

RESEARCH

Open Access



Response to written commentary in preparation for high-stakes second language writing assessment

William S. Pearson*

*Correspondence:
wsp202@exeter.ac.uk

University of Exeter, St Luke's
Campus, Heavitree Road,
Exeter, UK

Abstract

Many L2 learners preparing for high-stakes, on-demand English language tests (e.g., IELTS, TOEFL) undertake classroom-based test preparation involving the provision of teacher written feedback commentary (WFC) on writing that simulates test tasks. The assumption is teachers' knowledge of both the language and testing system helps develop candidates' language/test-taking skills and familiarity with task expectations. Prior research has indicated features of WFC's content and delivery can impact on the extent and quality of student revisions, although preparation for writing assessment settings have yet to be explored. The present study investigated the effects of five WFC content and delivery characteristics (focus, length, explicitness, semantic function, and presence of mitigation) on three rehearsal essays written by eight candidates preparing for IELTS Writing Task 2. The qualities of content and delivery most associated with substantive, positive revisions included comments targeting Task Response, those 50 words or longer, when an explicit revision strategy was provided, the presence of mitigation through personal attribution, and question posing and criticism. The study found learners tended not to act upon descriptive end comments explicating written performance, praise, and comments below five words in length. The implications for teachers in classroom IELTS preparation contexts are discussed.

Keywords: Written feedback, Written teacher commentary, Preparation for L2 writing assessment, IELTS

Introduction

The impact of teacher written commentary on L2 writing

Written feedback commentary (WFC) has long been considered an important and meaningful area of English as a second language teachers' work (Ferris, 1995; Lee, 2009; Sugita, 2006). High quality written commentary, i.e., that is detailed, usable, and considers a learner's affective response (Dawson et al., 2019), identifies what revisions might strengthen her/his writing (Goldstein, 2004), helping developing writers perceive and reflect on how others read their work. While research has consistently shown L2 learners value teachers' written commentary (Elwood & Bode, 2014; Hyland & Hyland, 2006; Lee, 2008; Tang & Liu, 2018) and usually seek to engage constructively with it (Sugita,

2006; Uscinski, 2017; Zhang & Hyland, 2018), language or writing development across revised compositions can be unpredictable or underwhelming (Alsharif & Alyousef, 2017; Christiansen & Bloch, 2016; Conrad & Goldstein, 1999; Nurmukhamedov & Kim, 2009; Uscinski, 2017).

One reason for the often-disappointing learning potential of written commentary relates to features of its content and delivery (Goldstein, 2006). Studies have generated insights into the differential impacts of various language features of EFL teachers' comments, notably semantic function and tone (e.g., advisory, criticism, praise) (Conrad & Goldstein, 1999; Ferris, 1997; Gedamu & Gezahegn, 2021; Neupane Bastola, 2021), syntactic structure (e.g., declarative, interrogative, imperative) (Nurmukhamedov & Kim, 2009; Sugita, 2006), length in words (Ferris, 1997; Grouling, 2018), provision of explicit revision strategies (Conrad & Goldstein, 1999; Ene & Upton, 2014; Lee et al., 2018), and presence of mitigation strategies (e.g., hedging, personal attribution) (Ferris, 1997; Hyland & Hyland, 2001, 2019; Nurmukhamedov & Kim, 2009; Treglia, 2008). Some research (e.g., Alsharif & Alyousef, 2017; Christiansen & Bloch, 2016; Conrad & Goldstein, 1999; Ferris, 1997) frames impact as the strategies learners adopt to revise their texts and the outcomes of such revisions. Others (Cunningham, 2019; Gedamu & Gezahegn, 2021; Treglia, 2008; Zacharias, 2007) query students' perspectives reported in interviewing after responding to written feedback.

Descriptive or experimental research into the impact of various approaches to WFC content and delivery reveals comments that suggest or necessitate content changes offer significant value (Conrad & Goldstein, 1999; Ferris, 1997; Nurmukhamedov & Kim, 2009) because such comments constitute feedback's core function of contributing actionable feed forward information, helping learners close performance gaps (Price et al., 2010). This is accompanied by caveats that advisory comments should be clear and comprehensible (Zacharias, 2007), detailed and explanatory (Elwood & Bode, 2014), 'guiding' rather than 'telling' (Treglia, 2008), encouraging and motivational (Tang & Liu, 2018), and lacking in a terse or exasperated tone (Hyland & Hyland, 2001). Less consistency is apparent in L2 learners' responses towards comments across syntactic structures (Conrad & Goldstein, 1999; Ferris, 1997), the use of praise/criticism (Ferris, 1995; Hyland & Hyland, 2001; Treglia, 2008), and the potentially facilitating or controlling effects of explicit revision strategy provision (Ene & Upton, 2014). This is likely because of the interaction between various student (e.g., beliefs, feedback literacy, language proficiency) and contextual variables (L2 learning context, written tasks) (Ellis, 2010; Goldstein, 2006). Research has mostly addressed the impact of WFC in tertiary-level process writing contexts (Conrad & Goldstein, 1999; Cunningham, 2019; Ferris, 1997; Nurmukhamedov & Kim, 2009; Sugita, 2006). Learners in such settings can be considered successful to a degree, often possessing opportunities to act on teacher WFC and familiarity with various feedback and response strategies.

Teacher commentary in preparation for writing assessment settings

One learning-to-write context in which the impact of teacher commentary on L2 student writing has yet to be explored is preparation for high-stakes English language writing assessment, for example IELTS (International English Language Testing System) and TOEFL (Test of English as a Foreign Language). Such assessments exhibit

several consistent features to the context of writing. Strictly controlled test conditions require candidates to perform under a time pressure without the use of external sources of input. Topics are expected to be familiar to test-takers, while at the same time not favouring the subject matter expertise of particular disciplines (IELTS, 2019a). Task prompts, although not known in advance, feature consistent rhetorical specifications (Coffin, 2004; Liu & Stapleton, 2015), meaning the discursual purpose required in written response follows a predictable pattern (e.g., persuading the reader that something is the case). Owing to these demands, many candidates undertake test preparation at a language teaching organisation led by a teacher knowledgeable of the testing system (Alderson & Hamp-Lyons, 1996; He, 2010; Hu & Trenkic, 2019; Saif et al., 2021; Yu et al., 2017). It is not uncommon for such learners to be unsuccessful test veterans (Alsagoafi, 2018; Barkaoui, 2017; Hamid, 2016), who may believe that outside feedback is key to unlocking the necessary gains in test performance (Pearson, 2018a).

Classroom-based preparation for L2 writing assessment is often orientated around modelling genre features of texts (Hamp-Lyons, 1998; Yang & Badger, 2015). Coaching in some contexts (e.g., He, 2010) may focus heavily on a very narrow range of rhetorical functions (Liu & Stapleton, 2015) through ‘teaching to the test’. This reflects a perspective held by some candidates that such tests constitute obstacles to be overcome (Liu & Stapleton, 2015; Sinclair et al., 2019), rather than opportunities for meaningful language/skill development. Typically, textual modelling is followed by learner rehearsal of parallel or retired tasks written in simulated test conditions (Allen, 2016; Hu & Trenkic, 2019; Mickan & Motteram, 2009; Yang & Badger, 2015). Such tasks provide useful opportunities for WFC, particularly in how successfully learners address the task and their use of language. Yet research has seldom addressed the role of written feedback in preparation for high-stakes L2 writing assessments.

One such assessment is IELTS Writing, a high-stakes test used to screen applicants’ written English language proficiency for mostly academic enrolment purposes. Task 2 constitutes a direct test of writing, requiring candidates to establish, “a position [on an impromptu issue] which is then defended through the use of evidence, negotiation, logic,... persuading the reader (either a specific reader or the world generally) to adopt the writer’s position and (frequently) carry out an action” (Coffin, 2004, p. 231). Task 2 is assessed through a series of abstract judgements of a candidate’s general communicative ability (Davies, 2008), partially available to candidates in the *Task 2 public band descriptors* (IELTS, 2019b). The descriptors synthesise a multitude of textual features into a limited number of open-ended observations (e.g., ‘an adequate range of vocabulary’, ‘conclusions may become unclear or repetitive’) across four criterion-referenced band descriptors (Task Response [TR], Coherence and Cohesion [CC], Lexical Resource [LR], and Grammatical Range and Accuracy [GRA]).

Student response to teacher written feedback commentary in IELTS test preparation settings may differ noticeably vis-à-vis tertiary-level process writing classrooms. By virtue of not having yet obtained desired test outcomes, developing writers may be more likely to embrace teacher comments suggesting or ordering action, incorporate information in the form of explicit revision strategies and appropriations, and be less concerned with receiving praise. Owing to prominent product writing foundations of IELTS (Zareekbatani, 2015), learners may lack familiarity with responding

behaviourally to WFC through undertaking revisions. To address such uncertainties, the present study investigates the impact of teacher written commentary on the rehearsal writing of eight candidates preparing for IELTS. Guiding the study are the following three research questions:

1. What are the characteristics of written feedback commentary on IELTS Task 2 rehearsal essays, written in preparation for the test?
2. How successful are developing writers in addressing written commentary, measured as the extent of revisions and impact on textual quality?
3. What characteristics of commentary appear to influence student revision?

Method

This study repurposes data from a broader inquiry into student engagement with written feedback in preparation for high-stakes L2 writing assessment. Eight candidates preparing for IELTS completed a bespoke learning-to-write project that featured sequentially writing two drafts of three Task 2 rehearsal essays, with form- and content-focused commentary provided by the researcher to help them reach their band score goals. Text-analytic descriptions of written commentary and student revisions generate a quantitative picture of feedback and response (Ferris, 2012) with respect to five features of commentary (focus, length, explicitness, pragmatic intent, and provision of mitigation) drawing upon existing schema in the literature (Christiansen & Bloch, 2016; Conrad & Goldstein, 1999; Ene & Upton, 2014; Ferris, 1997; Ferris et al., 1997; Hyland & Hyland, 2019; Lee et al., 2018; Nurmukhamedov & Kim, 2009).

The participants

Eight individuals preparing independently for IELTS were recruited in response to an advert placed on the public wall of an IELTS-orientated Facebook group advertising project participation in exchange for feedback. The participants originated from a range of countries, including three from India and one each from Algeria, Indonesia, Korea, Russia, and Sri Lanka. Five were males, three were females. All were young adults in their twenties. The participants were undertaking the test for the purposes of academic study in an Anglophone tertiary institution (five), permanent emigration (two), or professional registration (one). Reflecting these divergent purposes and the diversity in test-user requirements, the participants needed scores of between 6.0 and 7.5, although most required 6.5 or 7.0. No placement test was undertaken prior to admitting participants onto the study. Instead, an impressionistic judgement of suitability was made based on individuals' stated band score goals, any disclosed prior test scores, and the general quality of initial spoken and written interactions in English. It transpired four of the participants were two-time test veterans who had not yet achieved their target scores. Approval to undertake the study was granted from the ethics committee of the researcher's institution. All participants provided their written consent to participate before any data was collected.

The learning programme

The learning programme centred on the participants sequentially writing three Task 2 rehearsal essays using prompts selected by the researcher. Students were instructed to write first drafts in simulated test conditions (e.g., within 40 min, no recourse to dictionaries, spelling and grammar tools, test guidance, or other sources of input), which were submitted to the researcher by email for written feedback in Microsoft Word. Written feedback comprised indirect error treatment using a metalinguistic code (e.g., word choice, verb tense) based on Han and Hyland (2015), and commentary. Error corrections not included in a comment (i.e., merely the metalinguistic codes) were not analysed in this study. First draft commentary targeted textual features where there was a deficit between learners' written performance and their stated band score goals, with reference to the public band descriptors. All feedback was transferred from Word to an unedited version of the document hosted in Kaizena, an application that allows users in a virtual classroom space to comment on a shared document. The participants were requested to consider the written feedback and act on it in a second draft in non-simulated conditions, offering a lower-stakes opportunity to reach their band score goals. The learners submitted their second drafts for summative feedback, which was not analysed in this study. The participants undertook the project in a 'closed' Kaizena classroom with just the researcher present and chose when and how quickly they wished to proceed through the essays. Upon completion of the learning programme, all written feedback was imported from Word into Excel for analysis.

Data analysis

Characteristics of teacher commentary

This study features text-analytic description to generate a quantitative picture of written feedback commentary and student response (Ferris, 2012). Initially, coding centred on delineating the written feedback into discrete comments. A comment was defined semantically as, "a stretch of discourse having a unified intended function" (Conrad & Goldstein, 1999, p. 153). Comments were either contained within a single sentence or cut across multiple sentences (see Additional file 1). In the case of multi-sentential comments, a judgement was made whether the discourse consisted of single or multiple idea units (Ferris et al., 1997), the latter of which were treated as discrete comments.

At the time of research, there existed no comprehensive framework elaborating the content and delivery characteristics of written feedback commentary on L2 writing. Consequently, and to aid comparison between the study's findings and prior literature, five previously researched variables were incorporated into the design. The first was the focal area of written feedback (Alsharif & Alyousef, 2017; Christiansen & Bloch, 2016; Ene & Upton, 2014; Grouling, 2018; Lee et al., 2018), using values unique to the context of writing, i.e., the IELTS Task 2 assessment criteria. Next, the length of comments (in words) was automatically calculated using an Excel formula, with the resulting values categorised according to Ferris' (1997) scheme. Third, comments featuring an explicit revision directive were differentiated from (implicit) WF that did not suggest a particular strategy (Conrad & Goldstein, 1999; Ene & Upton, 2014; Lee et al., 2018). Comments that did not indicate a revision was required were counted separately. Fourthly, the semantic (or pragmatic) function of comments

was identified. While there exist several overlapping schema for coding the function of comments (Conrad & Goldstein, 1999; Ferris, 1997; Gedamu & Gezahegn, 2021; Grouling, 2018; Neupane Bastola, 2021; Treglia, 2008), Grouling's (2018) seven modes were adopted since they seemed to best suit the writing context. One change was made to the selection of code labels, with 'neutral' being replaced with 'descriptive' to better reflect the function of such comments. Finally, instances of feedback mitigation were coded according to Hyland and Hyland's (2001, 2019) four techniques. The full analytical model is described and illustrated in Table 1.

To provide more refined insights into the characteristics and impact of feedback commentary, a further distinction was made between *marginal* and *end comments*. In practice, comments targeting a highlighted feature of the text in a comment bubble in Word were coded as marginal, whereas summary prose written after the essay were deemed end comments. Additionally, comments were categorised into *overall*, *content*, and *form* according to Ene and Upton (2014), since marginal form comments required discrete categories of revision operations (see below). Comment focus, length, and explicitness were coded categorically, i.e., using one value only. In contrast, some (especially multi-sentential) comments contained more than one approach to mitigation and/or featured multiple structures of varying pragmatic intent, all of which were coded. All coding and analysis was undertaken by the researcher, who at the time possessed a Master's degree in Applied Linguistics, ten years' experience practising TESOL, and professional experience assessing Writing Task 2 essays.

Student revision operations

A subjective rating scale was developed to assess the impact of written feedback commentary on students' second draft essays, inspired by existing literature (Christiansen & Bloch, 2016; Conrad & Goldstein, 1999; Ferris, 1997; Ferris et al., 1997; Nurmukhamedov & Kim, 2009; Sugita, 2006). 286 comments addressing content and 128 end comments targeting form were assessed through a dual measure of the extent and effect of revisions (see Fig. 1). The extent of textual changes were labelled as *substantive*, *minimal*, or *no change* (Ferris, 1997), while effects were categorised as *positive*, *mixed*, or *negative* (Ferris, 1997; Sugita, 2006). Only revisions that worsened the text were classified as negative. If textual quality stayed roughly the same or featured strengths and weaknesses, the outcome was considered mixed. Textual improvements were not linked directly to changes in IELTS band scores since these are not sensitive to short-term improvements in writing proficiency (Rao et al., 2003). For the 183 marginal form comments, a judgement of whether the revision was *target-like*, *non-target-like*, *omitted*, or *no change* (Nurmukhamedov & Kim, 2009) was deemed more suitable. Examples illustrating each revision category are outlined in the Additional file 1.

The commentary characteristic and student response schema were applied on a range of practice texts before each authentic comment was defined and analysed by the researcher. A second pass of the whole data was carried out after two months to improve the accuracy and consistency of coding. The findings are presented as both raw frequency counts and proportions of comment characteristics and revision operations (Ferris, 1997; Ferris et al., 1997; Nurmukhamedov & Kim, 2009; Sugita, 2006). Twenty-one overall comments and a further 11 end comments that mostly explained how to

Table 1 Analytic model for written commentary

Category	Value	Example comment
Focus	Task response	<i>This seems to be very similar to the previous point you made in the previous paragraphs</i>
	Coherence and cohesion	<i>Just one thing, the pronoun reference word 'it' is a bit unclear in main body paragraph 2 and 'them' in the conclusion</i>
	Lexical resource	<i>Try to change the unusual general vocabulary that I've highlighted, which sound unnatural</i>
	Grammatical range and accuracy	<i>I was impressed with the range and quality of structures</i>
	Other	<i>This is a very good essay, good job!</i>
Length	Short (1–5 words)	
	Average (6–19)	
	Long (20–49)	
	Very long (50+)	
Explicitness	Explicit	<i>Perhaps add 'people from different countries or cultures' to be more precise</i>
	Implicit	<i>I feel this point is rather similar to your previous main argument</i>
	Revisions not required	<i>Body paragraph 1 has a good concluding sentence</i>
Semantic function	Advisory	<i>I'd use cautious language here to avoid overgeneralising: 'that can result'</i>
	Criticism	<i>I don't understand what you mean here</i>
	Descriptive	<i>The points you make are relevant, extended, and supported</i>
	Imperative	<i>Combine into one word</i>
	Praise	<i>Another good topic-specific item</i>
	Question posing	<i>Could you give an example of this?</i>
	Reflective	<i>Wow. I had no idea the complex was so large. But I'm not surprised, there are so many amazing sights there</i>
Provision of mitigation strategy	Hedges	<i>Perhaps say 'need to'</i>
	Interrogative syntax	<i>Is there a more academic sounding word?</i>
	Paired act/positive softener	<i>This is OK, but is there a more academic sounding word?</i>
	Personal attribution	<i>I think 'Take for example the country Spain, where...' sounds more appropriate as you're writing only about Spain</i>
	None	<i>Replace with 'the personality of the actor'</i>

undertake the revision process or provided information about the assessment criteria were not included in the revision analysis since it was not clear a measurable student response was possible.

Results and discussion

Commentary characteristics

There were 618 discrete first draft comments provided across the three rounds of Task 2 rehearsal essay writing. This works out as an average of approximately 25 comments per essay, indicating the feedback was extensive (Pearson, 2018b; Yang & Badger, 2015). Table 2 shows the frequencies and proportions of the various characteristics of written

	Extent of the revision	
Effect of the revision	No discernible change in response to the comment	
	Minimal change to address the comment, effect generally negative	Substantive change to address the comment, effect generally negative
	Minimal change to address the comment, effect mixed	Substantive change to address the comment, effect mixed
	Minimal change to address the comment, effect generally positive	Substantive change to address the comment, effect generally positive

Fig. 1 Rating scale for content and form end comment revisions, adapted from Ferris (1997)

feedback commentary. The majority encompassed marginal comments targeting issues in the text body ($n=321$), although as can be seen, their distribution across the four assessment criteria varied. Comments addressing how learners responded to the task, including the clarity, development, and relevance of the ideas presented constituted 38.6% of all marginal comments and just over a quarter of end comments. As in other test preparation activities, this reflects the role of written feedback in helping test-takers improve their awareness of and cope with the demands of the testing system (Brown, 1998; Hayes & Read, 2008; Mickan & Motteram, 2008; Saif et al., 2021; Yang & Badger, 2015). Language was not ignored (Yang & Badger, 2015), with 31.2% of all marginal comments addressing Lexical Resource, often the appropriacy and naturalness of word forms, word choices, or collocations. Coherence and Cohesion was rarely brought up in marginal commentary, reflecting the relative inaccessibility of such features compared to lexicogrammar and TR (Cotton & Wilson, 2011; Riazi & Knox, 2013).

There were a number of consistent characteristics to the content and delivery of marginal comments. They most frequently featured an advisory pragmatic function (39.7%), often explicitly outlining a strategy the learner could adopt to improve an aspect of their essay (61.1%), usually in regard to Task Response or Lexical Resource. They were typically short, with 67% being below 20 words, though hedging (the most frequent mitigation strategy and one which generally increases the lengths of comments) occurred in 24.3% of marginal comments. The 33% of marginal comments that were 20 words or longer usually affixed criticism or praise to an advisory statement, with the former serving to problematise the issue and the latter to soften the blow (Hyland & Hyland, 2001). A less common function overall, although still notable, were critical comments (20.9%), often co-occurring with implicit feedback, requiring the learner to work out how to address the issue.

In contrast, end comments ($n=297$) were more evenly distributed across the assessment criteria, reflecting their contrasting role in Task 2 rehearsal essay feedback. The pragmatic function of describing learner performance vis-à-vis the public band descriptors featured more prominently (22.3%), as did praise (24.2%), coded in comments that reiterated or explained the key messages of marginal comments, often in unmitigated statements of 20 words or more (70%). End comments did not always require revisions (40.1%), perhaps because they consisted of a general remark on performance (6.4%) or constituted praise. Those that did mostly conveyed an advisory sentiment implicitly (36%), as such comments often addressed students' texts from a global perspective, making it difficult to provide specific strategies. End comments did not forcefully instruct learners to make changes through the use of the imperative, unlike a small number of marginal comments (5.4%).

Table 2 Characteristics of teacher commentary

Category	Value	Marginal content comments		End comments	
		n	%	n	%
Focus	Task response	124	38.6%	77	25.9%
	Coherence and cohesion	34	10.6%	74	24.9%
	Lexical resource	100	31.2%	69	23.2%
	Grammatical range and accuracy	63	19.6%	58	19.5%
	Overall	0	0.0%	19	6.4%
Length	Short (1–5 words)	61	19.0%	4	1.3%
	Average (6–19)	154	48.0%	85	28.6%
	Long (20–49)	78	24.3%	161	54.2%
	Very long (50+)	28	8.7%	47	15.8%
Explicitness	Explicit	196	61.1%	71	23.9%
	Implicit	84	26.2%	107	36.0%
	Revisions not required	41	12.8%	119	40.1%
Semantic function	Advisory	163	39.7%	107	30.1%
	Criticism	86	20.9%	78	22.0%
	Descriptive	28	6.8%	79	22.3%
	Imperative	22	5.4%	0	0.0%
	Praise	46	11.2%	86	24.2%
	Question posing	61	14.8%	3	0.8%
	Reflective	5	1.2%	2	0.6%
Provision of mitigation strategy	Hedges	87	24.3%	62	20.3%
	Interrogative syntax	54	15.1%	5	1.6%
	Paired act/positive softener	25	7.0%	11	3.6%
	Personal attribution	29	8.1%	12	3.9%
	None	163	45.5%	215	70.5%

The propensity of advisory and critical comments reflects the highly evaluative nature of WFC on IELTS Task 2 rehearsal essays (Pearson, 2018b), where the teacher judges the correctness of the work and justifies the marks given (Weaver, 2006), in an effort to improve future test outcomes from a deficit perspective. The teacher, with superior knowledge of both the language and testing system, is legitimised as the ultimate authority on the test (Saif et al., 2021). Such an imbalance is visible in the prevalence of explicit commentary, often appropriations encouraging the writer to shift her/his position by injecting the teacher’s own meaning into the students’ words through reformulations or suggested topic ideas/development (Goldstein, 2004; Tardy, 2019). Clearly, there is a propensity for critical WFC to constitute a threat to students’ self-concept (Hyland & Hyland, 2001), which might explain the presence of extensive mitigation. However, the pressure to achieve goals may help ‘immunise’ some students against the potential harm of critical WFC (Han & Hyland, 2019).

Ratings of student revisions

The outcomes of student revisions in response to actionable content and end form-focused comments, measured as both the extent of the revision and the effect on textual quality are presented in Table 3. It can be seen that notable proportions of marginal

(26.1%) and end comments (41.1%) were not acted upon by learners, possibly indicating a lack of engagement (Han & Hyland, 2015). These figures are higher than studies of teacher WFC in tertiary-level process writing environments (Christiansen & Bloch, 2016; Conrad & Goldstein, 1999; Ene & Upton, 2014; Hyland, 2003; Nurmukhamedov & Kim, 2009; Ranalli, 2021; Zhang & Hyland, 2018), where comments across multiple drafts serve as scaffolding to help students develop their texts with an initial focus on content and organisation and later, grammar and mechanics. Two studies uncovered higher rates of no response (Ferris, 1997; Sugita, 2006), although Ferris (1997) acknowledges the frequency of praise reduced students’ agency to revise. Additionally, the participants of this study were generally more disposed towards undertaking minimal textual changes (41.2%) than they were substantive ones (22.8%), mirroring some studies in non-writing for assessment purposes contexts (Nurmukhamedov & Kim, 2009; Sugita, 2006). Non or perfunctory resolutions were notable responses to end comments, of which only 17.8% were met with a substantive revision.

Students’ unwillingness or inability to revise can be considered surprising given that they rarely met their band score goals across first drafts and due to the high amount of explicit, advisory WFC that outlined ways forward. However, significant proportions of (especially end) comments constituted praise or described aspects of written performance, implying no revision was necessary (Ferris, 1997; Hyland & Hyland, 2001). Non-simulated writing may not have been considered reflective of authentic written outcomes, resulting in participants not perceiving a valid purpose in response (Zareekbatani, 2015; Zheng et al., 2020). It could also be the case that, as in other writing settings, there was disagreement with the commentary (Goldstein & Kohls, 2002; Pratt, 1999), stemming from a lack of trust in the credibility of the feedback provider (Ranalli, 2021), who was not initially known to the participants. This is not an insignificant issue since purported experts on IELTS preparation abound on social networking groups (Pearson, 2018a), along with much ‘folk knowledge’ passed off as test-taking gospel (Allen, 2016). Alternatively, it was possible the comprehensive WFC posed difficulties for the learners to respond to all messages, while frequent mitigation may have lessened the impetus to revise by diluting the importance in which a textual issue was framed (Hyland & Hyland, 2001).

Rates of successful response to WFC appeared lower than several prior studies (Conrad & Goldstein, 1999; Ferris, 1997; Nurmukhamedov & Kim, 2009; Sugita, 2006),

Table 3 Ratings of content and form-focused end comment revisions

Rating	Marginal content comments		End comments	
	<i>n</i>	%	<i>n</i>	%
No change	36	26.1%	109	41.1%
Minimal change / negative effect	3	2.2%	2	0.8%
Substantive change / negative effect	0	0.0%	0	0.0%
Minimal change / mixed effect	24	17.4%	64	24.2%
Substantive change / mixed effect	19	13.8%	14	5.3%
Minimal change / positive effect	30	21.7%	43	16.2%
Substantive change / positive effect	26	18.8%	33	12.5%
Total	138	100.0%	265	100.0%

painting an unclear picture of feedback effectiveness in this context. Marginal comments about content were effective at inducing a positive effect on students' texts 40.5% of the time, although many successful revisions were minimal in scope (21.7%). In terms of raw frequency, end comments resulted in more numerous instances of enhanced textual quality ($n=76$), though a greater percentage induced a mixed impact as opposed to definitively improving it. It could be that certain key messages became diluted in the lengthy end comment descriptions, deeply coded using the conventions of language assessment specialists (Weaver, 2006). Alternatively, the comprehensiveness of the information may have proved overwhelming and unmanageable (Evans et al., 2010; Lee, 2019), especially for the weaker learners with ambitious IELTS band score targets.

Greater success was exhibited by the learners in addressing marginal form comments. Target-like revisions occurred at a rate of 51.4%, although many comments directly treated student errors or contained explicit reformulations. 24% of marginal form comments resulted in deletion of the problematic feature, suggesting student avoidance of the issue (Han & Hyland, 2015), while a somewhat concerning 14.2% remained in draft two. However, as found elsewhere (Ferris, 1997; Nurmukhamedov & Kim, 2009; Sugita, 2006), occurrences of content revisions that worsened the text were rare ($n=5$), explained by the advisory, appropriating nature of the WFC and that the students appeared reluctant to take risks in response to comments. Likewise, just 10.4% of form revisions lead to non-target-like outcomes.

The influence of comment characteristics on student revisions

Table 4 shows the influence of the five categories of commentary characteristics investigated in the present study, with negative effects on textual quality excluded owing to the infrequency of such occurrences. First, with regard to textual focus, the greatest frequency of substantive revisions was brought about by comments targeting Task Response (34.6%). This is not surprising since such comments typically encouraged learners to improve the clarity, support, or extension of main or supporting ideas, requiring substantial changes. With the notable rate of 21.8% substantive, positive changes, TR constituted the criterion most likely to bring about tangible textual improvements through WFC. However, it was also the case that TR comments were frequently ignored (30.3%) or resulted in minimal revisions that did not definitively improve the text (19.7%). As in other contexts (Christiansen & Bloch, 2016; Ferris, 1997; Sugita, 2006; Uscinski, 2017), students preparing for Task 2 both paid attention to written feedback that helps them make substantive, effective revisions, but also disregarded suggestions, highlighting the salience of addressing individual student factors in conjunction with content and delivery attributes of WFC (Conrad & Goldstein, 1999) in this context.

Substantive changes in a text's Coherence and Cohesion occurred at the much lower rate of 20.6%, with a 6.6% lower proportion of positive revision outcomes, suggesting learners struggled to address CC issues, mirroring the experiences of teachers (Cotton & Wilson, 2011; Riazi & Knox, 2013). While a significant proportion (55.7%) of comments addressing Lexical Resource (a significant focal of area of WFC), resulted in no change, LR constituted the criteria in which learners tended to perform closest to their targets, meaning many such comments praised the overall resource or specific

Table 4 Relationship between content and end form comment characteristics and revision ratings

Category	Value	No change		Minimal change/ mixed effect		Substantive change/ mixed effect		Minimal change/ positive effect		Substantive change/positive effect	
		n	%	n	%	n	%	n	%	n	%
Focus ^d	Task response	57	30.3%	37	19.7%	24	12.8%	29	15.4%	41	21.8%
	Coherence and cohesion	41	44.6%	17	18.5%	5	5.4%	15	16.3%	14	15.2%
	Lexical resource	39	55.7%	12	17.1%	1	1.4%	18	25.7%	0	0.0%
Length	Grammatical range and accuracy	8	16.7%	22	45.8%	3	6.3%	12	25.0%	3	6.3%
	Short (1–5 words)	16	76.2%	1	4.8%	0	0.0%	4	19.0%	0	0.0%
	Average (6–19)	53	47.7%	21	18.9%	4	3.6%	23	20.7%	10	9.0%
Explicitness	Long (20–49)	64	33.2%	47	24.4%	18	9.3%	37	19.2%	27	14.0%
	Very long (50 +)	12	16.4%	19	26.0%	11	15.1%	10	13.7%	21	28.8%
	Explicit	28	20.0%	36	25.7%	13	9.3%	31	22.1%	32	22.9%
Semantic function ^b	Implicit	24	16.6%	47	32.4%	20	13.8%	32	22.1%	22	15.2%
	Revisions not required	93	82.3%	5	4.4%	0	0.0%	11	9.7%	4	3.5%
	Advisory	34	20.9%	45	27.6%	17	10.4%	41	25.2%	26	16.0%
Provision of mitigation strategy	Criticism	18	14.2%	39	30.7%	15	11.8%	22	17.3%	33	26.0%
	Descriptive	30	37.0%	14	17.3%	6	7.4%	21	25.9%	10	12.3%
	Praise	76	74.5%	6	5.9%	5	4.9%	11	10.8%	4	3.9%
Other ^a	Question posing	10	23.8%	7	16.7%	7	16.7%	7	16.7%	11	26.2%
	Reflective	5	71.4%	0	0.0%	0	0.0%	1	14.3%	1	14.3%
	Hedges	29	25.2%	32	27.8%	11	9.6%	20	17.4%	23	20.0%
None	Interrogative syntax	11	27.5%	7	17.5%	9	22.5%	8	20.0%	5	12.5%
	Paired act/positive softener	7	25.0%	7	25.0%	5	17.9%	6	21.4%	3	10.7%
	Personal attribution	6	24.0%	1	4.0%	3	12.0%	8	32.0%	7	28.0%
		102	44.0%	48	20.7%	15	6.5%	38	16.4%	29	12.5%

^aOther^a was omitted since it was not considered possible to measure students' responses to such comments

^bThe imperative was omitted because of its low overall frequency

items used. In contrast, grammar-focused end comments were seldom addressed with substantive or positive revisions, although as such comments tended to be descriptive, they may have helped reinforce or explicate the messages contained in marginal form-focused comments and indirectly treated errors (Ferris, 1997).

In terms of length, it was found comments of 1–5 words offered little utility, particularly in facilitating successful revisions, which occurred only four times. This is because they often featured praise (Ferris, 1997) or were facile (Treglia, 2008; Walker, 2009; Weaver, 2006). Average comments also featured a low take up rate, (47.7%) though did contribute to small-scale improvements in textual quality (20.7%). In comparison, long comments seemed to offer more utility, with 14.5% fewer being ignored and 3.5% more positive outcomes. Importantly, substantive revisions with a positive effect were the most frequent outcome of very long comments, accompanied by low rates of no change (16.4%). This could be because longer comments tended to combine description of problematic textual features with advisory information to help the learners resolve the issue (Conrad & Goldstein, 1999) or because the amount of text dedicated to the issue conveyed its seriousness to the learner. Nevertheless, the high rate of mixed effects (41.1%) indicates learners experienced difficulties acting on detailed WFC, a phenomenon not unique to this context (Christiansen & Bloch, 2016; Conrad & Goldstein, 1999; Ferris, 1995).

Comparable rates of substantive revisions resulted from explicit (32.2%) and implicit WFC (29%). This could be because criticism, a key semantic function underlying implicit feedback (Ene & Upton, 2014), served to highlight something was wrong thereby triggering a substantive revision attempt. Nevertheless, the 7.7% higher frequency of substantive, positive revisions suggests feedback that explains and scaffolds what learners need to do to better meet their goals is more helpful at encouraging revisions than merely criticising the work (Treglia, 2008). Identical rates of marginal, positive responses show the inclusion of specific revision strategies did not always significantly affect the quality of subsequent revisions (Conrad & Goldstein, 1999), perhaps because learners lacked the assessment literacy to translate commentary deeply coded in the language of assessment into actionable strategies (Weaver, 2006). Alternatively, since there was a higher rate of explicit feedback not being acted upon (by 3.4%), learners may have disagreed with the information (Goldstein & Kohls, 2002; Pratt, 1999) as it did not align with their schema of what constituted an effective response or a workable approach in test conditions. Perhaps unsurprisingly, in 82.3% of cases, if WFC did not outline or imply a response to a problematic issue, no revision attempt was made on the highlighted issue.

Several salient patterns emerged in learners' responses to comments of varying semantic function and mitigation. The functions least able to induce a revision response were, unsurprisingly, praise (74.5% no change), mirroring the findings of Ferris (1997), and reader reflection (71.4%), albeit the latter was a far less frequent comment type. In contrast, criticism constituted a polarising pragmatic function, accounting for both a high proportion of substantive revisions with positive effects (26%), but also the most occurrences of marginal changes with mixed effects (30.7%). This is perhaps because learners lacked understandings of the framing of problematic textual issues (in relation to the band descriptors) (Conrad & Goldstein, 1999; Goldstein, 2004), did not agree the issues

were problematic (Goldstein & Kohls, 2002), or perceived a reduced self-concept stemming from repeatedly performing below their target (Estaji & Tajeddin, 2012). Interestingly, unmitigated comments were likely to be ignored (44%) or acted upon perfunctorily (20.7%), suggesting the participants appreciated the sting being taken out of face-threatening feedback (Hyland & Hyland, 2001; Treglia, 2008). Personal attribution exhibited the highest rates of substantive, positive changes, perhaps because test preparation candidates are known to highly value the input of outside experts (Allen, 2016; Mickan & Motteram, 2009), and thus perceived such messages as insider information.

Descriptive comments that characterised learners' texts did not act as a catalyst for extensive revisions, with 37% not being acted on and 25.9% resulting in marginal, positive effects. A likely explanation is that the absence or implicitness of a revision imperative combined with the generality of such comments made them difficult to act upon (Ferris, 1997). Interestingly, 26.2% of all questions posed led to substantive, positive textual changes, possibly because learners were encouraged to think more deeply about the identified issue and/or consult the assessment criteria/test preparation materials. It may not be the case that merely rephrasing WFC in the interrogative triggers such a response, as the equivalent outcomes of comments hedged using interrogative syntax were significantly lower (12.5%). Comparable rates of positive (41.2%) and mixed revision effects (38%) and the 9.2% lower rate of substantive, positive outcomes for the most common function, advisory, provides further evidence learners struggled to act on WFC requesting changes to their essays (Christiansen & Bloch, 2016; Conrad & Goldstein, 1999; Ferris, 1995), a phenomenon requiring additional exploration.

Conclusions

The study found that, as in other L2 learning-to-write settings (Christiansen & Bloch, 2016; Ferris, 1997; Uscinski, 2017), notable proportions of marginal content and end comments were either not acted upon or addressed minimally. In the present study, this behaviour was associated with end comments targeting Grammatical Range and Accuracy, praise, comments below 20 words in length, when no imperative to revise was present, and with unmitigated comments. Given that most participants underperformed vis-à-vis their band score targets, such approaches may have been resisted because they were not considered valid and appropriate to the student's point of view or purpose for writing (Straub, 1997). As such, practitioners in preparation for writing assessment settings may wish to avoid such techniques if a notable revision imperative exists. The poor response rate to unmitigated comments suggests feedback providers should be wary of the affective impacts of WFC (Dawson et al., 2019; Treglia, 2008), even in the high-stakes settings of preparing for IELTS, and not assume the need to achieve goals immunises learners against the harmful effects of critical feedback. Since praise did not seem to induce revisions (Ferris, 1997), practitioners may seek to adopt alternative strategies to mitigate the impact of comments that request revisions (Treglia, 2008). Of particular value might be the strategy of personal attribution, as learners in this context highly value the input of outside experts (Allen, 2016; Saif et al., 2021).

Comments targeting Task Response induced the most substantive, positive effects across the four assessment criteria, establishing this criterion as a more malleable feature of learner writing compared to LR and GRA, which may be better dealt with using

focused marginal form comments that directly treat errors to avoid overloading learners (Bitchener & Knoch, 2009). Textual improvements stemmed from longer, more explicit comments that posed questions and criticised problematic textual issues. Consequently, there appears merit in practitioners providing detailed, constructive, and thought provoking WFC that diagnoses and treats deficient textual features framed within the public band descriptors. Nevertheless, the noteworthy portion of explicit, advisory WF that was ignored suggests participants did not always find WFC that attempted to point the way forward understandable or usable (Dawson et al., 2019; Treglia, 2008). Since soliciting advice on the response to a task from a knowledgeable outsider is a key factor underlying learners' participation on an IELTS test preparation programme, teachers are advised to openly discuss students' expectations and preferences towards the explicitness of strategy provision, including appropriations and reformulations (Tardy, 2019) and tailor their approach accordingly. Recorded oral feedback or teacher-student conferences could improve the feasibility of extensive written feedback versus extended commentary (Moore & Wallace, 2012).

Even though the written commentary drew on both the Task 2 public band descriptors and feedback approaches reported in prior studies (Brown, 1998; Pearson, 2018b; Yang & Badger, 2015), the findings are limited by possible idiosyncrasies of the researcher's approach and the small sample of learners. Furthermore, the content and delivery of WFC (and its impact) was heavily influenced by contextual and learner factors (Ellis, 2010; Goldstein, 2004, 2006). The challenge to respond effectively was raised by the likelihood the participants were unfamiliar with or even reluctant to revise their essays (Zareekbatani, 2015) as well as the comprehensiveness and unfocusedness of the WFC, stemming from notable gaps to band score goals across multiple criteria. As such, the findings may tentatively transfer only to the segment of the IELTS candidature linguistically unready to achieve their goals, comprising test repeaters and learners who perceive test preparation as a shortcut to success (Alsagoafi, 2018; Barkaoui, 2017; Hamid, 2016; Hu & Trenkic, 2019; Sinclair et al., 2019). Future research that investigates a larger sample of students' responses to written feedback or explores the phenomenon from the perspective of the learner, e.g., through an approach encompassing 'talking around texts' (Ivanić & Satchwell, 2007) via semi-structured interviewing, could yield more complete and nuanced insights into the characteristics of WFC that help or hinder student response.

Abbreviations

CC: Coherence and Cohesion; GRA: Grammatical Range and Accuracy; IELTS: International English Language Testing System; L2: Second language; LR: Lexical Resource; TESOL: Teaching English to speakers of other languages; TOEFL: Test of English as a Foreign Language; TR: Task Response; WFC: Written feedback commentary.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40862-022-00145-6>.

Additional file 1. Supplementary data.

Acknowledgements

Not applicable.

Author contributions

All data collection, analysis, and writing up of the report were undertaken by the author. The author read and approved the final manuscript.

Funding

The author received no specific funding for this work.

Availability of data and materials

The dataset used and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations**Ethics approval and consent to participate**

Ethical approval to undertake the study was obtained from the author's institution (University of Exeter, reference number D1920-049). All participants provided their written consent to participate in the study before any data was collected.

Consent for publication

Not applicable.

Competing interests

The author declares that he/she has no competing interests.

Received: 14 December 2021 Accepted: 3 March 2022

Published online: 16 August 2022

References

- Alderson, J. C., & Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of washback. *Language Testing*, 13(3), 280–297. <https://doi.org/10.1177/026553229601300304>
- 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>
- Alsagoafi, 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>
- Alsharif, A. A., & Alyousef, H. S. (2017). Investigating ESL/EFL students' approaches in response to revision processes: A case study. *Journal of Language and Linguistic Studies*, 13(2), 153–181.
- Barkaoui, K. (2017). Examining repeaters' performance on second language proficiency tests: A review and a call for research. *Language Assessment Quarterly*, 14(4), 420–431. <https://doi.org/10.1080/15434303.2017.1347790>
- Bitchener, J., & Knoch, U. (2009). The value of a focused approach to written corrective feedback. *ELT Journal*, 63(3), 204–211. <https://doi.org/10.1093/elt/ccn043>
- Brown, J. D. H. (1998). *Does IELTS preparation work? An application of the context-adaptive model of language program evaluation* (IELTS Research Reports, Volume 1). IELTS Australia Pty Limited. https://www.ielts.org/-/media/research-reports/ielts_rr_volume01_report2.ashx
- 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.
- 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>
- 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)
- Cotton, F., & Wilson, K. (2011). *An investigation of examiner rating of coherence and cohesion in the IELTS Academic Writing Task 2* (IELTS Research Reports, Volume 12). IDP: IELTS Australia and British Council. https://www.ielts.org/-/media/research-reports/ielts_rr_volume12_report6.ashx
- Cunningham, J. M. (2019). Composition students' opinions of and attention to instructor feedback. *Journal of Response to Writing*, 5(1), 4–38.
- Davies, A. (2008). Assessing academic English language proficiency: 40+ years of U.K. language tests. In J. Fox, M. Wesche, D. Bayliss, L. Cheng, C. E. Turner, & C. Doe (Eds.), *Language testing reconsidered* (pp. 73–86). University of Ottawa Press.
- Dawson, P., Henderson, M., Mahoney, P., Phillips, M., Ryan, T., Boud, D., & Molloy, E. (2019). What makes for effective feedback: Staff and student perspectives. *Assessment and Evaluation in Higher Education*, 44(1), 25–36. <https://doi.org/10.1080/02602938.2018.1467877>
- 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>
- Ene, E., & Upton, T. A. (2014). Learner uptake of teacher electronic feedback in ESL composition. *System*, 46(1), 80–95. <https://doi.org/10.1016/j.system.2014.07.011>
- 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>
- Evans, N. W., Hartshorn, K. J., & Tuioti, E. A. (2010). Written corrective feedback: Practitioners' perspectives. *International Journal of English Studies*, 10(2), 47–77. <https://doi.org/10.6018/ijes/2010/2/119191>
- 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. (1997). The influence of teacher commentary on student revision. *TESOL Quarterly*, 31(2), 315–339. <https://doi.org/10.2307/3588049>
- 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., Pezone, S., Tade, C. R., & Tinti, S. (1997). Teacher commentary on student writing: Descriptions & implications. *Journal of Second Language Writing*, 6(2), 155–182. [https://doi.org/10.1016/S1060-3743\(97\)90032-1](https://doi.org/10.1016/S1060-3743(97)90032-1)
- Gedamu, A. D., & Gezahegn, T. H. (2021). EFL supervisors' written feedback focus and language functions: A mixed methods study. *Asian-Pacific Journal of Second and Foreign Language Education*. <https://doi.org/10.1186/s40862-021-00125-2>
- Goldstein, L. M. (2004). Questions and answers about teacher written commentary and student revision: Teachers and students working together. *Journal of Second Language Writing*, 13(1), 63–80. <https://doi.org/10.1016/j.jslw.2004.04.006>
- Goldstein, L. M. (2006). Feedback and revision in second language writing: Contextual, teacher, and student variables. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 185–205). Cambridge University Press.
- Goldstein, L. M., & Kohls, R. (2002). Writing, commenting and revising: The relationship between teacher feedback and student revision online. Paper Presented at the American Association of Applied Linguistics Conference.
- Grouling, J. (2018). The genre of teacher comments from hard copy to iPad. *Journal of Response to Writing*, 4(1), 70–99.
- 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/10.1080/01596306.2015.1061978>
- Hamp-Lyons, L. (1998). Ethical test preparation practice: The case of the TOEFL. *TESOL Quarterly*, 32(2), 329–337. <https://doi.org/10.2307/3587587>
- 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>
- Hayes, B., & Read, J. (2008). IELTS test preparation in New Zealand: Preparing students for the IELTS academic module. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 97–111). Lawrence Erlbaum Associates Publishers.
- He, L. (2010). The Graduate School Entrance English Examination. In L. Cheng & A. Curtis (Eds.), *English language assessment and the Chinese learner* (pp. 145–157). Routledge.
- 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. (2003). Focusing on form: Student engagement with teacher feedback. *System*, 31(2), 217–230. [https://doi.org/10.1016/S0346-251X\(03\)00021-6](https://doi.org/10.1016/S0346-251X(03)00021-6)
- 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](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>
- Hyland, K., & Hyland, F. (2019). Interpersonality and teacher-written feedback. *Feedback in second language writing: Contexts and issues* (pp. 165–183). Cambridge University Press.
- 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.ashx>
- IELTS. (2019b). *IELTS Task 2 Writing band descriptors (public version)*. https://takeielts.britishcouncil.org/sites/default/files/ielts_task_2_writing_band_descriptors.pdf
- Ivanič, 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>
- Lee, I. (2008). Understanding teachers' written feedback practices in Hong Kong secondary classrooms. *Journal of Second Language Writing*, 17(2), 69–85. <https://doi.org/10.1016/j.jslw.2007.10.001>
- Lee, I. (2009). Ten mismatches between teachers' beliefs and written feedback practice. *ELT Journal*, 63(1), 13–22. <https://doi.org/10.1093/elt/ccn010>
- Lee, I. (2019). Teacher written corrective feedback: Less is more. *Language Teaching*, 52(4), 524–536. <https://doi.org/10.1017/S0261444819000247>
- Lee, J. J., Vahabi, F., & Bikowski, D. (2018). Second language teachers' written response practices: An in-house inquiry and response. *Journal of Response to Writing*, 4(1), 34–69.
- Liu, F., & Stapleton, P. (2015). Writing prompt convergence in high-stakes tests: Exploring rhetorical functions and objects of enquiry. *Studies in Educational Evaluation*, 47, 68–75. <https://doi.org/10.1016/j.stueduc.2015.09.001>
- Mickan, P., & Motteram, J. (2008). *An ethnographic study of classroom instruction in an IELTS preparation program* (IELTS Research Reports, Volume 8). IELTS Australia Pty Limited. https://www.ielts.org/-/media/research-reports/ielts_rr_volume08_report1.ashx
- Mickan, P., & Motteram, J. (2009). *The preparation practices of IELTS candidates: Case studies* (IELTS Research Reports, Volume 10). IELTS Australia Pty Limited and British Council. https://www.ielts.org/-/media/research-reports/ielts_rr_volume10_report5.ashx
- Moore, C., & Wallace, I. P. H. (2012). Personalizing feedback for feed-forward opportunities utilizing audio feedback technologies for online students. *International Journal of E-Education, e-Business, e-Management and e-Learning*, 2(1), 6–10.
- Neupane Bastola, M. (2021). Formulation of feedback comments: Insights from supervisory feedback on master's theses. *Innovations in Education and Teaching International*, 58(5), 565–574. <https://doi.org/10.1080/14703297.2020.1804985>

- Nurmukhamedov, U., & Kim, S. H. (2009). "Would you perhaps consider...": Hedged comments in ESL writing. *ELT Journal*, 64(3), 272–282. <https://doi.org/10.1093/elt/ccp063>
- Pearson, W. S. (2018a). Utilising Facebook community groups for IELTS preparation: A thematic analysis. *The Asian EFL Journal*, 20(12.2), 5–59.
- Pearson, W. S. (2018b). Written corrective feedback in IELTS Writing Task 2: Teachers' priorities, practices, and beliefs. *TESL-EJ*, 21(4), 1–32.
- Pratt, E. E. (1999). *A qualitative study of peer and teacher response in an ESL writing classroom in Puerto Rico* [Indiana University of Pennsylvania]. <https://search.proquest.com/pqdtglobal/docview/304540663/fulltextPDF/B35D35B61EC64234PQ/1?accountid=10792>
- Price, M., Handley, K., Millar, J., & O'Donovan, B. (2010). Feedback: All that effort, but what is the effect? *Assessment and Evaluation in Higher Education*, 35(3), 277–289. <https://doi.org/10.1080/02602930903541007>
- Ranalli, J. (2021). L2 student engagement with automated feedback on writing: Potential for learning and issues of trust. *Journal of Second Language Writing*, 52, 100816. <https://doi.org/10.1016/j.jslw.2021.100816>
- Rao, C., McPherson, K., Chand, R., & Khan, V. (2003). *Assessing the impact of IELTS preparation programs on candidates' performance on the General Training reading and writing test modules* (IELTS Research Reports, Volume 5). IELTS Australia Pty Limited. https://www.ielts.org/-/media/research-reports/ielts_rr_volume05_report5.ashx
- Riazi, A. M., & Knox, J. S. (2013). *An investigation of the relations between test-takers' first language and the discourse of written performance on the IELTS Academic Writing Test, Task 2* (IELTS Research Report Series, No. 2). IDP: IELTS Australia. https://www.ielts.org/-/media/research-reports/ielts_online_rr_2013-2.ashx
- Saif, S., Ma, J., May, L., & Cheng, L. (2021). Complexity of test preparation across three contexts: Case studies from Australia, Iran and China. *Assessment in Education: Principles, Policy and Practice*, 28(1), 37–57. <https://doi.org/10.1080/0969594X.2019.1700211>
- Sinclair, J., Larson, E. J., & Rajendram, S. (2019). "Be a machine": International graduate students' narratives around high-stakes English tests. *Language Assessment Quarterly*, 16(2), 236–252. <https://doi.org/10.1080/15434303.2019.1628238>
- Straub, R. (1997). Students' reactions to teacher comments: An exploratory study. *Research in the Teaching of English*, 31(1), 91–119. <https://doi.org/10.2307/40171265>
- Sugita, Y. (2006). The impact of teachers' comment types on students' revision. *ELT Journal*, 60(1), 34–41. <https://doi.org/10.1093/elt/ccj079>
- Tang, C., & Liu, Y.-T. (2018). Effects of indirect coded corrective feedback with and without short affective teacher comments on L2 writing performance, learner uptake and motivation. *Assessing Writing*, 35, 26–40. <https://doi.org/10.1016/j.asw.2017.12.002>
- 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.
- Treglia, M. O. (2008). Feedback on feedback: Exploring student responses to teachers' written commentary. *The Journal of Basic Writing*, 27(1), 105–137.
- 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.
- Walker, M. (2009). An investigation into written comments on assignments: Do students find them usable? *Assessment and 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 and Evaluation in Higher Education*, 31(3), 379–394. <https://doi.org/10.1080/02602930500353061>
- 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, G., He, L., Rea-Dickins, P., Kiely, R., Lu, Y., Zhang, J., Zhang, Y., Xu, S., & Fang, L. (2017). Preparing for the speaking tasks of the TOEFL iBT® Test: An investigation of the journeys of Chinese test takers. *ETS Research Report Series*, 2017(1), 1–59. <https://doi.org/10.1002/ets2.12145>
- Zacharias, N. T. (2007). Teacher and student attitudes toward teacher feedback. *RELC Journal*, 38(1), 38–52. <https://doi.org/10.1177/0033688206076157>
- Zareekbatani, A. (2015). *Technology and L2 writing: EFL student perspectives on electronic feedback using online learning logs* [University of Exeter]. <https://ore.exeter.ac.uk/repository/handle/10871/17297>
- Zhang, Z., & 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., 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>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.