Student Engagement with Teacher Written Feedback on Rehearsal Essays Undertaken in Preparation for IELTS

William S. Pearson

Abstract

Due to pressure to meet goals, some test-takers preparing for the IELTS (International English Language Testing System) Writing test solicit written feedback (WF) from an expert provider on their rehearsal essays, in order to identify and close gaps in performance. The extent self-directed candidates are able to utilize written feedback to enhance their language and writing skills in simulated Task 2 essays has yet to be investigated. The present study addresses one learner factor deemed prominent in mediating the learning potential of WF, student engagement. The study used assessments of student writing according to the public band descriptors and text-analytic descriptions from three Task 2 rehearsal essays triangulated with five rounds of semi-structured interviews to explore how four learners preparing for IELTS Writing engaged affectively, behaviorally, and cognitively with asynchronous, electronic written feedback provided using the Kaizena app. The study found that, while the learners highly valued WF, they were not always able to understand the intentions behind comments or envisage an appropriate response, leading to negative emotional reactions from two learners in the form of anxiety and frustration. Written progress across the essays was limited, stemming from an initial lack of buy-in to making content revisions and surface-level approaches to WF processing. Moderate behavioral engagement with indirect error treatment was exhibited, although meaningful accuracy gains were apparent for only one learner and content changes meant many errors went uncorrected. The implications for practitioners of IELTS Writing preparation are discussed.

Keywords

written feedback, student engagement with written feedback, high-stakes English writing assessment, IELTS, learning to write

Introduction

There has been increasing interest in L2 student engagement as the key to unlocking how and why learners do or do not gain utility from written feedback (Han & Hyland, 2015; Han & Xu, 2021; Ranalli, 2021; Uscinski, 2017; Yu & Jiang, 2020; Zhang, 2020). A multi-dimensional approach comprising affective, behavioral, and cognitive dimensions, built on the work of Fredricks et al. (2004) and Ellis (2010), offers researchers a robust theoretical model that accounts for the richness and complexity of student engagement (Han & Hyland, 2015). By triangulating textual evidence (e.g., revision operations, document editing time) with analysis of student verbal reports, scholars can gain insights into the quantity and quality of student processing and response as well as attitudinal and emotional reactions to written feedback (WF). As such, multi-dimensional studies of student engagement have gained widespread traction in recent years (Fan & Xu, 2020; Tian & Zhou, 2020; Zhang, 2020; Zheng et al., 2020; Zheng & Yu, 2018), albeit in mostly Chinese tertiary-level writing contexts. In line with Ellis’ (2010) initial heuristic, much research has focused on engagement with form-focused written feedback (FFWF). The extent to which students engage with content-focused written feedback (CFWF) on a diverse array of global and local textual issues combined with unfocused FFWF remains unexplored.

One learning-to-write setting where L2 developing writers may need to engage with comprehensive content- and form-focused written feedback is preparation for the IELTS (International English Language Testing System) Writing test. In order to meet test-user determined cut-off scores, many candidates undertake self-directed or teacher-led IELTS preparation (Alsagoafi, 2018; Chappell et al., 2019; Gan, 2009; Hu & Trenkic, 2019; Winke & Lim, 2014), a key feature of which is simulated practice test tasks (Allen, 2016; Chappell et al., 2019; Hu & Trenkic, 2019; Mickan & Motteram, 2009). Test-takers perceive feedback from a
provider with expertise (usually a teacher or successful test veteran) on simulated writing tasks integral to enhancing performance (Pearson, 2019), particularly since IELTS Writing is underscored by certain idiosyncrasies that do not cohere with the expectations of academic writing generally, (see Moore & Morton, 2005). Yet the extent test-takers make sense of and respond to comprehensive, unfocused CFWF and FFWF to develop language and writing skills in preparation for high-stakes English writing assessment is largely unknown. To address this gap, the present mixed methods case study explores how four candidates preparing for IELTS affectively, behaviorally, and cognitively engage with written feedback on two drafts of three Task 2 rehearsal essays.

**Literature Review**

**Rehearsing for IELTS Writing Task 2**

IELTS is a high-stakes English language proficiency (ELP) test undertaken to generate evidence that a candidate meets the linguistic requirements determined by a test-user for the target language use (TLU) domain in which successful candidates are expected to operate (IELTS, 2019a). The primary TLU domain of IELTS is academic study with English as the medium of instruction (the Academic test), and to a lesser extent, everyday social, transactional, and workplace language use for migrants to Anglophone countries (General Training). Both forms of the Writing test feature two tasks, to be completed within 1 hour with no recourse to external sources. Task 2, the focus of this study, consists of a 250-word essay requiring candidates to present an argument on an issue with the goal of convincing the reader that something “is the case” or “should be the case,” termed **analytical and hortatory exposition** respectively (Mickan & Slater, 2003, p. 63). Candidates are required to adopt a position on topics of “general interest and suitability for test-takers entering undergraduate or postgraduate studies” (Academic) or of “general interest” (General Training) (IELTS, 2019a, p. 13). Prompts and the conditions in which responses are written result in particular idiosyncratic academic conventions, such as personalized opinion-giving, a focus on real world situations, actions, and processes and evidence as anecdote and experience (Moore & Morton, 2005). As such textual features generally diverge from the expectations of tertiary-level academic writing, test outcomes should be interpreted as indicators of pre-study academic readiness or language learning aptitude (Davies, 2008).

Writing Task 2 is assessed using four equally weighted analytical assessment criteria across nine performance bands unique to the test (see IELTS, 2019b). As shown in Table 1, Task Response (TR) and Coherence and Cohesion (CC) mostly encompass the global features of response, while Lexical Resource (LR) and Grammatical Range and Accuracy (GRA) concern the surface-level. Underscoring the assessment of IELTS Writing is a general proficiency theoretical approach. The test eschews specific language structures and functions, instead claiming the existence of generic academic language use, generalizable to any academic domain (Davies, 2008). This is reflected in the public band descriptors (PBDs) (see IELTS, 2019b). For instance, at TR band 7.0 (“good user”), candidates are expected to present “a clear position throughout the response,” with main ideas that are “extended” and “supported.” If “a relevant position” with “relevant main ideas” is presented, albeit with some ideas that are “inadequately developed/unclear,” band 6.0 (“competent user”) is awarded. There is minimal elaboration of what such statements entail in official practice materials, requiring learners to infer features of written performance from the designated overall score and summary comments on sample essays, a challenging prospect for many candidates (and practitioners of IELTS preparation).

Prior research has shown the practice of undertaking rehearsal tasks in simulated test conditions is a central feature of self-directed and teacher-led IELTS Writing preparation (Allen, 2016; Chappell et al., 2019; Hu & Trenkic, 2019; Mickan & Motteram, 2009; Smirnova, 2017). Rehearsal essays are popular with candidates because they provide opportunities to self-evaluate test format familiarity (Pearson, 2019), enhance understanding of typical topics and prompt frames (Coffin, 2004; Craven, 2012), and practice skills and strategies used to lower the burden of test taking (Chappell et al., 2019; Winke & Lim, 2014). Nevertheless, improving

---

**Table 1. Textual Features Referenced in the IELTS Writing Task 2 Public Band Descriptors.**

| Task response | Coherence and cohesion | Lexical resource | Grammatical range and accuracy |
|---------------|------------------------|-----------------|-------------------------------|
| How fully the prompt is addressed | Organization of ideas | Range of lexis | Range and accuracy of structures |
| Clarity and relevance of the position presented | Progression | Use of uncommon items | Use of complex structures |
| Support, relevance, and extension of main ideas | Paraproaching | Accuracy of word choice and collocation | Accuracy of complex forms |
| Clarity and justification of conclusions | Use of referencing and substitution | Accuracy of spelling and word formation | |
| Appropriacy of the format | Use of cohesive devices | |

*Note. Adapted from (IELTS, 2019b).*
future test performance via rehearsal essay writing is not assured. It may take over 200 hours of English language learning to tangibly improve writing outcomes (Green, 2005), while variability in topics and the absence of syllabi of structures or functions limits the use of memorized material. As non-experts in how Task 2 is assessed, developing writers may struggle to diagnose problematic textual issues, a situation compounded by the vague phrasing of performance standards in the public band descriptors. As a consequence, dependence on the assistance of experts who offer feedback is often reported (Allen, 2016; Estaji & Tajeddin, 2012; Mickan & Motteram, 2009). There is also a danger that, instead of undertaking what is really required—upgrades in ELP (Ingram & Bayliss, 2007)—candidates entrap themselves in extensive test preparation, believing that exploiting design features of the test will yield the requisite gains (Alsagoafi, 2018; Chappell et al., 2019; Hamid & Hoang, 2018; Hu & Trenkic, 2019).

**Student Engagement With Written Feedback**

Rehearsal writing for test preparation purposes provides important opportunities for written feedback. The existence of an elaborated series of assessment criteria may facilitate understandings of how a learner performed in a task through written feedback’s correction, reinforcement, forensic diagnosis, benchmarking, and feed-forward roles (Price et al., 2010). Since IELTS preparation is typically undertaken by learners possessing emergent language skills for acculturation purposes or as a means of improving future test performance, feedback’s forensic-diagnosis and benchmarking functions are salient. Learners frequently seek explanations and justifications of existing performance in combination with illustrative guidance on extending outcomes to the next band (Pearson, 2019). Nevertheless, Task 2 feedback providers are faced with upwards of 17 discrete textual features across the four criteria to attend to (Table 1), not to mention complex content and delivery written feedback decisions (Chong, 2020), most notably how much to provide, whether to refer to directly to the PBDs, how to illustrate target performance, and whether to treat errors directly, indirectly, or at all.

A multi-dimensional approach to exploring student engagement with written feedback has been adopted in several recent studies, particularly in relation to WCF (written corrective feedback) provided on in-sessional English support programs at Chinese tertiary-level institutions (Han & Hyland, 2015; Tian & Zhou, 2020; Yu et al., 2018; Zhang & Hyland, 2018; Zheng et al., 2020; Zheng & Yu, 2018). Such research is theoretically underpinned by Fredricks et al. (2004) and Ellis’ (2010) componential frameworks for investigating engagement. The three dimensions, the affective, behavioral, and cognitive were originally operationalized in L2 writing settings by Han and Hyland (2015), with modifications by later authors. The affective dimension denotes students’ attitudes toward WF and the emotional reactions it might evoke. The behavioral dimension concerns students’ textual responses (usually operationalized as quantifiable revision operations, time spent editing their work, or number of submissions for feedback) as well as the activities and strategies learners undertake to improve the accuracy of drafts or develop their L2. Lastly, the cognitive dimension encompasses students’ awareness and understanding of WF, and cognitive and metacognitive operations involved in processing and response. Learners’ mental states and processes are usually explored qualitatively, either through interviewing or immediate or retrospective verbal reports. The model has yet to be applied to preparation for high-stakes English writing assessment, where there are notable pressures to perform in simulated tasks that resemble product writing.

The present literature has shown that student engagement with WF is complex, dynamic, and often-times unsatisfactory or inconsistent (Han & Hyland, 2015; Han & Xu, 2021; Yu et al., 2018; Yu & Jiang, 2020; Zhang & Hyland, 2018; Zheng & Yu, 2018). Affectively, learners with positive attitudes, high self-confidence, and agreement with WF are more likely to invest time in responding to WF (Zhang & Hyland, 2018). However, the negative emotional reactions invoked by disappointing textual outcomes can suppress a student’s motivation to respond (Yu & Jiang, 2020; Zhang & Hyland, 2018). Behaviorally, learners undertake revisions in order to develop their writing or employ more L1-accepted linguistic forms. Students may avoid responding if they disagree with the WF (Yu et al., 2018) or perceive no purpose or benefit in response (Zheng et al., 2020). Cognitively, learners’ abilities to notice or understand errors are associated with their language proficiency (Zhang & Hyland, 2018; Zheng & Yu, 2018) and feedback literacy (Han & Xu, 2021), although learners of any ability may misinterpret the intent underlying WF, particularly if it is meaning focused or not explained clearly and concisely.

**The Study**

This case study stems from a larger research project (see Pearson, 2021) where candidates preparing for IELTS were recruited to take part in a remote learning-to-write program centered on the provision of asynchronous, electronic content- and form-focused written feedback on rehearsal writing undertaken as preparation for the test. The design of the study was guided by the following research question: How do student writers preparing for IELTS engage affectively, behaviorally, and cognitively with asynchronous, electronic written feedback provided on three rounds of Task 2 rehearsal essays?

**Participants**

Four individuals preparing for IELTS were purposively selected from a wider pool of 8 students who completed the
Table 2. Participants’ Background Characteristics.

|                      | Kushal          | Yuri            | Min Jung        | Chandrika       |
|----------------------|-----------------|-----------------|-----------------|-----------------|
| **Nationality**       | India           | Russian Federation | Republic of Korea | Sri Lanka       |
| **Age**              | Late 20s        | Late 20s        | Early 20s       | Mid 20s         |
| **Gender**           | Male            | Male            | Female          | Female          |
| **Education level**  | Undergraduate   | Undergraduate   | Undergraduate   | Undergraduate   |
| **IELTS Module**     | Academic        | General Training | Academic        | Undecided       |
| **Occupation**       | Doctor          | Draftsman       | Student         | Teacher         |
| **Reason for doing IELTS** | Professional training abroad | Migration to Canada | Higher education abroad | Migration to Canada |
| **Academic discipline** | Medicine       | N/A             | Nursing         | Undecided       |
| **Desired Writing band score** | 7.5             | 7.0             | 6.5             | 7.0             |
| **No. of test attempts** | 2               | 2               | 2               | 0               |
| **Previous test score(s)** | 6.5, 7.0       | 6.5, 6.5        | 5.5, 5.5        | N/A             |

The study utilizes textual evidence in combination with computer-mediated semi-structured interviewing (a preliminary and closing interview as well as one after each round of writing) to generate knowledge of how the learners behaviorally and cognitively engaged. Revision operations and error density measures were deductively coded, drawing on pre-existing schema (Christiansen & Bloch, 2016; Han & Hyland, 2015; van Beuningen et al., 2012; Zhang & Hyland, 2018). Interviews (with the exception of the preliminary and closing encounters) comprised ‘talking around the text’ (Ivanič & Satchwell, 2007), featuring screen sharing of learners’ rehearsal essays and the accompanying written feedback, used as a stimulus, along with a schedule that took inspiration from the current literature (Han & Hyland, 2015; Zheng et al., 2020; Zheng & Yu, 2018). Students’ affective responses were explored through interviewing only. All interviews were conducted in English using videoconferencing software (Zoom), with the transcripts analyzed thematically (Braun & Clarke, 2006).

There were five stages to data collection, represented in a flowchart in Figure 1. Stage 1 comprised an initial semi-structured interview (see Supplemental Material, part A), undertaken to get to know the participants, query their writing needs, identify WF preferences and expectations, and address questions they had about the study. Thereafter, the learners undertook three rounds of stages 2–4 of the project, consisting of writing a Task 2 rehearsal essay in conditions that simulated the test, which were submitted to the researcher by email or Facebook Messenger for written feedback (stage 2). FFWF and CFWF were provided on essay drafts targeting aspects of students’ written performance that fell short of goals. At the end of this stage, the essays and written feedback were returned electronically through Kaizena, an online application that allows content and surface-level textual features to be highlighted and commented upon through an interactive feedback “conversation” between the researcher and student (explained in more detail below).

In stage 3, learners attended to the written feedback in a revised version of their essay. The inclusion of a second draft, while atypical of IELTS Writing preparation, provided an opportunity for the participants to engage with the written feedback textually as well as to meet their goals in a less pressurized writing context. Additional summative WF was provided on participants’ second drafts, which were also uploaded to Kaizena. Form- and content-focused revision operations were coded during this stage, while a list of salient textual issues to address in the interview was drawn up. Within 1 or 2 days of being returned, a computer-mediated semi-structured interview was held to explore participants’ affective, behavioral, and cognitive engagement with written feedback (stage 4, Supplemental Information, part B). Interviews centered around discussion of the identified salient issues, presented via screen sharing of learners’ compositions and the accompanying feedback for use as a
stimulus. Textual features that had not been cleared up along with new issues were addressed in the form of a tutorial. After the third round of stage 4, a closing interview was held with learners (stage 5, Supplemental Information, Part C), addressing their progress on the project and evaluations of participation. The elapsed time for data collection varied according to participants’ availability, the speed in which they submitted their drafts and attended to the written feedback, and their keenness to undertake the real test. Yuri and Chandrika completed the research activities in 21 and 22 days, while Kushal took 37 days and Min Jung 44.

**Task Prompts and Written Feedback**

Multiple rounds of writing were chosen, partly to address the current lack of longitudinalness in engagement research (Han & Hyland, 2015; Yu et al., 2018). Three parallel Academic Task 2 prompts (see Supplemental Information, Part D) were selected by the researcher, based on the diversity of topics and frames. It was felt three rounds of writing offered a suitable balance between providing opportunities to engage with written feedback/(possibly) change the approach to first draft writing and mitigating the prospect of participant attrition and excessive study complexity. The learners were instructed to write first drafts in conditions that simulate Writing Task 2, that is, to spend about 40 minutes on the response, write at least 250 words, not refer to external sources of information, and include relevant examples from their knowledge or experience. There was no obligation to complete second drafts under these conditions. Upon submission, participants’ essays were assessed in relation to the PBDs by the researcher, with a score being assigned in the four criteria and overall. Thereafter, WF was generated by the researcher, explicitly orientated toward helping the learners better meet the demands of the test at their stated band score level (see in Table 2) in light of the pre-eminent problematic textual features uncovered in the
assessment. Facilitating the evaluation and feedback provision, the researcher possessed an MA in Applied Linguistics, several years’ experience of teaching IELTS preparation (involving providing WF on simulated practice tasks), and prior training and experience assessing authentic Task 2 scripts in a professional capacity.

FFWF was provided to address participants’ Lexical Resource and Grammatical Range and Accuracy. First draft errors were comprehensively corrected using a metalinguistic code (sometimes including additional explanation) based on Han and Hyland (2015) (Supplemental Information, Part E). Indirect correction was adopted to encourage students to notice patterns in error types and take responsibility for self-correcting lexicogrammatical problems. Complex uninterpretable errors (see Ferris et al., 2011) poorly suited to the code were addressed directly, as were all second draft errors. Written commentary in the text body diagnosed and explained textual features deemed problematic in relation to participants’ desired scores. To promote feed-forward in student revisions, explicit strategies and/or sample reformulations were provided when possible. Additionally, a summative outline of the learner’s performance in the four criteria was provided along with advice on how to improve. While the majority of comments stemmed from criteria outlined in the PBDs, occasionally, formative feedback beyond the descriptors was applied. A sample of the written feedback is outlined in the Supplemental Material (Part F).

All written feedback was provided to the participants working with the researcher in a closed, virtual classroom space on the Kaizena app (free registration required). This involved the WF, originally generated on a Word document version of the essay, being transferred to the application’s “conversation” bar (essentially resembling Microsoft Word’s Reviewing Pane), linked to a clean version of the essay uploaded to the space. This allowed textual features to be highlighted and targeted with form- and content-focused “comments” (like Word), but also global comments not linked to any selected text. The rationale for using Kaizena was to provide a singular online space for interactions and because content-focused messages were predicted to be complex or controversial, requiring clarification or discussion. Additionally, it was hoped learners’ written responses to Kaizena comments would offer further insights into their cognitive engagement. However, the expected dialogic interactions did not materialize, perhaps because the information was perceived in receptive-transmission terms, or the participants lacked the confidence to challenge messages. Participants were only able to download a feedback-less version of their essays from Kaizena, preventing them from “accepting changes,” as in Word.

**Data Analysis**

To analyze students’ textual responses to WF, the study drew on the text-analytic descriptive tradition (Ferris, 2012). Learners’ form-focused revision operations (FFROs) were deductively coded using categories established in prior multi-dimensional engagement with written feedback research (Han & Hyland, 2015; Zhang & Hyland, 2018). Responses to indirect and direct error treatments were coded using separate but overlapping concepts (illustrated along with the schema for coding content-focused revision operations [CFROs] in Supplemental Material, Parts G and H), reflecting the different cognitive processes involved. Additionally, an error density measure of “number of errors/total number of words × 100” was calculated (van Beuningen et al., 2012) to track students’ accuracy across drafts and compositions. CFROs in response to actionable first-draft comments were deductively coded according to two categories. First, a judgment of how closely the learner had followed feedback instructions was made according to the scheme developed by Christiansen and Bloch (2016). The second was a measure of the success of the revision in terms of whether it made the text “much better,” “better,” “the same,” or “worse” (Christiansen & Bloch, 2016). The public band descriptors were drawn on to facilitate the coding. Results of the text-analytic descriptions are presented as proportions of the overall raw frequency of CFWF and FFWF points. 10% of revision operations (n=30) were recoded six months later to assess intra-rater reliability. The initial and recoded data were compared using an Excel formula, which generated an agreement figure of 0.929, evidence of “good” reliability of coding.

Thematic analysis (Braun & Clarke, 2006; Terry et al., 2017) was applied to the transcribed interview data. The transcripts were read and re-read to attain familiarity, followed by the inductive and iterative development of codes, describing important segments of the data that addressed the three dimensions of engagement for each participant. Codes were then clustered into cross-case themes orientated around a central organizing concept (Terry et al., 2017). For example, “Genuine WF told her what her mistakes were,” WF assisted in providing direction to writing,” “WF helped provide a foundation” along with six other codes were amalgamated into the theme “Unsophisticated explanations of how WF helped writing development.” While the analysis was undertaken without a priori categories in mind, the multi-dimensional model of engagement suggested concepts around which relevant data was coded. These included participants’ value judgments of the WF and emotional responses (the affective dimension), descriptions of WF processing strategies (behavioral), and understandings of the WF (cognitive). Themes were recursively reviewed with reference to the codes and transcripts (Braun & Clarke, 2006). The final themes (see Table 3) were defined once they were deemed to possess the necessary internal homogeneity and external heterogeneity. The results are presented and discussed together, organized according to the themes present across the three dimensions (preceded by an outline of students’ assessed written performance across the project). Content- and form-focused revision operations are addressed under behavioral engagement, mirroring prior research (Han & Hyland, 2015).
Table 3. Themes Present in the Data Across the Three Dimensions of Engagement.

| Dimension          | Theme                                                                 |
|--------------------|----------------------------------------------------------------------|
| Affective engagement | WF collectively was judged positively                               |
|                    | Unsophisticated explanations of how WF helped writing development    |
|                    | WF contributed to increases in confidence                             |
| Behavioral engagement | Negative affective reactions stemmed from performance deficits        |
| Cognitive engagement | Routine, uninformed strategies employed in processing WF             |
|                    | Decoding intended meaning of WF was not straightforward              |
|                    | Not understanding a textual problem highlighted by WF or an item of WF |
|                    | Misinterpreting a textual problem highlighted by WF or an item of WF  |
|                    | Not understanding how to act on WF                                   |

Table 4. Assessed Band Score Outcomes of the Participants’ Three Essays.

|                       | Essay one                   | Essay two                   | Essay three                  |
|-----------------------|-----------------------------|-----------------------------|-----------------------------|
|                       | Draft one | Draft two | Draft one | Draft two | Draft one | Draft two |
| Kushal                 | 7.5      | 7.5       | 7.0      | 7.5       | 7.5      | 7.5       |
| Yuri                  | 6.5      | 6.5       | 6.0      | 6.5       | 6.5      | 7.0       |
| Min Jung              | 5.5      | 5.5       | 5.5      | 5.5       | 5.5      | 6.0       |
| Chandrika             | 6.5      | 6.0       | 6.0      | 6.5       | 6.0      | 6.5       |

Extracts of participants’ authentic utterances mostly serve to illustrate the interpretive assertions of the researcher (Terry et al., 2017).

Results and Discussion

Students’ Written Progress Across the Three Essay Rounds

As shown in Table 4, only Kushal was consistently able to meet his desired overall Writing score, failing to achieve 7.5 just once. In contrast, Yuri managed his target only via revisions in response to WF, falling short by 0.5 on four occasions and by 1.0 in draft one of essay two. Chandrika and Min Jung exhibited consistent shortfalls in performance, with the former achieving 6.0 or 6.5, while the latter was consistently one band below her required 6.5. In terms of progress across the project, Chandrika, Kushal, and Min Jung’s scores in TR, CC, and LR either wholly flattened or exhibited a dip across essays. In just two second drafts did these participants enhance the quality of their writing in any of the criteria by a band (e.g., Chandrika’s LR in essay two and CC in essay three). Yuri was more successful, raising his sub-scores in five separate instances. While this suggests he was better able to engage with WF, most features of his writing seemed borderline 6.0/7.0. That the learners were not able to make short-term upgrades in band scores through WF coheres with the findings of prior research into IELTS Writing preparation (Alsagoafi, 2018; Estaji & Tajeddin, 2012; Hamid, 2016; Rao et al., 2003). However, caution should be exercised in concluding poor written progress since IELTS bands lack sensitivity to smaller-scale changes in L2 writing (Green, 2005; Rao et al., 2003). It may be that additional cycles of essay writing were needed to provide opportunities for responding to WF and/or greater time between essays/drafts to account for delayed uptake.

Affective Engagement

WF collectively was judged positively. All participants consistently stated they highly valued a holistic conception of the written feedback. Representative judgments included, “I have to say I rely on that. So because of that, I made these improvements” (Chandrika), “they were really helpful to be honest, and I would really appreciate if I have some, for example, doubts in the future” (Kushal), and, “it’s really helpful when I write down and you feedback your explanation” (Min Jung). This finding contributes to a large body of research that demonstrates the high regard in which L2 developing writers hold teacher WF generally (Cunningham, 2019a; Ferris, 2011; Hyland & Hyland, 2006; Zacharias, 2007). The value placed on WF in these settings is not surprising since the learners were recruited on the basis of their desire for written feedback, which was noted as difficult to achieve locally (Allen, 2016; Chappell et al., 2019). Nevertheless, attitudes were not unequivocally positive, reflecting the complexity of individuals’ affective responses to WF (Han & Hyland, 2018; Mahfoodh, 2017). Yuri professed that he found 85% of the points helpful, with the other 15% attributed to the need for further alternatives, underscoring that some learners value choices in how to respond (Treglia, 2008). Interestingly, Min Jung (unprompted) posited a comparable figure of 80% that she
deemed helpful but was referring to the tutorial element of the interviews. While this seems a rather damning verdict, it provides support for the claim that L2 students value face-to-face follow up on written feedback (Han, 2017; Hedgcock & Lefkowitz, 1994; Saito, 1994).

The participants disclosed few specific facets of the content and delivery of WF they were positively disposed toward. One feature that was praised was the indirect treatment of errors using the metalinguistic code: “the error code makes us that very like a teaching that we learned from the school” (Chandrika) and, “it’s very structured and helpful” (Kushal). This was unexpected, as it was anticipated the participants would privilege task-related understandings (Alsagoafi, 2018; Chappell et al., 2019; Hamid & Hoang, 2018; Hu & Trenkic, 2019) over general grammar feedback. This finding likely reflects the tendency of L2 learners to perceive correction as crucial in the strive for error-free writing (Ferris, 2011; Hedgcock & Lefkowitz, 1994; Saito, 1994). As in other studies (e.g., Elwood & Bode, 2014), detailed descriptions and explanations of the qualities of student writing through commentary were also highly valued. This stemmed from learners’ impatience to better understand how to achieve their IELTS performance goals (Alsagoafi, 2018; Saïf et al., 2021). In other settings, developing writers sometimes feel unhappy receiving a lot of feedback as it suggests an increased workload during revisions (Mahfoodh, 2017; Yu et al., 2018; Zacharias, 2007).

**Unsophisticated explanations of how WF helped writing development.** In addition to the characteristics of WF provision that were deemed beneficial, a further attitudinal theme encompassed explanations of how WF added value to their writing. These covered a narrow range of feedback features and lacked sophistication, reflecting learners’ limited awareness of the steps required to meet their band score requirements (Allen, 2016; Chappell et al., 2019; Mickan & Motteram, 2009). One explanation, that WF demystified IELTS through explicating textual deficiencies and the means to remedy them (Treglia, 2008), cohered with a belief in WF’s critical role in this context: “Because after this revision, I know more about my flaws. What are the specific areas that I should focus in writing?” (Chandrika). Interestingly, it was the weaker writers, Chandrika and Min Jung, who stated that the process of undertaking revisions added value to their participation in the project: “For me is much more helpful, revise and review again. So it can be what I got what was wrong, what I have to not do again in the future” (Min Jung). The engagement of lower-level learners with WF is documented to take place at the surface level (Barkaoui, 2007; Porte, 1997; Radecki & Swales, 1988). However, since both Chandrika and Min Jung fell short of their target scores, they had good reason to perceive value in undertaking content revisions. CFROs enabled them to hone the quality of their writing incrementally, as achieving their goals in authentic test conditions appeared unrealistic.

In contrast, Kushal and Yuri’s initial lack of value attached to revising drafts stemmed from limited experience undertaking revisions in these settings and a belief in simulating rehearsed writing in authentic test-like settings. For Kushal, this meant much problematic content in essay two was addressed through renewed simulation, as opposed to meaningful engagement with the comments. As a cognitively overwhelmed writer, it is perhaps of little surprise that Yuri merely stated, “it’s hard to say something concrete about this”. He perceived baby steps across the essays, unable to provide explanations of how WF usefully served his needs. Finally, the clarity, structure, and ease of navigation underlying WF presentation in the Kaizena platform using the comment and highlight functions were cited as adding value (with the exception of Yuri), providing further evidence L2 learners are open to new technologies that facilitate writing development (Chong, 2020; Cunningham, 2019b). The participants found the possibility to respond to individual comments reassuring, although rarely utilised this functionality.

**WF Contributed Increases in Test-Taking Confidence.** It is perhaps unrealistic to expect intermediate-level learners to elaborate the precise mechanism(s) in which comprehensive WF enhanced their written development. A more prevalent theme underlying learners’ appraisals was the value added through raising their self-confidence in meeting task expectations, a finding mirroring studies into candidates’ perspectives toward IELTS Listening and Speaking preparation (Chappell et al., 2019; Winke & Lim, 2014; Yang & Badger, 2015). Interestingly, the confidence-boosting power of WF was even stated by Chandrika and Min Jung, who never met their goals in a rehearsal composition: “it’s so much valuable time for me, because I really got to know. So I think from now, I feel quite confidence” (Min Jung). As in the Speaking module (see Yang & Badger, 2015), greater confidence accompanied learners’ growing familiarity with what it was like responding to topics parallel to test tasks: “I have discovered some ideas in my mind regarding this topic, for example, and if I am to face this all at least similar topic again” (Yuri). Furthermore, improved confidence derived from tackling a range of prompt topics and frames (although Yuri stated, “participants should experience all the type of the essays”), being more aware of one’s own mistakes, and gaining a “feel” for higher-level content through reading the reformulations.

Since the written feedback tended to be highly evaluative, there was a propensity for it to constitute a threat to self-esteem. The lower-than-expected essay scores accompanied by critical feedback undermined the confidence of Min Jung and Yuri. In the case of Min Jung, this resulted in instances where she was afraid to make revisions because, “I don’t want to be the wrong anymore.” Similarly, despite meeting his targets in several criteria and overachieving in the second draft of essay three, Yuri felt concerned about replicating
such performance in future: “It's formidable to understand, what exactly do they need to do in order to achieve the same result”. Self-doubts persisted over the project owing to the unpredictability of prompt topics/frames and worries over an inability to replicate successful performance in authentic settings (Estaji & Tajeddin, 2012). It is probably only after achieving their desired outcomes in the actual test that many learners truly feel confident in their abilities to perform at the requisite level.

Negative Affective Reactions Associated with Not Knowing How to Adjust Written Performance. Min Jung and Yuri’s emotional reactions across the project demonstrated that, like in other contexts (Hyland & Hyland, 2001; Lee, 2008; Zacharias, 2007), WF can exert a detrimental affective impact on learners. It is hardly unsurprising that these learners evinced negative reactions as they consistently performed 0.5 to 1.0 band lower than their targets (especially in first drafts). Both exhibited disappointment stemming from the frustration and despair associated with not being able to accomplish a task to the required level (Estaji & Tajeddin, 2012), as exemplified by Yuri: “I’ve grown a bit weary of this. . . due to the fact that I’ve been struggling for this for two years with this.” To be told repeatedly they were constantly making mistakes negatively affected Min Jung and Yuri. It progressively exasperated Min Jung, lowering her confidence in her abilities and creating self-doubts: “I don’t know when I can finish about IELTS.” Yuri’s case illustrates where a mismatch between teachers’ responding behaviors and students’ desires can induce negative affective responses (Hyland, 1998). The abundance of feedback that did not contribute concrete understandings of expected test performance at band 7.0 bred anxiety: “I’m almost driven crazy with IELTS uncertainty. . . because nobody can say for sure for certain that it's 100% way to do this,” which eventually resulted in disagreement with some WF.

It was evident Min Jung and Yuri’s feelings of frustration and disappointment also stemmed from lower-than-expected written performance. IELTS test-takers’ assumption of linear gains across multiple test undertakings and confusion with inconsistent scores is documented in the literature (Hamid, 2006; Pearson, 2019). It may have been accentuated in the present study since the learners had recruited outside expertise to supplement existing preparation activities and implicitly expected progression over the project. Min Jung, particularly, held the faulty assumption that written progress directly matched between teachers’ responding behaviors and students’ reactions that Yuri was the most compliant learner in addressing CFWF, attending to 77.7% of comments. His obstinacy to revise is more visible in the low proportion of text omitted in response to CFWF (2.8%) and the 19.4% of comments that were ignored. Additionally, while evincing the highest amount of content revisions that made the texts “much better” (19.4%), 41.7% provided no improvement to text omitted in response to CFWF (2.8%) and the 19.4% of comments that were ignored. Furthermore, the high stakes of the learning context served to induce an activating response (Pekrun, 2006), mitigating the discouragement that can stem from comprehensive critical feedback (Lee, 2008; Zacharias, 2007). This motivated Min Jung and Yuri to persevere in spite of their disappointment or distrust in the assigned scores.

**Behavioral Engagement**

**Content- and form-focused revision operations.** A key measure of engagement is the extent learners are willing and able to undertake textual revisions in response to WF, operationalized textually as frequency counts of deductively coded revision operations and their outcomes (Han & Hyland, 2015; Yu et al., 2018; Zhang & Hyland, 2018). Active behavioral engagement is important since learners are unlikely to make progress unless there is an underlying sense of personal agency and willingness to use WF (Barkaoui, 2007; Price et al., 2010). Of particular note were CFROs, as a relatively small number of TR and CC issues made the difference in whether the participants achieved their desired band scores or not. Reluctance to address textual features that contributed to tangible band score deficits attenuated both assessed outcomes and writing development. Much CFWF in this context suggested how a text may conform more closely to a conception of higher-level writing, constituting the phenomenon of teacher appropriation (Tardy, 2019). The literature characterizes L2 writers as either willing to grant the teacher absolute power to appropriate, complying with their demands submissively, or by “contesting” comments through non or perfunctory revisions (Radecki & Swales, 1988). Investing in CFROs reveals developing writers’ willingness to accept the expertise of the feedback provider and regard criticisms as suggestions that can help polish compositions (Orsmond & Merry, 2013).

The CFROs uncovered in the study varied notably among the participants, a finding consistent with other multi-dimensional engagement studies (e.g., Yu et al., 2018; Zhang & Hyland, 2018). However, the results should be interpreted cautiously as it appeared certain patterns in learners’ revision operations were not always consistent with the interview findings. Indeed, this is one of the advantages of the multi-dimensional model, as one perspective alone provides insufficient insights (Han & Hyland, 2015). As shown in Table 5, it is perhaps surprising in light of his negative affective responses that Yuri was the most compliant learner in addressing CFWF, attending to 77.7% of comments. His obstinacy to revise is more visible in the low proportion of text omitted in response to CFWF (2.8%) and the 19.4% of comments that were ignored. Additionally, while evincing the highest amount of content revisions that made the texts “much better” (19.4%), 41.7% provided no improvement to textual quality, stemming from issues envisaging what acting on the WF entailed.

A learner who, behaviorally, was a WF receptor (Radecki & Swales, 1988) was Chandrika, whose CFROs followed instructions in 65.1% of instances, with only 13% of
comments being ignored. As an inexperienced candidate, high levels of WF compliance likely reflected self-doubts in her abilities and high levels of trust in the expertise of the WF provider. Among the participants, Chandrika recorded the greatest amount of revision operations that made her texts “better” or “much better” (69.6%). An important caveat is that this figure reflects her notable first draft underperformance vis-à-vis what she could achieve in less controlled conditions. Kushal’s proportions of revision operations that followed instructions were 44.4%, with an equal figure that improved his texts. These were disappointing outcomes in light of his test experience and language proficiency. Omitted text constituted 33.3% of all CFROs, notably higher than the other three learners. This does not necessarily suggest a lack of engagement (Uscinski, 2017), rather, that Kushal perceived revisions as second attempts to generate essay content in simulated test conditions.

Min Jung followed the instructions of 39.9% of comments (5.7% fully), evidently a concern resulting from weaknesses in ELP and cognitive engagement. The figure highlights that following the feedback provider’s instructions should not be equated to willingness to revise, as response to commentary is clearly contingent on understandings of the underlying intent of the WF. This is further illustrated by the lowly 37.2% of revisions that resulted in textual improvements. The worryingly high 42.9% of comments that were ignored should also be interpreted cautiously. Disregarding comments can be an indication of disengagement (Han & Hyland, 2015) or growing writer autonomy and agency (Mahfoodh, 2017), characteristics more applicable to Yuri and Kushal. For Min Jung, this behavior stemmed ultimately from noticeable weaknesses in ELP since problematic textual issues often cut across multiple assessment criteria, increasing the difficulties of clear and concise written feedback provision and its successful resolution.

Table 6 shows participants’ lexicogrammatical error rates per 100 words across the essays. Kushal was the sole learner to exhibit a reduced error rate across first drafts, probably because he was “only” required to lower the frequency of errors from “a few” to “occasional.” As Müller (2015) stresses, the IELTS band 7.0 writer is vastly different in quality to the band 6.0 writer, with six times fewer distracting errors, none of which impinge on communicative quality. At 6.0, there are a diffuse array of treatable and untreatable errors that require learners’ attention, taxing their cognitive capabilities during processing. That three learners did not demonstrate meaningful gains in lexicogrammatical accuracy coheres with the findings of several studies investigating the impact of cognitively-demanding unfocused WCF (Frear & Chiu, 2015; Sheen et al., 2009). Kushal and Yuri consistently reduced their respective error rates in each second draft, providing further evidence of the effectiveness of indirect error treatment (Ferris, 1995; Ferris et al., 2011), even when the pressure to resolve surface-level issues is heightened by
notable global concerns. However, this outcome did not apply to learners whose accuracy levels were lower (i.e., Chandrika and Min Jung) owing to ELP limitations and significant content issues vis-à-vis their desired band.

Looking at the form-focused revision operations in Table 7, it is apparent a low proportion of errors were ignored in comparison to other studies of student engagement (Han & Hyland, 2015; Uscinski, 2017; Zheng & Yu, 2018). This is likely because the participants were motivated to resolve errors as a means of improving performance in LR and GRA. Incorrect resolutions were low for all participants (below 21%), although it is surprising that the linguistically strongest learner (Kushal) accounted for the highest proportion of failed resolutions. Deletions encompassed notable proportion of Kushal (45%), Yuri (43.5%), and Chandrika’s (33.3%) error resolutions. Deletions often resulted from the requirement for substantial content revisions, calling into question the merits of instructors investing time and effort in metalinguistic corrections that may ultimately be ignored. Kushal and Min Jung appeared the most compliant in incorporating direct corrections on first drafts, in contrast to Yuri who usually removed directly treated errors.

**Routine, Uninformed Strategies Employed in Processing WF.** The other aspect to learners’ behavioral engagement with written feedback was the explicit strategies and skills they employed when processing and responding to written feedback, explored in the interviews. It was found, regardless of proficiency level, the learners reported routine, uninformed strategies and skills when processing the WF. Typically, they conceptualized processing as a matter of “going through” feedback points one-by-one (Chandrika, Kushal) or merely contemplating how to respond (Min Jung, Yuri). Cunningham (2019a) cautions that the mere reading of feedback should not necessarily be considered insufficient (in comparison to not reading the feedback), although in the context of Task 2 underperformance and the restrictions to achieving goals that result, learners ought to be relied upon to read the feedback. Processing the feedback was accompanied by mostly surface level skills and strategies (Porte, 1997; Yu et al., 2018; Zheng & Yu, 2018). These included making sense out of the information (Chandrika), determining if the WF point required action or not (Kushal), and deciding which sentences to add or remove (Min Jung). Not much time was spent during processing and response, although Min Jung’s pre-occupation with IELTS preparation and significant performance deficits resulted in a remarkable 3 hours spent revising essay two after a 5-hour self-study session.

There are a number of explanations of the surface-level approaches to written feedback processing evinced by the participants. Since the revision of rehearsal essays is uncommon in these settings, the participants lacked knowledge and experience utilizing WF to develop their writing and test-taking skills ( Barkaoui, 2007; Porte, 1997). They rarely undertook supplementary learning activities, few of which addressed language proficiency, a characteristic of IELTS preparation generally (Gan, 2009; Saif et al., 2021; Smirnova, 2017). Secondly, perhaps because the revision of essay drafts was considered unusual and perceived as for the benefit of the researcher, there was a lack of initial participant buy-in (Yu et al., 2018). Finally, the presence of largely facile revision strategies is perhaps not surprising since, other than a basic set of guidelines for undertaking revisions, the participants did not receive training on how to act on the written feedback.

**Cognitive Engagement**

Decoding intended meaning of WF not straightforward. Owing to its role in explaining/modeling desired performance where a gap was exhibited, the written feedback ended up being comprehensive and unfocused. As a consequence, engagement was cognitively taxing for all participants, regardless of performance level and goals. Kushal noted, “I need to be more, like, focused when I’m going through those feedbacks,” while Chandrika felt, “I have to read more times your feedback when I get something new.” For Yuri, not knowing how to respond resulted in him spending, “15 or 20minutes to think what I’m gonna do what?” This finding mirrors studies situated in other L2 writing settings that posit successful

### Table 7. Participants’ Form-Focused Revision Operations.

|                | Kushal (n = 20) | Yuri (n = 22) | Min Jung (n = 36) | Chandrika (n = 57) |
|----------------|----------------|--------------|------------------|-------------------|
| Indirect corrections | Correct revision | 25.0% | 43.5% | 47.2% | 36.8% |
|                 | Incorrect revision | 20.0% | 4.3%  | 19.4% | 12.3% |
|                 | Deletion | 45.0% | 43.5% | 22.2% | 33.3% |
|                 | Substitution | 0.0%  | 0.0%  | 2.8%  | 12.3% |
|                 | No revision | 10.0% | 4.3%  | 8.3%  | 5.3%  |
| Direct corrections | Disregard | 0.0%  | 10.0% | 7.7%  | 30.0% |
|                 | Incorporate | 75.0% | 10.0% | 92.3% | 50.0% |
|                 | Remove | 25.0%  | 80.0% | 0.0%  | 20.0% |
cognitive engagement with WF requires considerable mental effort (Kim & Bowles, 2019; Yu et al., 2018; Zheng & Yu, 2018), largely in consideration of how and to what extent texts should be revised (Storch & Wigglesworth, 2010; Zheng & Yu, 2018). Such difficulties may have been mitigated had more thorough guidance been provided on how to process and use the feedback, which can reduce misunderstandings (Elwood & Bode, 2014; Uscinski, 2017). Alternatively, a mid-focused WF approach, defined by a focus on two to five pre- eminent textual features, may have lessened the cognitive burden, although at the expense of certain textual issues being attended to and in contravention of learners’ desires for comprehensive WF.

An added difficulty was developing shared understandings of criterion-referenced statements of written performance. Notably, the learners struggled to understand how characteristics of their writing cohered with statements outlined in the PBDs. This encompassed localized concerns such as Kushal bemoaning, “occasional errors I don’t know what does that mean” as well as issues understanding marks awarded overall: “I’ve got only also this 6 band in task response, but actually, I can’t understand how is it does it work? You know this? Why 6? Why not 5 or why not 7?” (Yuri). This was because the students lacked access to the discourse in which the information was presented (Chanock, 2000), which is targeted toward language assessment specialists. Unfortunately, detailed descriptive/explanatory written feedback could not overcome the lack of clarity inherent in IELTS’ general proficiency model of language assessment (Davies, 2008). As such, the merits of the public band descriptors as a pedagogical tool to facilitate learners’ written development seem limited.

In fact, Chandrika, Min Jung, and Yuri varyingly became cognitively overwhelmed by the WF as the project progressed (Han & Hyland, 2015; Yu et al., 2018; Zhang & Hyland, 2018). Yuri noted, “it's hard to develop these ideas when you have this mess in your head with this lots of questions why this? Why not this?” a perspective echoed by Chandrika: “there are so many things that I can’t tell everything that are running in my mind”. Feeling cognitively overwhelmed stemmed from the learners being unable to utilize WF to address performance deficits. Overloaded with comprehensive and unfocused information that lacked a clear focus (Bitchener, 2008; Sheen, 2007), these three participants lacked sufficient attentional capacity to process and respond to the multitude of global- and surface-level textual features (Bitchener, 2008; Sheen, 2007). Repeated over a series of drafts, the outcome was cognitive overload, confusion, and discouragement, which may have resulted in the quiet resistance to some written feedback (Han & Hyland, 2015; Radecki & Swales, 1988), particularly with regard to TR-focused points on Yuri’s essay three and Min Jung’s subsequent re-use of memorized material. That WF in these settings has the propensity to cause affective harm raises serious questions over the impact of IELTS preparation on learners who are not close to attaining their goals.

**Not understanding a textual problem highlighted by WF.** Successful behavioral engagement with CFWF was hampered by the fact that the learners did not always understand what particular comments “meant,” a common theme in both L1 (Chanock, 2000; Weaver, 2006) and L2 written feedback research (Conrad & Goldstein, 1999; Hyland, 1998; Mahood, 2017). Failure to understand CFWF encompassed misinterpreting or not understanding the nature of a textual issue highlighted by WF or an item of WF itself and not knowing how to act on a comment (accompanied by awareness of the issue). Incidences where the participants professed or demonstrated non-understanding of a textual issue or item of WF were relatively low, occurring just 14 times across the dataset. This was somewhat surprising given the outcomes of learners’ CFROs (Table 5) and the relatively poor progress across the project as measured in band scores. On the other hand, the learners were not complete novices to writing in this context, which might explain why not understanding how to act on commentary constituted a noticeably more common source of cognitive difficulty (23 instances).

**Misinterpreting the Nature of a Textual Issue Targeted by WF.** Since the written feedback was comprehensive, unfocused, and married to linguistically complex criterion-referenced descriptors that specify performance in general terms, there was great potential for misunderstanding criterion-referenced descriptors of performance. As indicated in other studies of WF on L2 writing (Conrad & Goldstein, 1999; Hyland, 1998; Hyland & Hyland, 2001). It is, thus, perhaps a positive indicator that “only” 14 instances emerged in the interviews where the participants had misunderstood the intent of the WF or a textual issue it was targeting. Yet since the rates of fully following the directions of CFWF were between 5.7% to 22.2%, clearly the learners encountered difficulties interpreting what was expected of them. It is likely that linguistic shortcomings caused them to misunderstand the intention behind CFWF, perhaps because they were unable to decode messages obfuscated by burdensome language assessment terminology or hedging that was employed to “soften the blow” (Hyland & Hyland, 2001).

One trend evinced by the weaker writers, Chandrika and Min Jung, was the conflation of content- and form-focused concerns. Min Jung understood the feedback condemned memorized generic academic-sounding vocabulary. However, this resulted in a perception that improving her response to the task was a matter of solely using the right vocabulary: “Task response always a problem, how to describe about the making suitable words, like making suitable using words.” Similarly, Chandrika understood her second drafts as lengthy
but felt her inability to write in, “a proper way” was due to a low range of vocabulary. Indeed, she noted, “if you don’t have that much of vocabulary to express one example, so you can that mean I can go for two [supporting ideas].” This divergence is reflective of the dichotomy whereby teachers perceive revision as a generative process where meaning is reassessed and a text is reshaped, yet inexperienced learners view it as the correction of surface-level errors (Barkaoui, 2007; Radeck & Swales, 1988). Radeck and Swales (1988) conclude that such a narrow attitude “can only hinder their development as L2 writers” (p. 364), a perspective that coheres with the findings of this study.

Not Understanding How to Act on Written Feedback. The most notable characteristic of participants’ cognitive engagement was a lack of purported understandings of how to act on CFWF, despite awareness of the issue(s) shown. Failure to generate suitable revisions to address teacher commentary is not uncommon in the literature (Christiansen & Bloch, 2016; Conrad & Goldstein, 1999; Hyland & Hyland, 2006; Mahfooodh, 2017), although is a complex, highly learner-dependent issue. This is illustrated in the present study by the inability of Chandrika to characterize her written ideas with relevance, clarity, and development simultaneously, of Kushal to translate understandings of needing to be more focused into reality, of Min Jung to textually realize how to structure a clearly developed argument regardless of topic, and of Yuri to develop ideas that seemed complete and incorporate content suggestions he disagreed with. As Price et al. (2010) stress, being able to act on written feedback is crucial, underscoring its role in closing performance gaps in skill-based settings through feed-forward (Price et al., 2010) and usability (Walker, 2009). To do either, WF must be designed to “help the student to reduce or close the gap” (Walker, 2009, p. 68). Nevertheless, the association between ELP limitations and notable performance gaps made specifying expected revision strategies beyond native-speaker reformulations in linguistically simplified terms challenging.

Another explanation was participants’ inability to conceive of what response to WF at their desired band score entailed, a finding of prior research (Chappell et al., 2019). One consistent explanation across multiple participants was a deficit of topic knowledge (Craven, 2012). Chandrika explained in regard to prompt two, “there’s a law and age limitation in our countries, but still, I don’t know how they use those things,” a point also made by Yuri; “I can’t be aware of everything. You know, how it works in other countries.” There was also a linguistic dimension to this theme, that is, the participants whose ELP deficiencies resulted in essay ideas characterized as consistently unclear or inadequately developed (Chandrika, Min Jung, and Yuri) lacked the linguistic ability to discern why this was the case. Instead of Remedying a deficit of content knowledge, for example, through reading about the essay topics online, as shown in Table 5, the learners often deleted content highlighted as deficient (Conrad & Goldstein, 1999; Hyland, 1998), perhaps to avoid the issue. The situation was exacerbated by the nebulous Task 2 assessment criteria, exemplified by Kushal’s struggle to comprehend how examiners delineate a “well-developed” response and Yuri’s skepticism of his ideas being “unclear.”

Conclusions

As the study explored how a small sample of candidates preparing for IELTS (with an overrepresentation of test veterans) engaged with written feedback using a case study approach, the findings are not generalizable. Additionally, the research was simplified for the manageability of data collection, analysis, and reporting by a single researcher. A prominent omission was participants’ cognitive engagement with FFWF, which was beyond the time constraints of the interviews, particularly as discussion of often-complex CFWF points proved time consuming. A further limitation stems from the multi-dimensional model of student engagement itself. While it has exhibited increasing acceptance among scholars in tertiary-level learning-to-write contexts (Yu et al., 2018; Yu & Jiang, 2020; Zheng et al., 2020), facets of the approach were not entirely well suited to exploring engagement in this context. These include the emphasis on surface-level issues (which comprise half of the Task 2 criteria), behavioral engagement operationalized predominantly as revision operations (since content changes influenced FFROs, especially deletions and removals, and revisions are uncommon in this context), and a lack of sophisticated schema for coding CFROs. Similarly, coding qualitative data as one particular dimension seemed reductive and compartmentalizing. It was not always possible to say for certain that a code was positioned within a single dimension, with the nebulous border between behavioral/cognitive engagement particularly problematic.

The study found that, despite detailed written feedback that explained and exemplified performance at participants’ target band, student engagement was insufficient to bridge performance deficits across the project. Cognitively, the learners did not always understand the intentions behind comments or what actions to take to resolve problematic textual features within how the task is assessed (Conrad & Goldstein, 1999; Hyland, 1998). Behaviorally, despite successful error resolutions across drafts, meaningful improvements in lexicogrammatical accuracy were not apparent, while notable content changes between drafts meant many errors went uncorrected. Processing WF tended to be at the surface-level (Yu et al., 2018; Zheng & Yu, 2018), accompanied by a lack of buy-in to making content-focused revisions in the early stages of the project and a tendency to omit faulty content rather than meaningfully address it. Affectively, the WF was highly valued and deemed confidence building
(Chappell et al., 2019; Yang & Badger, 2015), contributing an activating effect (Pekrun, 2006). Nevertheless, underperformance accompanied by comprehensive, unfocused WF information resulted in two learners feeling anxious, frustrated, and sometimes overwhelmed, although the pressure to meet their IELTS goals mitigated the propensity of WF in these settings to be deactivating.

The findings of the study feature implications for practitioners of IELTS preparation who provide written feedback to prospective test-takers. First, feedback providers are advised to be cautious about what can be achieved and to manage students’ expectations of WF being a panacea to notable deficits in performance. Practitioners may seek to focus on a few key features across learners’ compositions in relation to the public band descriptors, rather than overloading them with comprehensive WF that lacks a clear focus (Bitchener, 2008; Sheen, 2007). To address TR and CC issues, it may be preferable to model performance at the learners’ required level by appropriating students’ texts, rather than time-consuming descriptions and explanations of performance. Additionally, teachers should consider whether comprehensive grammar correction is worth their time, particularly if students are tasked with undertaking content-based re-writes. It might be advisable to adopt a ‘mid-focused’ approach to FFWF, highlighting a handful of noteworthy recurring errors, or focusing on ones that disturb comprehension as these are penalized more heavily. Alternatively, FFWF could be delivered separately once students have had the opportunity to engage with CFWF. Future research in face-to-face settings with a larger cohort of learners is required before more equivocal judgments concerning the effectiveness of WF in IELTS Writing Task 2 preparation settings can be made.

**Declaration of Conflicting Interests**
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Funding**
The author(s) received no financial support for the research, authorship, and/or publication of this article.

**Ethics Statement**
Approval of the University of Exeter’s Research Ethics and Governance Office was obtained prior to the commencement of data collection (Reference Number D1920-049).

**Consent**
All participants provided their written consent to participate in the study by digitally signing a bespoke participant information sheet.

**ORCID iD**
William S. Pearson [https://orcid.org/0000-0003-0768-8461](https://orcid.org/0000-0003-0768-8461)

---

**Supplemental Material**
Supplemental material for this article is available online.

**References**
Allen, D. (2016). Investigating washback to the learner from the IELTS test in the Japanese tertiary context. *Language Testing in Asia, 6*(7), 1–20. [https://doi.org/10.1186/s40468-016-0030-z](https://doi.org/10.1186/s40468-016-0030-z)

Alsagoofi, A. (2018). IELTS economic washback: A case study on English major students at King Faisal University in Al-Hasa, Saudi Arabia. *Language Testing in Asia, 8*(1), 1–13. [https://doi.org/10.1186/s40468-018-0058-3](https://doi.org/10.1186/s40468-018-0058-3)

Barkaoui, K. (2007). Revision in second language writing: What teachers need to know. *TESL Canada Journal, 25*(1), 81–92. [https://doi.org/10.18806/tesl.v25i1.109](https://doi.org/10.18806/tesl.v25i1.109)

Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of Second Language Writing, 17*(2), 102–118. [https://doi.org/10.1016/j.jslw.2007.11.004](https://doi.org/10.1016/j.jslw.2007.11.004)

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. [https://doi.org/10.1191/1478088706qp063oa](https://doi.org/10.1191/1478088706qp063oa)

Chanock, K. (2000). Comments on essays: Do students understand what tutors write? *Teaching in Higher Education, 5*(1), 95–105. [https://doi.org/10.1080/135625100149489](https://doi.org/10.1080/135625100149489)

Chappell, P., Yates, L., & Benson, P. (2019). Investigating test preparation practices: Reducing risks (IELTS Research Reports Online Series, No. 3). British Council, Cambridge Assessment English and IDP, IELTS Australia. [https://www.ielts.org/-/media/research-reports/2019-3-chappell_et_al_layout.aspx](https://www.ielts.org/-/media/research-reports/2019-3-chappell_et_al_layout.aspx)

Chong, S. W. (2020). A research report: Theorizing ESL community college students’ perception of written feedback. *Community College Journal of Research and Practice, 44*(6), 463–467. [https://doi.org/10.1080/10668926.2019.1610675](https://doi.org/10.1080/10668926.2019.1610675)

Christiansen, M. S., & Bloch, J. (2016). “Papers are never finished, just abandoned”: The role of written teacher comments in the revision process. *Journal of Response to Writing, 2*(1), 6–42. [http://www.journalrw.org/index.php/jrw/article/view/32](http://www.journalrw.org/index.php/jrw/article/view/32)

Coffin, C. (2004). Arguing about how the world is or how the world should be: The role of argument in IELTS tests. *Journal of English for Academic Purposes, 3*(3), 229–246. [https://doi.org/10.1016/j.jeap.2003.11.002](https://doi.org/10.1016/j.jeap.2003.11.002)

Conrad, S. M., & Goldstein, L. M. (1999). ESL student revision after teacher-written comments: Text, contexts, and individuals. *Journal of Second Language Writing, 8*(2), 147–179. [https://doi.org/10.1016/S1060-3743(99)80126-X](https://doi.org/10.1016/S1060-3743(99)80126-X)

Craven, E. (2012). *The quest for IELTS Band 7.0: Investigating English language proficiency development of international students at an Australian university* (IELTS Research Reports, Volume 13). IDP: IELTS Australia and British Council. [https://www.ielts.org/-/media/research-reports/ielts_rr_volume13_report2.aspx](https://www.ielts.org/-/media/research-reports/ielts_rr_volume13_report2.aspx)

Cunningham, J. M. (2019a). Composition students’ opinions of and attention to instructor feedback. *Journal of Response to Writing, 5*(1), 4–38. [https://journalrw.org/index.php/jrw/article/view/133/90](https://journalrw.org/index.php/jrw/article/view/133/90)

Cunningham, K. J. (2019b). Student perceptions and use of technology-mediated text and screencast feedback in ESL writing. *Computers and Composition, 52*, 222–241. [https://doi.org/10.1016/j.compcom.2019.02.003](https://doi.org/10.1016/j.compcom.2019.02.003)
Davies, A. (2008). Assessing academic English language proficiency: 40+ years of U.K. language tests. In J. Fox, M. Wescie, D. Bayliss, L. Cheng, C. E. Turner, & C. Doe (Eds.), Language testing reconsidered (pp. 73–86). University of Ottawa Press.

Ellis, R. (2010). Epilogue: A framework for investigating oral and written corrective feedback. *Studies in Second Language Acquisition, 32*(2), 335–349. https://doi.org/10.1017/S0272263109990544

Elwood, J. A., & Bode, J. (2014). Student preferences vis-à-vis teacher feedback in university EFL writing classes in Japan. *System, 42*(1), 333–343. https://doi.org/10.1016/j.system.2013.12.023

Estaji, M., & Tajeddin, Z. (2012). The learner factor in washback context: An empirical study investigating the washback of the IELTS academic writing test. *Language Testing in Asia, 2*(1), 5–25. https://doi.org/10.1186/2229-0443-2-1-5

Fan, Y., & Xu, J. (2020). Exploring student engagement with peer feedback on L2 writing. *Journal of Second Language Writing, 50*, 100775. https://doi.org/10.1016/j.jslw.2020.100775

Ferris, D. R. (1995). Student reactions to teacher response in multiple-draft composition classrooms. *TESOL Quarterly, 29*(1), 33–53. https://doi.org/10.2307/3587804

Ferris, D. R. (2011). *Treatment of error in second language student writing* (2nd ed.). The University of Michigan Press.

Ferris, D. R. (2012). Written corrective feedback in second language acquisition and writing studies. *Language Teaching, 45*(4), 446–459. https://doi.org/10.1017/S0261444812000250

Ferris, D. R., Brown, J., Liu, H. S., & Stine, M. E. A. (2011). Responding to L2 students in college writing classes: Teacher perspectives. *TESOL Quarterly, 45*(2), 207–234. https://doi.org/10.5054/tq.2011.247706

Frear, D., & Chiu, Y. H. (2015). The effect of focused and unfocused indirect written corrective feedback on EFL learners’ accuracy in new pieces of writing. *System, 53*, 24–34. https://doi.org/10.1016/j.system.2015.06.006

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74*(1), 59–109. https://doi.org/10.3102/00346543074001059

Gan, Z. (2009). IELTS preparation course and student IELTS performance: A case study in Hong Kong. *RELJ Journal, 40*(1), 23–41. https://doi.org/10.1177/0033688208101449

Green, A. (2005). EAP study recommendations and score gains on the IELTS Academic Writing test. *Assessing Writing, 10*(1), 44–60. https://doi.org/10.1016/j.asw.2005.02.002

Hamid, M. O. (2016). Policies of global English tests: Test-takers’ perspectives on the IELTS retake policy. *Discourse: Studies in the Cultural Politics of Education, 37*(3), 472–487. https://doi.org/10.1080/1596306.2015.1069178

Hamid, M. O., & Hoang, N. T. H. (2018). Humanising language testing. *TESL-EJ, 22*(1), 1–20. http://www.tesl-ej.org/wordpress/issues/volume22/ej85/ej85a5/

Han, Y. (2017). Mediating and being mediated: Learner beliefs and learner engagement with written corrective feedback. *System, 69*, 133–142. https://doi.org/10.1016/j.system.2017.07.003

Han, Y., & Hyland, F. (2015). Exploring learner engagement with written corrective feedback in a Chinese tertiary EFL classroom. *Journal of Second Language Writing, 30*, 31–44. https://doi.org/10.1016/j.jslw.2015.08.002

Han, Y., & Hyland, F. (2019). Academic emotions in written corrective feedback situations. *Journal of English for Academic Purposes, 38*, 1–13. https://doi.org/10.1016/j.jeap.2018.12.003

Han, Y., & Xu, Y. (2021). Student feedback literacy and engagement with feedback: A case study of Chinese undergraduate students. *Teaching in Higher Education, 26*(2), 181–196. https://doi.org/10.1080/13562517.2019.1648410

Hedgecock, J., & Lefkowitz, N. (1994). Feedback on feedback: Assessing learner receptivity to teacher response in L2 composing. *Journal of Second Language Writing, 3*(2), 141–163. https://doi.org/10.1016/1060-3743(94)00012-4

Hu, R., & Trenkic, D. (2019). The effects of coaching and repeated test-taking on Chinese candidates’ IELTS scores, their English proficiency, and subsequent academic achievement. *International Journal of Bilingual Education and Bilingualism, 24*(10), 1486–1501. https://doi.org/10.1080/13670050.2019.1691498

Hyland, F. (1998). The impact of teacher written feedback on individual writers. *Journal of Second Language Writing, 7*(3), 255–286. https://doi.org/10.1016/S1060-3743(98)90017-0

Hyland, F., & Hyland, K. (2001). Sugaring the pill: Praise and criticism in written feedback. *Journal of Second Language Writing, 10*(3), 185–212. https://doi.org/10.1016/S1060-3743(01)00038-8

Hyland, K., & Hyland, F. (2006). Feedback on second language students’ writing. *Language Teaching, 39*(2), 83–101. https://doi.org/10.1017/S0261444806003399

IELTS. (2019a). *Guide for educational institutions, governments, professional bodies and commercial organisations*. Cambridge Assessment English, The British Council, IDP Australia. https://www.ielts.org/-/media/publications/guide-for-institutions/ielts-guide-for-institutions-2015-uk.aspx

IELTS. (2019b). *IELTS Task 2 Writing band descriptors* (Public version). https://takeielts.britishcouncil.org/sites/default/files/ielts_task_2_writing_band_descriptors.pdf

Ingram, D., & Bayliss, A. (2007). *IELTS as a predictor of academic language performance, part 1* (IELTS Research Reports, Volume 7). British Council and IELTS Australia Pvt Limited. https://www.ielts.org/-/media/research-reports/ielts_rr_volume07_report3.ashx

Ivančič, R., & Satchwell, C. (2007). Boundary crossing: Networking and transforming literacies in research processes and college courses. *Journal of Applied Linguistics, 4*(7), 101–124. https://doi.org/10.1558/japl.v4i1.101

Kim, H. R., & Bowles, M. (2019). How deeply do second language learners process written corrective feedback? Insights gained from think-alouds. *TESOL Quarterly, 53*(4), 913–938. https://doi.org/10.1002/tesq.522

Lee, I. (2008). Student reactions to teacher feedback in two Hong Kong secondary classrooms. *Journal of Second Language Writing, 17*(3), 144–164. https://doi.org/10.1016/j.jslw.2007.12.001

Mahfoodh, O. H. A. (2017). “I feel disappointed”: EFL university students’ emotional responses towards teacher written feedback. *Assessing Writing, 31*, 53–72. https://doi.org/10.1016/j.jslw.2016.07.001

Mickan, P., & Motteram, J. (2009). *The preparation practices of IELTS candidates: Case studies* (IELTS Research Reports,
Saito, H. (1994). Teachers’ practices and students’ preferences for feedback on second language writing: A case study of adult ESL Learners. *TESL Canada Journal, 11*(2), 46–70. https://doi.org/10.18806/tesl.v11i2.633

Sheen, Y. (2007). The effect of focused written corrective feedback and language aptitude on ESL learners’ acquisition of articles. *TESOL Quarterly, 41*(2), 255–283. https://doi.org/10.1002/j.1545-7249.2007.tb00059.x

Sheen, Y., Wright, D., & Moldaw, A. (2009). Differential effects of focused and unfocused written correction on the accurate use of grammatical forms by adult ESL learners. *System, 37*(4), 556–569. https://doi.org/10.1016/j.system.2009.09.002

Smirnova, E. A. (2017). Using corpora in EFL classrooms: The case study of IELTS preparation. *RELJ Journal, 48*(3), 302–310. https://doi.org/10.1177/0033688216684280

Storch, N., & Wigglesworth, G. (2010). Learners’ processing, uptake, and retention of corrective feedback on writing: Case studies. *Studies in Second Language Acquisition, 32*(2), 303–334. https://doi.org/10.1017/S0272263109990332

Tardy, C. M. (2019). Appropriation, ownership, and agency: Negotiating teacher feedback in academic settings. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 64–82). Cambridge University Press.

Terry, G., Hayfield, N., Clarke, V., & Braun, V. (2017). Thematic analysis. In C. Willig (Ed.), *The SAGE handbook of qualitative research in psychology* (pp. 17–36). SAGE Publications Ltd.

Tian, L., & Zhou, Y. (2020). Learner engagement with automated feedback, peer feedback and teacher feedback in an online EFL writing context. *System, 91*, 102247. https://doi.org/10.1016/j.system.2020.102247

Treglia, M. O. (2008). Feedback on feedback: Exploring student responses to teachers’ written commentary. *The Journal of Basic Writing, 27*(1), 105–137. https://wac.colostate.edu/docs/jbw/v27n1/treglia.pdf

Uscinski, I. (2017). L2 learners’ engagement with direct written corrective feedback in first-year composition courses. *Journal of Response to Writing, 3*(2), 36–62. http://journalrw.org/index.php/jrw/article/view/68

van Beunnigen, C., de Jong, N. H., & Kuiken, F. (2012). Evidence on the effectiveness of comprehensive error correction in second language writing. *Language Learning, 62*(1), 1–41. https://doi.org/10.1111/j.1467-9922.2011.00674

Walker, M. (2009). An investigation into written comments on assignments: Do students find them usable? *Assessment & Evaluation in Higher Education, 34*(1), 67–78. https://doi.org/10.1080/02602930801895752

Weaver, M. R. (2006). Do students value feedback? Student perceptions of tutors’ written responses. *Assessment & Evaluation in Higher Education, 31*(3), 379–394. https://doi.org/10.1080/02602930500353061

Winke, P., & Lim, H. (2014). The effects of testwiseness and test-taking anxiety on L2 Listening test performance: A visual (eye-tracking) and attentional investigation (IELTS Research Report Online Series, No. 3). British Council, Cambridge English Language Assessment and IDP: IELTS Australia. https://www.ielts.org/-/media/research-reports/ielts_online_rr_2014-3.ashx

Yang, Y., & Badger, R. (2015). How IELTS preparation courses support students: IELTS and academic socialisation. *Journal of Further and Higher Education, 39*(4), 438–465. https://doi.org/10.1080/0309877X.2014.953463
Yu, S., & Jiang, L. (2020). Doctoral students’ engagement with journal reviewers’ feedback on academic writing. *Studies in Continuing Education*, 1–18. https://doi.org/10.1080/0158037X.2020.1781610

Yu, S., Zhang, Y., Zheng, Y., Yuan, K., & Zhang, L. (2018). Understanding student engagement with peer feedback on master’s theses: A Macau study. *Assessment & Evaluation in Higher Education*, 44(1), 50–65. https://doi.org/10.1080/02602938.2018.1467879

Zacharias, N. T. (2007). Teacher and student attitudes toward teacher feedback. *RELC Journal*, 38(1), 38–52. https://doi.org/10.1177/0033688206076157

Zhang, Z. (Victor). (2020). Engaging with automated writing evaluation (AWE) feedback on L2 writing: Student perceptions and revisions. *Assessing Writing*, 43, 100439. https://doi.org/10.1016/j.asw.2019.100439

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

Zheng, Y., & Yu, S. (2018). Student engagement with teacher written corrective feedback in EFL writing: A case study of Chinese lower-proficiency students. *Assessing Writing*, 37, 13–24. https://doi.org/10.1016/j.asw.2018.03.001

Zheng, Y., Zhong, Q., Yu, S., & Li, X. (2020). Examining students’ responses to teacher translation feedback: Insights from the perspective of student engagement. *SAGE Open*, 10(2), 1–10. https://doi.org/10.1177/2158244020932536