.

During the analysis, a prevalent issue was the participants' tendency to provide a rationale for their preferences, either upon request or naturally. The respondents thought that WF was the most practical one because there were fewer steps to access the feedback. That is, they could quickly revise their paper based on the WF provided in the marginal comments. As for the SCF, they considered it more comprehensible and motivating. Those who found AF less useful or practical recounted a negative

experience with the audio add-on, stating that it was very difficult to launch it. Figure 5 illustrates the interview participants' preferences for and their perceptions of the feedback modes. The horizontal axis represents the participants, while the vertical one stands for the rank of preference for the three modes. The participants' perceptions of each feedback mode is specified using acronyms, explained in the legend. For example, IP02 preferred WF to SCF, and AF was the least preferable one for him or her because it was not only difficult to understand but also time consuming (DU, TC). Some respondents seemed to prefer WF predominantly for its practicality, while others preferred SCF due to its multimodality, which promoted comprehensibility, and AF was the least preferred. Poorer comprehensibility was cited as a reason by all the respondents who clearly explained why they did not prefer it (Figure 5). A typical complaint was "I wasn't able to revise my writing as I couldn't understand some of the audio feedback." IP02 also considered this mode time-consuming, as he or she had listened to the portions of the AF multiple times, due to its poorer comprehensibility. While some respondents had problems with the AF tool itself, some others had problems understanding some of the audio comments. These two factors made the AF less preferable. Using an if-then tactic, the researcher rechecked the qualitative data to further understand this lower preference for this mode and found that lower comprehensibility due to lack of multimodality was more of a problem than the audio tool itself.

C: Comprehensible   **EA:** Easily Accessible   **SP:** Teacher's social presence
DU: Difficult to understand **M:** Multimodal   **TC:** Time-consuming
E: Effective   **P:** Practical   **NR:** No reason mentioned

Notes. (1) To create this figure, the researcher initially produced a matrix table in NVivo (with the features and participants as rows and columns) to compare what each participant said about each feedback mode and then drew this figure using the data from the table. (2) The labels in the legend were assigned when the respondent uttered the same adjective/phrase in the legend (or a phrase that explains the adjective or a synonym of it). For example, more technical terms, particularly "social presence," and "multimodal" were never used by the participants. (3) IP02 stated that paper-based written feedback is equally preferable for her (more preferable than the AF or SCF). (4) IP03 recursively stressed that written feedback was more effective but did not mention any reasons for this and made no positive comments about the SCF. She talked about the time she spent for it, yet she seemed unsure. This was accepted as an indication of lack of preference, so it was ranked as the third.

Figure 5. Rank order of preference for the three feedback modes.

Further analysis of the data revealed a network of relations among the constructs that shaped the respondents' preferences (Figure 6). The interrelationships among the variables in Figure 6 were mostly identified by checking the phrases expressing relations, particularly in the extracts coded in more than one code in this theme. For instance, the participants' comments about "multimodality" were accompanied by those regarding "comprehensibility," as one participant reported, "As there were no visual elements in the audio, I had difficulty understanding it" (IP01). This comment mirrored those of two others, and it was coded into both "multimodality" and "comprehensibility." Social presence and interactivity increased the respondents' motivation, and motivation in turn seemed to have an impact

upon their preference of digital feedback. They thought that comprehensibility and practicality were the two key predictors of effectiveness. They established an indirect link between effectiveness and multimodality because the latter, as they reported, boosted comprehensibility (Figure 6). While working on these constructs, the researcher checked multiple times what each of these constructs might mean for the respondents because it was recognized that they might have used some of these concepts interchangeably (e.g., practicality and effectiveness). For example, meaning was not clear in the following statement by IP05: "In WF, you only change the portion commented on, but in video…well, some general problems can be noticed as you show it. Therefore, video is more practical." (IP05). Such issues were resolved by examining the larger context of such utterances further. In this particular example, practicality was used to mean "effectiveness," yet in other instances it referred to user-friendliness.
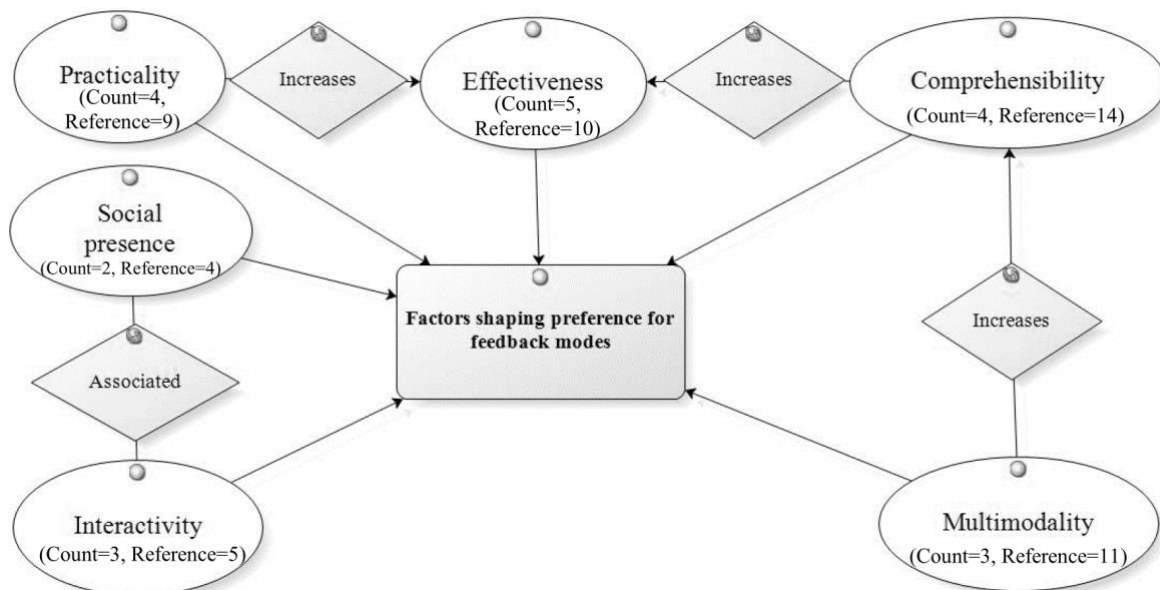
Notes. (1) This figure presents a visual summary of the theme "factors shaping the preferences for feedback modes" and the relations among the codes. (2) It was created by visualizing the codes and the relationships among these codes by using NVivo. Each of the ovals represents a code created using the data from the interview transcripts. The count and reference in each construct refer to the number of participants who talked about it and the number of mentions, respectively. (3) Three participants mentioned "interactivity" but they did not refer to a particular mode of feedback; this concept was not included in Figure 5.

Figure 6. The interaction of the factors that shaped the respondents' preference for feedback modes.

Interaction and Engagement

Theme 3. The respondents usually followed similar procedures and ran into similar problems.

Almost all of the respondents used computers to work on the feedback because they felt less comfortable with the smaller screens of mobile devices. Some reported that they used their mobile phones to check if the researcher provided the feedback. Still, a few respondents noted that they used their mobile phones to work on WF rather than videos or audio recordings because it was easier for them to open documents in mobile devices.

Another significant issue was the nature of the interaction between the researcher and students. A few participants sometimes responded to the comments in the AF and WF groups. Most of the respondents just revised their work based on the comments provided. The respondents in the AF group used the feature of written comments in the audio tool, while those in the WF group used comment balloons. In SCF, the students did not interact with the researcher, although they could have done it using marginal comment balloons. In general, the interaction was poor for both the AF and WF and almost none for the SCF. Yet, probably due to the social presence of the researcher, they found digital feedback interactive.

The analysis of the screen recordings (140 minutes) seemed to support the respondents' recount of how

they worked with the feedback they received. They watched the SCF multiple times as needed. Most of them watched it carefully without making any revisions at first. They later went back to their writing to fix the problems. In the AF and WF, it seemed easier for the participants to listen to or read portions of comments and revise the related part as the feedback was divided into many parts. In short, the data from the verbal probes seemed to verify the respondents' recounts of how they worked with the feedback.

Discussion

The first quantitative research question investigated to what extent each feedback mode helped the participants to address the microlevel and macrolevel problems correctly. The audio group made the highest number of revisions and the written group followed it in the essay-writing task. However, as in Orlando's (2016) study, there was a mismatch between the interview participants' preferences and the ranking of mean scores in the essay-writing task, as the least favorite mode was the audio. The second question sought to identify which mode was more effective in helping the learners to revise the microlevel and macrolevel problems in the essays supplied to them. In contradiction to the earlier finding that SCF was more effective in addressing macrolevel issues (Ali, 2006; Morra & Asis, 2009; Silva, 2012; Tuzi, 2004) or in microlevel issues (Mitchell, 2012), the present study has not found a significant difference between the three feedback modes for the essay-revision task in any of the revision scores. On the other hand, this finding lends support to a few studies in the literature (Ducate & Arnold, 2012 cited in Elola & Oskoz, 2016; Elola & Oskoz, 2016). However, it should be treated with caution as it is based on a revision task in which students revised essays supplied to them, which is quite different from a personalized essay-writing task.

The third research question investigated the participants' preferences for the feedback modes. A particularly significant result obtained in this study is that several important constructs could account for learners' preferences. This could also help explain the mismatch mentioned above because, as the qualitative data indicated, a particular feedback mode could be preferable depending on learners' priorities, such as practicality, comprehensibility, multimodality, effectiveness, interactivity, and the researcher's social presence (Figure 6). In other words, students' preferences do not correspond to effectiveness or any other factor alone; effectiveness itself could be the result of combined benefits of the constructs above. As the qualitative data indicated, comprehensibility, for example, might be an issue in AF. Although both written and audio comments are associated with particular portions of the text, the WF was considered clearer-cut than the AF. For the SCF, the respondents frequently stressed multimodality as a factor that improved comprehensibility. This in turn was influential in their preferences.

Moreover, it is worth stressing that the concept of practicality might be misleading because as students get used to a digital tool, it seems more practical to them. Qualitative data verified this, because the respondents who reported technical problems also unanimously accepted that working with digital feedback was actually easy. Therefore, advanced digital skills could improve one's perception of practicality, which is highly influential in preferences, as students consider supposedly impractical and time-consuming tools and methods as a threat to their comfort zones. This is in line with Godwin-Jones' (2018) warning that both learners and teachers might have a difficult time getting used to new forms of digital writing, and teachers could encounter resistance against innovations. Having to deal with new feedback modes seemed to have forced some respondents to leave their longstanding habits, causing them to consider such practices an intrusion into their comfort zones.

A significant contribution of digital feedback was that the students thought that the researcher had put more effort into the process of providing feedback (globally in digital feedback and more specifically in the SCF). As the data indicated, although interaction was poor due to their long-standing habits, it gave the respondents the feeling that they are being noticed and cared about. This is probably because, as some have noted, they felt the presence of the researcher while they were working on their writing, and this motivated them to work on their writing, which concurs with Harper et al., (2015). Globally, these findings correlate favorably to the findings of several researchers that highlighted the affective benefits of SCF (Cavanaugh & Song, 2014; Ducate & Arnold, 2012; cited in Elola & Oskoz, 2016; Ice et al.,

2007; Sipple, 2007). Another significant finding was that although SCF and AF could potentially increase the amount of interaction, personal characteristics along with culturally imposed limitations could be highly influential on the amount of interaction with the teacher or among students. For example, a student recounted his or her experience related to hesitation to respond to the researcher's comments, thinking that this might be inappropriate. He or she said, "You provided WF and told me lots of things... Frankly, I was shy, so I did not respond. I felt as if I was being a smart mouth with you" [IP05].

Despite their positive perceptions of digital feedback, the respondents felt anxiety, particularly at initial stages of the process. However, they seemed to have gradually become more positive and considered digital feedback more valuable than paper-based feedback. As the interview data indicated, the main reason for anxiety was the feeling of discomfort caused by unfamiliar procedures. This result has substantiated Orlando's (2016) recommendation for providing learners with training on new tools, as it is difficult to introduce novelty into the classroom in a short time. In other words, the participants' familiarity with new feedback tools could make it easier to use them in the classroom. In the present study, some of the participants were not so willing to adopt new practices by breaking longstanding habits, which is in agreement with previous research (Elola & Oskoz, 2016; Lin & Yang, 2011).

The fourth research question sought to explore what digital devices they used and how they worked with them. The respondents reported that they did not prefer mobile devices. This is probably because some tasks lend themselves better to computers than smaller devices, which could lower the quality of interaction. The interview respondents cited "smaller screen size" as the reason for less frequent use of mobile phones when receiving and responding to feedback. In this sense, there seems to be a long way to go for students to be able to use mobile devices efficiently for feedback purposes, particularly to respond to feedback.

Pedagogical Implications

As several researchers (Elola & Oskoz, 2016; Lee, 2017) noted, using digital feedback modes in combination could help eliminate limitations and bring together the benefits of different modes. The lack of statistically significant differences between the modes with respect to global scores and microlevel or macrolevel scores might justify the use of a combined approach. The interview data supported this, as the respondents highlighted certain characteristics of the modes, rather than praising a particular mode extensively (Figure 5). Such an approach could also appeal to different learning styles in a classroom. However, using several modes together could cause anxiety as well, or as Chang et al. (2017) noted, it might be overwhelming for students. As a solution to this problem, adequate training on how to use digital feedback modes should be provided (Chang, et al., 2017; Orlando, 2016). Teachers could also stress the importance of intelligibility of speech rather than its accuracy if students are to respond orally (e.g., by using an AF tool), or written responses could also be an option for those students who are anxious about their pronunciation.

Digital feedback, probably regardless of the mode, could help promote interaction. However, although feedback should be dialogic from a sociocultural perspective (Lantolf, 2006), it could take time to establish such a dialogic environment as students might be unwilling to interact at first. Technology could help as students interact and get used to receiving digital feedback. This is highly important because the inability to understand comments, which was a problem in the present study, could be solved through interaction. This is because, although there were very few such instances, some students stressed that they used the comment function to ask for clarification, particularly through written comments, upon having difficulty understanding feedback. However, establishing a dialogic environment is more strongly related to the value placed on interaction and questioning, rather than the technology itself. The idiom "being smart mouth" used by IP05 provided support for this.

Limitations

There are several limitations of this study. For example, although the learners were trained on how to revise their papers based on digital feedback, computer skills might have impacted the quality of their work. Another limitation of the study was that those who recorded their screens were not the same as

those interviewed, with the exception of one participant, because the interview participants were recruited based on their scores, which were calculated after doing the tasks. In addition, while the participants' experimentation with all three feedback modes was one of the strengths of the quantitative strand, the small number of participants was a limitation. Moreover, although the participants took the same proficiency test and those who passed it were accepted to the department, their writing proficiency might vary to a certain extent. However, the crossover design in the essay-revision task could largely eliminate problems regarding sample size and potential differences between the essays or participants' writing proficiency (Figure 2).

Finally, despite its advantages mentioned earlier, the essay-revision task was somewhat poor in ecological validity because one can expect that learners could more easily understand and respond to feedback given on their own writing than that of others. Therefore, learners might be less motivated to revise essays written by their peers. However, the researcher tried to eliminate this issue by assigning grades for both tasks, meaning that a motivational factor was present. This seems positive, yet it also comes with a potentially negative side effect. That is, grades could harm learners' objectivity in the interviews. Therefore, the interviews were carried out after the grades were announced to minimize such a possible negative effect. Despite these precautions, a possible task effect might have been at work in the study.

Recommendations for Further Research

Based on the results and limitations of this study, several further research topics could be suggested. The perception-based findings from the six interview participants could further be investigated both quantitatively and qualitatively in prospective studies with larger samples, and as noted in a recent meta-analysis (Kang & Han, 2015), long-term studies are needed for more reliable results. Moreover, as the participants' learning styles were not taken into consideration, prospective studies could investigate the impact of learning styles when providing feedback in different modes to students. Similarly, further studies could focus on how providing feedback based on student preferences could affect the quality of writing. Peer feedback practices using different modes could also be investigated. Finally, as this study focused on asynchronous audio and video feedback, another avenue for further research could be the use of synchronous written, audio, and video chat for feedback as online versions of writing conferences.

Acknowledgements

References

Ali, A. D. (2016). Effectiveness of using screencast feedback on EFL students' writing and perception. *English Language Teaching, 9*(8), 106–121. https://doi.org/10.5539/elt.v9n8p106

Alvira, R. (2016). The impact of oral and written feedback on EFL writers with the use of screencasts. *PROFILE: Issues in Teachers' Professional Development, 18*(2), 79–92. https://doi.org/10.15446/profile.v18n2.53397

Bakla, A. (2017). An overview of screencast feedback in L2 writing: Fad or the future? International Foreign Language Education and Turkish as a Foreign Language Education Symposium, April 27–28, Ankara University, TÖMER, Bursa.

Biber, D., Nekrasova, T., & Horn, B. (2011). The effectiveness of feedback for L1-English and L2-writing development: A meta-analysis. Retrieved from http://www.ets.org/Media/Research/pdf/RR-11-05.pdf https://doi.org/10.1002/j.2333-8504.2011.tb02241.x

Bitchener, J., Young, S., & Cameron, D. (2005). The effect of different types of corrective feedback on ESL student writing. *Journal of Second Language Writing*, *14(*3), 191–205.

https://doi.org/10.1016/j.jslw.2005.08.001

Boswood, T., & Dwyer, R. (1996). From marking to feedback: Audio-taped responses to student writing. *TESOL Journal*, *5*(2), 20–23.

Cavanaugh, A., & Song, L. (2014). Audio feedback versus written feedback: Instructors' and students' perspectives. *MERLOT Journal of Online Learning and Teaching*, *10*(1), 122–38.

Chang, C., Cunningham, K. J., Satar, H. M., & Strobl, C. (2017). Electronic feedback on second language writing: A retrospective and prospective essay on multimodality. *Writing & Pedagogy, 9*(3), 405–428. https://doi.org/10.1558/wap.32515

Cranny, D. (2016). Screencasting, a tool to facilitate engagement with formative feedback? *AISHE-J: The All Ireland Journal of Teaching & Learning in Higher Education, 8*(3), 2911–2927.

Creswell, J. W. (Ed.) (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th Ed.). Boston, MA: Pearson Education.

Creswell, J. W. (Ed.) (2013). *Qualitative inquiry and research design: Choosing among five approaches (3rd Ed.).* London: Sage Publications.

Ducate, L., & Arnold, D. (2012). Computer-mediated feedback: Effectiveness and students' perceptions of screen-casting software vs. the comment function. In G. Kessler, A. Oskoz, & I. Elola (Eds.), Technology across writing contexts and tasks [Monograph]. *CALICO Monograph Series, 10,* 31–55.

Edwards, K., Dujardin, A., & Williams, N. (2012). Screencast feedback for essays on a distance learning MA in professional communication: An action research project. *Journal of Academic Writing*, *2*(1), 95–126. https://doi.org/10.18552/joaw.v2i1.62

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

Elola, I., & Oskoz, A. (2016). Supporting second language writing using multimodal feedback. *Foreign Language Annals, 49*(1), 58–74. https://doi.org/10.1111/flan.12183

Ferris, D. (2003). *Response to student writing: Implications for second language students*. Mahwah, NJ: Lawrence Erlbaum.

Garrison, D. R., & Anderson, T. (2003). *E-Learning in the 21st century*. London, New York: Routledge.

Ghosn-Chelala, M., & Al-Chibani, W. (2018). Screencasting: supportive feedback for EFL remedial writing students. *The International Journal of Information and Learning Technology*. Retrieved from: https:// doi.org/10.1108/IJILT-08-2017-0075

Godwin-Jones, R. (2018). Second language writing online: An update. *Language Learning & Technology*, *22*(1), 1–15. https://dx.doi.org/10125/44574

Harper, F., Green, H., & Fernandez-Toro, M. (2015). Using screencasts in the teaching of modern languages: Investigating the use of Jing® in feedback on written assignments. *The Language Learning Journal*, *43*, 1–18. https://doi.org/10.1080/09571736.2015.1061586

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112.

Huang, S. (2000). The nature of an EFL teacher's audiotaped and written feedback on student writing: A case study. ERIC Document 438728.

Hynson, Y. T. A. (2012). An innovative alternative to providing writing feedback on students' essays. *Teaching English with Technology*, *12*(1), 53–57.

Ice, P., C. Reagan, P. Perry, & Wells, J. (2007). Using asynchronous audio feedback to enhance teaching presence and students' sense of community. *Journal of Asynchronous Learning Networks*, *11*(2), 3–25. https://doi.org/10.2190/EC.43.1.g

Kang, E., & Han, Z. (2015). The efficacy of written corrective feedback in improving L2 written accuracy: A meta-analysis. *The Modern Language Journal*, *99*(1), 1-18.

Kirschner, P. A., van den Brink, H. and Meester, M. (1991). Audiotape feedback for essays in distance education. *Innovative Higher Education*, *15*(2), 185–195.

Lantolf, J. P. (2006). Sociocultural theory and second language learning: State of the art. *Studies in Second Language Acquisition*, *28*, 67–109.

Lee, I. (2017). *Classroom writing assessment and feedback in L2 school contexts.* Singapore: Springer Nature Singapore.

Li, K., & Akahori, K. (2008). Development and evaluation of a feedback support system with audio and playback strokes. *CALICO Journal*, *26*(1), 91–107.

Lin, W., & Yang, S. (2011). Exploring students' perceptions of integrating wiki technology and peer feedback into English writing courses. *English Teaching: Practice and Critique*, *10*(2), 88–103.

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition Vol. 2. Second language acquisition* (pp. 413–468). New York: Academic Press.

Mathieson, K. (2012). Exploring student perceptions of audiovisual feedback via screencasting in online courses. *American Journal of Distance Education, 26*(3), 143–156. https://doi.org/10.1080/08923647.2012.689166

Mayer, R. E. (2003). Theories of learning and their application to technology. In H. F. O'Neil & R. S. Perez (Eds.), *Technology applications in education: A learning view* (pp. 127–157). Mahwah, NJ: Erlbaum.

Mitchell, K. (2012). Screencast feedback on writing: Comments students can see and hear. Retrieved from: https://www.slideshare.net/ktmtchl/screencast-feedback

Moore, N. S., & Filling, M. L. (2012). iFeedback: Using video technology for improving student writing. *Journal of College Literacy & Learning, 38*, 3–14.

Morra, A. M., & Asis, M. I. (2009). The effect of audio and written teacher responses on EFL student revision. *Journal of College Reading and Learning*, *39*(2), 68–82. https://doi.org/10.1080/10790195.2009.10850319

Orlando, J. (2016). A comparison of text, voice, and screencasting feedback to online students. *American Journal of Distance Education, 30*(3), 156–166. https://doi.org/10.1080/08923647.2016.1187472

Patton. M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks, CA Sage Publications.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129–58.

Séror, J. (2012). Show me! Enhanced feedback through screencasting technology. *TESL Canada Journal*, *30*(1), 104-116. https://doi.org/10.18806/tesl.v30i1.1128

Silva, M. L. (2012). Camtasia in the classroom: Student attitudes and preferences for video commentary or Microsoft Word comments during the revision process. *Computers and Composition, 29*, 1–22. https://doi.org/10.1016/j.compcom.2011.12.001

Sipple, S. (2007). Ideas in practice: Developmental writers' attitudes toward audio and written feedback. *Journal of Developmental Education*, *30*(3), 22–31.

Stannard, R. (2008). A new direction in feedback. *Humanizing Language Teaching*, *10*(6), Available at: www.hltmag.co.uk/dec08/mart04.htm

Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principles and practice in applied linguistics* (pp. 125–144). Oxford: Oxford University Press.

Thompson, R., & Lee, M. J. (2012). Talking with students through screencasting: Experimentations with video feedback to improve student learning. *The Journal of Interactive Technology and Pedagogy*, *1*(1), 1-16.

Tuzi, F. (2004). The impact of e-feedback on the revisions of L2 writers in an academic writing course. *Computers and Composition*, *21*(2), 217–235. https://doi.org/10.1016/j.compcom.2004.02.003

Vincelette, E. J., & Bostic, T. (2013). Show and tell: Student and instructor perceptions of screencast assessment. *Assessing Writing*, *18*(4), 257-277.

Appendices

Appendix A. Teacher Feedback to the Parallel Essays Provided in the Three Modes

In all three modes, the learners were given written instructions in Google Drive telling them to revise the essays based on the feedback. Links are provided in the table below.

Table 4. *The parallel essays* (Click on the essays to view them.)

Written	Audio	Screencast
Essay 1	Essay 1*	Essay 1
Essay 2		Essay 2
Essay 3		Essay 3

Notes. (1) Instructions: Click on the link to access the feedback given on the three essays in each mode. (2) The participants were asked to read the comments, listen to them or watch the feedback supplied as a screencast, and revise the essays accordingly. (3) *After the study was carried out the user interface of the audio feedback tool was changed, so it was not possible to open the feedback provided using the previous version of the interface in Google Drive documents.

Appendix B. The Questions Used in the Interview Protocol

1. What feedback modes have you used while revising your essay?
2. Briefly describe your feelings about giving and receiving feedback in a digital environment.
   Probe: How engaging was it?
3. What differences do you think there are between paper-based feedback and digital feedback?
   Probe: Ease of giving feedback, time spent for it, motivation to give feedback and reading the feedback given, access to feedback and practicality and so on.
4. What digital devices did you use to give and receive feedback (read, listen, watch)? *Probe:* Why did you use that particular device (those particular devices)? Was there any particular reason? (Availability, internet access, ease of use and mechanical considerations)
5. How did you use the links provided in the video, audio, and comment balloons?
   Probe: Why?
6. Which feedback mode (video, audio, or written) is the most and least practical with respect to responding to it?
7. Which feedback mode (video, audio, or written) is the most effective one?
   Probe: Why?
8. What do you think about the quality of the feedback provided by your friend or your teacher?
9. What were the technological challenges that you have encountered while giving and receiving feedback?
   Probe: How did you overcome them?
10. Do you wish to receive and give feedback using audio comments and screen recording in the future?
    Probe: Why? Why not?
11. Additional Comments and Observations:

Note. The interview was carried out in Turkish and the quotes included in the results section were translated into English by the researcher and checked by a professional translator.

Appendix C. Additional Notes on Qualitative Analysis

1. During the coding, the researcher continuously created annotations and memos to keep track of ideas and to facilitate analysis. He coded them into groups and examined these groups and related portions of the interview transcripts to organize his thinking. The first step was to add annotations in the interview transcripts. Then the researcher pasted all the annotations into a memo and categorized them based on how they facilitate the analysis (See Figure 8). The examination of the groups of annotations also helped find out weaker points of analysis. For example, it clarified that despite a large number of annotations reserved for alternative explanations, interpretations, ideas for the report, suggestions for better analysis, and questioning, there were very few annotations for such categories as disconfirming evidence, if-then questions, or notes on credibility. Such information helped him to focus on weaker points in later phases of the analysis. Reflection during this process helped the researcher identify alternative meanings and accept more credible ones based on the data. In addition, exemplar quotes were coded into a node for ease of access and additional thinking on them.

Figure 8. The classifications of the annotations coded in the analytic codes in NVivo (top) and a screenshot from the code "Interpretations and ideas for the report" (bottom).

2. As a part of the validation process, the respondents were asked to read the transcripts to make the qualitative data more credible. During the design of the study and analysis of the data, ideas of peers were sought, and the results of the study were shared with them for peer debriefing. Finally, as a way of external auditing, two experienced researchers, one of them being outside applied linguistics, reviewed the paper.

Appendix D. Correction Rates in the Essay-Writing Task

Table 5. *Macrolevel Correction Rates in the Essay-Writing Task*

Macrolevel Issues	Audio			Screencast			Written		
	# E	# C	% C	# E	# C	% C	# E	# C	% C

	# E	# C	% C	# E	# C	% C	# E	# C	% C
Inappropriate title	2	1	50	3	1	33	2	2	100
Lack of an appropriate hook	2	2	100	4	1	25	2	1	50
Inability to connect the hook and the thesis statement	2	2	100	1	1	100	0	0	NA*
Lack of an appropriate thesis statement	2	1	50	4	2	50	2	1	50
Need for clarification or additional information	22	17	77	16	10	63	22	15	68
Logical fallacy and arguments too strong to be true	6	5	83	2	2	100	0	0	NA
Poorly developed paragraphs	13	10	77	16	7	44	5	2	40
Lack of support	2	2	100	3	1	33	2	2	100
Lack of unity/irrelevant ideas	2	2	100	4	4	100	2	1	50
Lack of cohesion	6	5	83	3	1	33	0	0	NA
Lack of appropriate topic sentences	1	1	100	1	1	100	0	0	NA
Repetition	5	4	80	4	3	75	4	4	100
Total number of errors/corrections	65	52	-	61	34	-	41	28	-
Overall correction rate**	-	-	80	-	-	56	-	-	67

*NA: Not applicable (No errors regarding that item in that particular mode)

** An overall correction rate was obtained by calculating the proportion of the total correct revisions to total number of errors (rather than averaging the individual percentages of correction).

*** E = Errors, C = Corrections

Table 6. *Macrolevel Correction Rates in the Essay-Writing Task*

Microlevel issues	Audio			Screencast			Written		
	# E	# C	% C	# E	# C	% C	# E	# C	% C
Verb tense error	4	3	75	10	7	70	0	0	NA*
Subject-verb agreement error	4	4	100	7	7	100	5	4	80
Active/Passive error	7	4	57	6	4	67	4	3	75
Word order error	7	5	71	4	2	50	4	3	75
Word form error	17	17	100	8	8	100	17	12	71
Wrong word choice error	17	12	71	17	11	65	18	10	56
Pronoun agreement error	4	3	75	1	1	100	5	4	80
Singularity/plurality error	15	15	100	8	6	75	13	13	100
Preposition error	15	12	80	13	8	62	13	9	69

	E	C	%	E	C	%	E	C	%
Spelling	6	3	50	0	0	NA	6	6	100
Fragments	0	0	NA	0	0	NA	1	1	100
Comma splice	3	3	100	5	3	60	0	0	NA
Comma use (before coordinating conjunctions)	3	2	67	1	1	100	0	0	NA
Punctuation of conjunctive adverbs	5	4	80	3	1	33	2	1	50
Punctuation of subordinate clauses	2	2	100	0	0	NA	0	0	NA
Punctuation of introductory clauses	1	1	100	1	1	100	0	0	NA
Space problem	3	1	33	7	5	71	4	3	75
Articles	0	0	NA	13	10	77	4	1	25
Missing words	12	9	75	4	2	50	6	4	67
Total number of errors/corrections	**125**	**100**	**-**	**108**	**77**	**-**	**102**	**74**	**-**
Overall correction rate*	**-**	**-**	**80**	**-**	**-**	**71**	**-**		**73**

*NA: Not applicable (No errors regarding that item in that particular mode)

** An overall correction rate was obtained by calculating the proportion of the total correct revisions to total number of errors (rather than averaging the individual percentages of correction).

*** E = Errors, C = Corrections

About the Author

Arif Bakla holds a PhD degree in ELT. Among his research interests are technology-enhanced language instruction, digital materials design, and feedback in L2 writing. He has recently published in various international journals, including ReCALL, Computers & Education, Instructional Science, and Qualitative Research. He is the author of the book *Putting Pen to Paper: Academic Paragraph Writing*.

E-mail: abakla@ybu.edu.tr