

Enhancing Writing Proficiency: The Role of Model Essays as Corrective Feedback Tools in IELTS Writing Task Achievement and Coherence/Cohesion

Mohamadreza Jafary¹, Seyedeh Fatemeh Amani² & Brian Benoit³

¹ Faculty of Education, McGill University, Montreal, Canada

² Faculty of Persian Literature and Foreign Language, Islamic Azad University, Roudehen Branch, Roudehen, Tehran, Iran

³ School of education, Bishop University, Canada

Correspondence: Mohamadreza Jafary, DISE, Faculty of Education, McGill University, Montreal, H3 A1Y2, Canada. E-mail: Mohamadreza.Jafary@mail.mcgill.ca

Received: September 16, 2023

Accepted: October 21, 2023

Online Published: October 27, 2023

doi: 10.5539/elt.v16n11p1

URL: <https://doi.org/10.5539/elt.v16n11p1>

Abstract

Proficiency in IELTS writing tasks is crucial for obtaining high scores in this internationally recognized test. This quasi-experimental study investigated the effectiveness of corrective feedback in improving the gain scores of both academic writing task 1 and task 2. Sixty Iranian students participated in pre/post-test administrations, with the experimental group receiving instruction based on the analysis made on 10 model essays, and the control group received reformulation on their own produced texts. Two skilled raters assessed the students' typewritten texts in terms of task response and "coherence/cohesion" – two writing band descriptors. After conducting ANOVA and Bonferroni post hoc tests, the results demonstrated that the treatment group achieved significantly higher scores in the two mentioned components. In both the post-test and delayed post-test, the Experimental Group (EG) consistently outperformed the Control Group (CG) in task response sub-scale of academic writing task 1 & 2 ($p < 0.05$). Further examination, employing a multi-mediator approach, accentuates the substantial impact of these writing rubrics on the overall IELTS score. Additionally, within the EG, there was a significant increase in coherence scores between the pretest and posttest ($p < 0.05$). However, no statistically significant change was observed between the post-test and the delayed post test ($p > 0.05$).

Keywords: writing proficiency, coherence/cohesion, IELTS, model essays, feedback

1. Introduction

Various approaches have been developed in the history of ESL writing, including the oral approach, current-traditional rhetoric, and process approach. However, criticisms arose, highlighting the approach's limitations in addressing individual differences, task variations, cognitive development, proficiency levels, and academic purposes. One approach that emerged to address these shortcomings was the English for Academic Purposes (EAP) approach. EAP aimed to socialize learners into academic contexts and enable them to succeed within academic communities. Regardless of the approach, the importance of corrective feedback in improving students' writing remains a common thread. For instance, while Ferris argues that "effective error correction, which is selective, prioritized, and clear," can benefit some students, Truscott disagrees. Feedback can take various forms, including direct and indirect, selective, and comprehensive. When assessing a skill, the evaluation criteria reflect the underlying theoretical framework. Holistic and analytic scoring are the primary methods for assessing essays, elucidating how raters react to writing tasks.

Corrective feedback in language learning and teaching is a topic of substantial interest, encompassing all four language skills. However, writing, as a productive skill, presents unique challenges for both teachers and students. The growing international migration to English-speaking countries for residence and education has increased the demand for high scores in international English exams. One such exam is the International English

Language Testing System (IELTS), taken by millions worldwide each year. IELTS is considered an international high-stakes test that provides reliable evidence of a test taker's English proficiency. It is administered, developed, and researched by organizations such as the British Council, University of Cambridge, and IDP - International Development Program of Australian Universities and Colleges.

Writing, being the most complex and demanding language learning skill, necessitates students to possess acceptable writing skills in an academic context (Hyland, 2011; Jahin & Idrees, 2012; Nueva, 2016). Several studies found that some characteristics of academic writing such as content, purpose, vocabularies, grammar mechanics, and well-organized format are problematic for students (Brown & Lee, 2015; Chen & Kent, 2020; Jafary, 2014; Jahin, 2012; Smith, 2015).

In the quest for effective corrective measures or tools, an empirical study was conducted in an intensive English program and the findings revealed that written feedback could positively affect the students' linguistic abilities (Evans & Morrison, 2011). Corrective feedback (CF) strategies have been classified in various ways, including explicit feedback, clarification requests, direct, indirect, metalinguistic feedback, recasts, repetition, elicitation, and reformulation. Such a sociocultural paradigm allows us to define written corrective feedback (WCF) as follows: it entails (a) reflecting what that individual learner needs most, as writing is a productive skill, and (b) a fundamental approach to teaching involves that writing tasks and feedback are relevant, timely, persistent, and practicable for both providers and students (Anderson, 2023; Burke & Pieterick, 2010; Easton et al., 2022; Ellis, 2009).

It is notable that the IELTS test has been the subject of extensive research. Several studies examined the wash back effect of IELTS, others focused on the internal rater reliability of IELTS writing, and Some investigated the authenticity of writing tasks in IELTS (Fenton-Smith & Humphreys, 2017; Johnson & Tweedie, 2021; Sulaiman, 2012). One particular study investigated the potential relationship between the analytical essay-scoring rubric in IELTS and thematic progression patterns criteria (Soleymanzadeh & Gholami, 2014). With 45 Iranian participants, Ahmadi, Riasati, and Bavali conducted a study to evaluate test takers' performance in IELTS academic writing, focusing on the table and bar chart format of Task 1 (Ahmadi et al., 2019). In a study more closely related to the feedback concerns of the present research, Alavi, Nemati, and Dorri Kafrani investigated the more problematic features of IELTS academic writing Task 2, emphasizing the need for corrective feedback and support (Alavi et al., 2020). Their results indicated that effective training, feedback, and practice were key to achieving better results. Familiarization with IELTS writing through sample analysis and timely feedback were also highlighted as important factors. In a similar vein, Pearson conducted a valuable study that aligns with the present research, which explored written corrective feedback specifically in Task 2 of the IELTS writing test based on teachers' reactions (Pearson, 2018). Results revealed that feedback types such as direct and prescriptive comments were found to be useful. Furthermore, teachers' theoretical knowledge about feedback, along with their experience and personal beliefs, played a role in selecting appropriate methods. The positive role of feedback was further affirmed in the related studies which investigated academic writing task 2 of the IELTS exam (Sanavi & Nemati, 2014).

Table 1. Advantages and Disadvantages of Analytic Scoring

Advantages	Disadvantages
Encourages raters to address the same features	May divert attention from overall essay effect
Allows more diagnostic reporting	Rating one scale may influence others
Assists reliability as candidates gets several scores	Very time consuming compared with the holistic method
Detailed criteria allow easier rater training	Writing is more than simply the sum of its parts
Prevents conflation of categories into one	Favors essays where scalable info easily extracted
Allows teachers to prioritize specific aspects	Descriptors may overlap or be ambiguous

Table 2. Assessing Test Usefulness: A Comparative Analysis of Holistic and Analytic Scales Across Six Key Dimensions (Ferris, 1999; Weigle, 2002)

Quality	Holistic Scales	Analytic Scales
Reliability	Lower than analytic, but still acceptable	Higher than holistic
Construct Validity	Holistic scale assumes that all relevant aspects of writing ability develop at the same rate and can thus be captured in a single score; holistic scores correlate with superficial aspects such as length and handwriting	Analytic scales more appropriate for L2 writers as different aspects of writing ability develop at different rates
Practicality	Relatively fast and easy	Time-consuming; expensive
Impact	The single score may mask an uneven writing profile and may be misleading for placement	More scales provide useful diagnostic information for placement and/or instruction more useful for rater training
Authenticity	White (1998) argues that reading Holistically is a more natural process than reading analytically	Raters may read holistically and adjust analytic scores to match the holistic impression
Interactiveness	n / a	n / a

Table 3 figure out some detail of IELTS band descriptors adapted from the IELTS task writing band descriptors public version.

Table 3. IELTS Writing Task Scoring Descriptor.

Task Response	Coherence and Cohesion
How the prompt is addressed properly	Arrangement and organization of main ideas
The relevance of the position presented in the text	Paragraphing
Support and extension of main ideas	Referencing and substitution
Clarity and justification of conclusions drawn	Use of cohesive markers

Adapted from Pearson (2018) based on the public IELTS Writing Task 2 band descriptor. In preparation for IELTS, candidates often engage in practice tasks, including writing, with feedback from teachers, which is the focus of this study. This study aims to elucidate effective strategies for advanced writing by analyzing a standardized exam like IELTS and also try to re-evaluate the role of feedback in academic writing by exploring the effectiveness of using model essays as a corrective feedback tool. Teachers' feedback, which may include error explanation, the introduction of new ideas, and the provision of examples or models, plays a crucial role in helping students correct their writing.

2. Method

The Method section describes in detail how the study was conducted, including conceptual and operational definitions of the variables used in the study, Different types of studies will rely on different methodologies; however, a complete description of the methods applied enables the reader to evaluate the appropriateness of our selected methods and the reliability and the validity of the associated results, It also presents a unique opportunity for experienced investigators to replicate the study with a focus on other related aspects pertinent to Feedback provision..

2.1 Research Design

The study adopts a quasi-experimental design encompassing pre-test, post-test, and delayed post-test phases, with the latter examination being administered in the subsequent month following the post-test. Furthermore, a mediation model has been employed to evaluate the impact of model essay instruction on the IELTS rubric via these avenues. The entire experimental procedure spans ten two-hour sessions and the dependent variable centers on the enhancement of the candidates' writing abilities, while the independent variables include IELTS writing sub-scales and model essays.

2.2 Participants

As delineated in Table 4, the overall participant cohort comprises 57 individuals, with thirty individuals allocated to the experimental group (EG) and twenty-seven individuals to the control group. The age spectrum of the participants ranges from 19 to 27, encompassing exclusively Iranian students across diverse academic disciplines, all of whom have attained an upper-intermediate level of English proficiency. The experimental group is composed of 16 females and 14 males, while the control group comprises 14 females and 13 males.

Table 4. The studied groups and the number of participants

Group	N Calculated	Attrition	Participants
Experimental	30	0	30
Control	30	3	27
Total	60	3	57

To investigate how this approach impacts two IELTS writing task scoring rubrics, task achievement/response and coherence / cohesion investigated. Our research questions and hypotheses are as follows:

2.3 Research Questions and Hypotheses

(1) Is there a significant difference in gain scores on the IELTS mock writing test between the experimental group (receiving model essays) and the control group for both Task 1 and Task 2?

Null Hypothesis: There is no significant difference in learners' gain scores on the IELTS mock writing test between the experimental group and the control group for both Task 1 and Task 2.

(2) Is there a significant difference in the 'task response/achievement' sub-scale of IELTS writing scoring rubric between the experimental group (receiving model essays) and the control group for both Task 1 and Task 2?

Null Hypothesis: There is no significant difference in learners' gain scores on the 'task response/achievement' sub-scale between the experimental group and the control group for both Task 1 and Task 2.

(3) Is there a significant difference in the 'coherence and cohesion' sub-scale of the IELTS writing scoring rubric between the experimental group (receiving model essays) and the control group for both Task 1 and Task 2?

Null Hypothesis: There is no significant difference in learners' gain scores on the 'coherence and cohesion' sub-scale between the experimental group and the control group for both Task 1 and Task 2.

The study draws on three main theories: the social constructivist theory of Vygotsky, the Interaction Hypothesis, and the cognitive theory and Schmidt's Noticing Hypothesis. Recent literature supports the idea that corrective feedback promotes learning by inducing learners to notice and correct errors, ultimately aiding in interlanguage development.

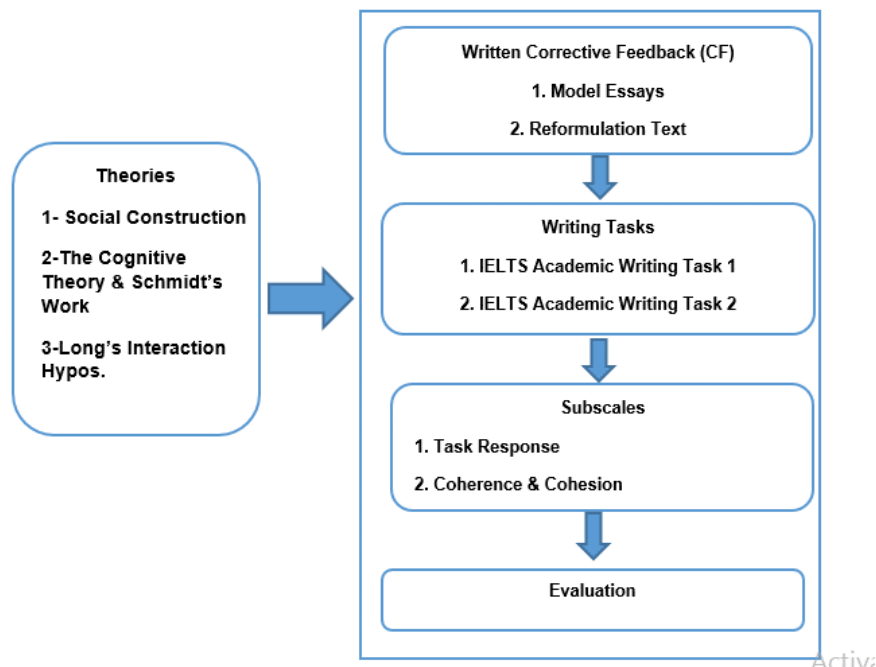


Figure 1. Theoretical relationship of the study

Figure 1 illustrates the research model, highlighting the intervention of model essay instruction, the use of reformulation as an analytical method, and the syntactic approach. To determine the primary impact of feedback on IELTS writing, several relevant articles have been reviewed to present the findings.

2.4 Instrument and Materials

Before commencing the study, a panel of experts meticulously scrutinized the validity of the questionnaire, lesson plan, teaching schedules, model essays, and writing tasks to ensure their robustness as research instruments. The selection of writing tasks was a deliberate process, carefully chosen from Taylor and Asmundson's repertoire, encompassing two distinct compositions: a descriptive essay and an argumentative piece (Taylor & Asmundson, 2008). A prescribed minimum word count requirement of 150 words for Task 1 and 250 words for Task 2 was enforced. Furthermore, the model essays utilized in this study were sourced from the same academic work. Two proficient native English raters validated the models as exemplars of native-level English writing. The study's array of materials included a TOEFL test (applied as the pretest), IELTS sample essays drawn from reputable sources such as "The IELTS Masterclass," "Cambridge IELTS Test 18," "IELTS on track," "Objective IELTS," and "IELTS Preparation Course" by Penny Cameron. Additionally, the IELTS Task 2 Writing band descriptors, accessible to the public, were incorporated into the study materials. The appointed raters were seasoned ESL lecturers, boasting a minimum of 12 years of experience in rating, with a track record of four years in evaluating IELTS essays at the British Council. The number of raters mirrored the approach employed by researchers, consisting of two raters alongside a reserved individual (Ahmadi et al., 2019; Sanavi & Nemati, 2014). The study's core IELTS criteria comprised Task Achievement/Task Response and "Coherence and Cohesion." Task Achievement/Task Response particularly examined the extent to which candidates' writing effectively conformed to the test's rigorous requirements, including adherence to prescribed word count limits. It critically evaluated whether candidates adeptly constructed persuasive arguments in direct response to the provided prompt, substantiated by a profound foundation of knowledge, personal experiences, concrete evidence, and compelling examples. On the flip side, the Coherence and Cohesion criterion delved into the text's lucidity and fluency. To be specific, coherence pertained to the consistent logical interconnection between each sentence and paragraph, whereas cohesion scrutinized the adept and judicious use of cohesive linguistic tools (e.g., conjunctions, pronouns, synonyms) to deftly establish referential relationships between sentences and paragraphs.

2.5 Procedure

As delineated previously, a TOEFL test and an IELTS writing test were administered as pretests for both the experimental and control groups. The strategic sequencing of the IELTS writing test following the TOEFL pretest was methodically chosen to nullify any potential 'fatigue' factor that could have skewed the results. To ensure uniformity and nullify the impact of handwriting variability on raters, all writing samples were systematically transcribed into typed format. Prior to initiating the rating process, an orientation session was conducted with both raters. During this session, the procedural intricacies, and the mechanics of using the scoring guide were thoroughly explained. It's worth noting that the study's underlying objectives remained intentionally undisclosed to the raters. Subsequently, during the second session, candidates underwent a comprehensive orientation to the various components of the writing tasks and the nuances of IELTS band scoring in both the experimental and control groups. However, it's essential to highlight that only the treatment group had the privilege of engaging in a thorough analysis of a model essay.

3. Results and Discussion

In terms of internal validity, a thorough examination of the demographic variables of participants in both groups was conducted. Statistical analyses were employed, including independent t-tests and chi-square tests, to ascertain the comparability of these variables.

As indicated by the statistical data presented in Tables 6, 7, and 8, there were no significant differences observed between the control and experimental groups in terms of gender, age, and English proficiency for both the experimental (EG) and control (CG) groups.

Table 5. The Rater Reliability Indices for Rater 1 and 2

	Pre-test	Post-test
Inter-rater Reliability	82.9	84.1
Rater No.1	95.1	97.4
Rater No.2	96.6	94.8

Table 6. Checking gender preexisting differences

		male	Female	Total
Control	Count	13	14	27
	% within group	48.10%	51.90%	100.00%
Experimental	Count	14	16	30
	% within group	46.70%	53.30%	100.00%
Total	Count	27	30	57
	% within group	47.40%	52.60%	100.00%

Table 7. Comparing the age differences between the control and experimental groups

Group	Min	Max	Mean	Std. Deviation	Std. Error Mean	t	p value
Control	20	26	22.814	1.545	0.297	0.239	0.812
Experimental	19	27	22.933	2.116	0.386		

Table 8. Comparison of English proficiency differences between the control and experimental groups

Group	N	Mean	Std. Deviation	Std. Error Mean	t	p value
Control	27	426.33	70.319	13.533	0.475	0.637
Experimental	30	417.3	72.897	13.309		

3.1 Research Hypothesis and Questions

To assess the hypotheses, a two-way repeated measure ANOVA was employed along with a post hoc test (Bonferroni) to examine potential group and test score variations, specifically for Task 1 and Task 2. Task 1 data

are presented in Tables 9, 10, and 11, accompanied by Figure 3, while Task 2 information is displayed in Tables 12, 13, and 14, along with Figure 4.

Table 9. Descriptive statistic (Mean, SD) of the holistic score for Task1

TEST	Group	N	Mean	SD
Pre-test	Control	27	4.056	0.467
	Experimental	30	4.05	0.480
Post test	Control	27	5.148	0.412
	Experimental	30	5.633	0.642
Follow up	Control	27	5.185	0.442
	Experimental	30	5.617	0.552

The results of the repeated measure ANOVA concerning the holistic scores for Task 1 indicated a statistically significant interaction between group and test ($F(2, 110) = 10.811, p < 0.05, \eta^2 = 0.164$). Subsequently, to assess the related hypothesis, a post hoc test (Bonferroni) was applied to compare the mean scores. Table 9 illustrates that the range of differences among the intra-group results for the three tests is consistent.

The first null hypothesis posits that the utilization of model essays has no impact on learners' writing proficiency, as defined by their gain scores in the IELTS mock writing test.

Table 10. The mean differences in holistic scores between the Experimental and Control Groups across the pre-test, post-test, and delayed post-test for Task 1

	Test	Mean Difference	SE	P value	Partial Eta Squared
Control Vs Experimental	pretest	0.006	0.126	0.965	0.0
	Post test	-0.485*	0.145	<0.05	0.17
	Follow up	-0.431*	0.133	<0.05	0.16

The results of the Bonferroni test revealed that the holistic scores obtained for Task 1 between the Control Group (CG) and Experimental Group (EG) in the pretest were not statistically significant ($p > 0.05$). However, significant differences emerged between the CG and EG for holistic scores in both the post-test ($p < 0.05$) and delayed post-test ($p < 0.05$). Consequently, the first hypothesis was not supported, indicating that the model essay approach effectively improved the candidates' writing proficiency (see Table 10). Table 11 provides an overview of the effectiveness of the model essay approach.

Table 11. The difference of Task 1 holistic mean scores between tests in Experimental and Control Groups for Task1

Group	(I) TEST	(J) TEST	Mean Difference	SE	P value	Partial Eta Squared
Control	Pre test	Post test	-1.093*	0.076	<0.05	
	Pre test	Follow up	-1.130*	0.094	<0.05	0.806
	Post test	Follow up	-0.037	0.081	1	
Experimental	Pre test	Post test	-1.583*	0.072	<0.05	
	Pre test	Follow up	-1.567*	0.089	<0.05	0.905
	Post test	Follow up	0.017	0.077	1	

The Bonferroni test results revealed that within the Experimental Group (EG), there was a significant difference in the holistic score of Task 1 between the pretest and post-test ($p < 0.05$). The mean holistic score for Task 1 in the EG increased by 1.58 units. However, there was no significant difference in the mean holistic score between the post-test and delayed post-test in the EG ($p > 0.05$).

In contrast, the results for the Control Group (CG) showed no significant difference in the mean holistic score of Task 1 between the pretest and post-test ($p > 0.05$). The mean holistic score in the CG improved by 1.09 units. Similarly, there was no significant difference in the mean holistic score between the post-test and delayed post-test in the CG ($p > 0.05$). It is notable that Figure 3 illustrates the improvement of the groups.

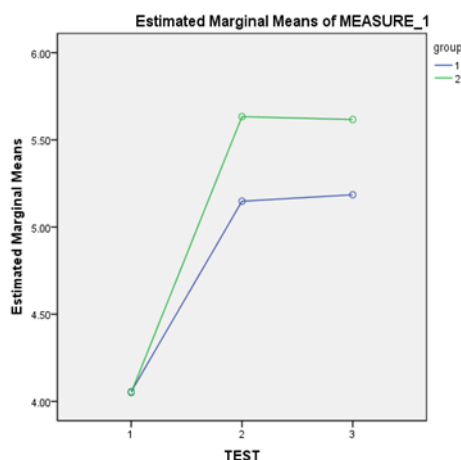


Figure 2. Mean plot of the holistic score of Task1 in both EG (2) and CG (1) groups

As can be seen, for both groups the Task1 holistic gain score increased in the posttest, however there is no considerable change with an interval of one month, when the delayed posttest was administered. Also, for Task 2 the overall procedures and changes were similar to Task 1.

Table 12. Descriptive statistic (Mean, SD) of the holistic score of Task2

TEST	Group	N	Mean	SD
pre test	Control	27	4.13	0.565
	Experimental	30	3.883	0.536
Post test	Control	27	5.407	0.555
	Experimental	30	5.833	0.634
Follow up	Control	27	5.315	0.540
	Experimental	30	5.783	0.486

Table 13. The holistic mean difference between Experimental and Control Groups in pre, post and delayed posttest for Task2

Test	Mean Difference	SE	P value	Partial Eta Squared
Control Vs Experimental pretest	0.246	0.146	0.097	0.049
Control Vs Experimental Post test	-0.426*	0.159	0.01	0.116
Control Vs Experimental Follow up	-0.469*	0.136	0.001	0.178

Table 14. The holistic mean score difference between the tests in Experimental and Control Group for Task2

Group	(I) TEST	(J) TEST	Mean Difference	SE	P value	Partial Eta Squared
Control	Pretest	Post test	-1.278*	0.07	<0.05	0.863
	Pretest	Follow up	-1.185*	0.077	<0.05	
	Post test	Follow up	0.093	0.059	0.369	
Experimental	Pretest	Post test	-1.950*	0.067	<0.05	0.944
	Pretest	Follow up	-1.900*	0.073	<0.05	
	Post test	Follow up	0.05	0.056	1	

Therefore, the first null hypothesis was rejected and it is concluded that the model essay was effective on the improvement of the holistic score of participants for Task 2.

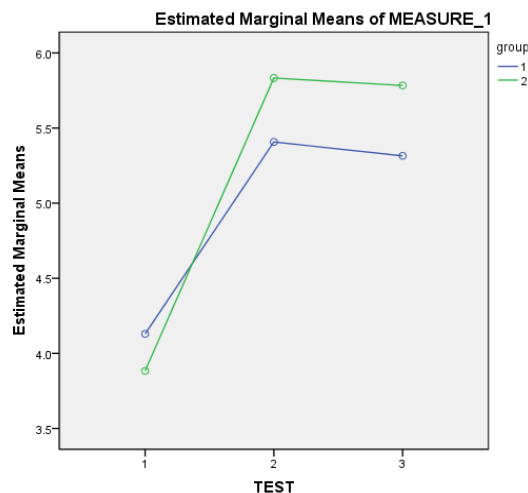


Figure 3. Mean plot for the holistic score of Task 2 in both EG (2) and CG (1) groups

3.2 Question 1

The primary research question addressed in this study referred to discerning the variance in gain scores within the IELTS holistic scoring system for mock writing assessments between two distinct groups: the experimental group (which received model essays) and the control group. This analysis encompassed both Academic writing Task 1 and Task 2. The salient findings from the study can be summarized as follows:

(1) **Statistical Significance in Gain Scores:** A statistically significant difference was observed when comparing the pretest and post-test holistic scores for both Task 1 and Task 2 within the experimental group. This indicates that the intervention involving the provision of model essays had a noticeable impact on the writing performance of participants.

(2) **Lack of Significance between Post-test and Delayed Test:** Interestingly, there was no statistically significant distinction in the mean holistic scores between the post-test and the subsequent delayed test for both Task 1 and Task 2 within the experimental group. This suggests that the improvements registered in the immediate post-test were not maintained over time.

(3) **Control Group's Parallel Results:** In a parallel manner, the control group also demonstrated statistically significant differences between their pretest and posttest holistic scores for both Task 1 and Task 2. This implies that some form of improvement was occurring even in the absence of the model essay intervention. However, like the experimental group, the control group did not exhibit a significant discrepancy between their post-test and delayed test scores.

(4) **Magnitude of Improvement Comparison:** While both groups displayed improvements, the experimental group (EG) outperformed the control group (CG). The mean holistic scores for the control group increased by 1.09 units for Task 1 and 1.27 units for Task 2. Conversely, the experimental group achieved greater improvements, with 1.58- and 1.95-unit improvements for Task 1 and Task 2, respectively.

(5) **Consistency with Prior L2 Writing Research:** These findings align with earlier studies in the field of second language (L2) writing. Specifically, they echo research that involved comparing students' original writing with its reformulated version following syntactic writing instruction, as well as research that compared students' original writing with model essays (Green, 2007; Truscott, 2007).

In conclusion, this study's results suggest that providing model essays to students in the experimental group led to statistically significant improvements in their writing scores for both Task 1 and Task 2, surpassing the gains observed in the control group. However, it's noteworthy that the progress achieved immediately after the intervention was not maintained when assessed in a delayed post-test. These findings are consistent with previous research in the realm of L2 writing.

3.2.1 Effect of both Writing Instruction on Mean Score of Task Response Score for Task 1 and Task 2 in EG and CG Statistics and Data Analysis

The Bonferroni test outcomes provide valuable insights into the task response scores, highlighting differences between the Control Group (CG) and the Experimental Group (EG) at different assessment stages. Tables 15, 16, and 17 show the relevant statistics. These key findings are summarized in Table 16:

(1) In the pre-test, the comparison of task response scores between the CG and EG yielded non-significant results ($p > 0.05$). This suggests that, initially, there was no discernible difference in task response performance between the two groups.

(2) In the post-test, the analysis revealed a significant contrast between the EG and CG for the task response rubric of Task 1 ($p < 0.05$). This indicates that after the intervention, the EG displayed a notably better task response performance compared to the CG.

(3) This distinction between the EG and CG persisted in the delayed post-test, with statistical significance observed ($p < 0.05$). It is evident that the EG continued to maintain a superior level of task response performance compared to the CG.

Further insights into the magnitude of these changes are provided in Table 17, which includes both p-values and Mean Differences:

(4) For the EG, there was a substantial improvement of 1.86 units in the mean task response score. However, it's noteworthy that there was no significant difference between the post-test and delayed post-test scores ($p > 0.05$). This implies that the improvements achieved in task response were sustained over time within the EG.

(5) Similarly, the CG also demonstrated a statistically significant improvement ($p < 0.05$) with a mean increase of 1.33 units in task response scores. Nevertheless, akin to the EG, there was no meaningful difference between the post-test and delayed post-test scores ($p > 0.05$) in the CG. This suggests that the improvement in task response within the CG was not time-sensitive but remained consistent.

Table 15. Descriptive statistic (Mean, SD) of task response score in Task1

TEST	Group	N	Mean	SD
pre test	Control	27	4.056	0.594
	Experimental	30	3.983	0.549
Post test	Control	27	5.389	0.543
	Experimental	30	5.85	0.800
Follow up	Control	27	5.333	0.679
	Experimental	30	5.817	0.748

Table 16. Task response mean the difference between Experimental and Control Groups in pre, post and delayed posttest for Task1

	Test	Mean Difference	SE	P value	Partial Eta Squared
Control Vs Experimental	pre test	0.072	0.151	0.635	0.049
	Post test	-0.461*	0.183	0.015	0.116
	Follow up	-0.483*	0.19	0.014	0.178

Table 17. The holistic mean score difference between tests in Experimental and Control Group for Task1

Group	(I) TEST	(J) TEST	Mean Difference	SE	P value	Partial Eta Squared
Control	Pretest	Post test	-1.278*	0.07	<0.05	
	Pretest	Follow up	-1.185*	0.077	<0.05	0.863
	Post test	Follow up	0.093	0.059	0.369	
Experimental	Pretest	Post test	-1.950*	0.067	<0.05	
	Pretest	Follow up	-1.900*	0.073	<0.05	0.944
	Post test	Follow up	0.05	0.056	1	

In summary, the Bonferroni test results and accompanying statistical values illustrate the significant improvements in task response scores for both the Experimental Group (EG) and Control Group (CG) following the intervention. These improvements were sustained over time, as indicated by the lack of significant differences between post-test and delayed post-test scores within both groups. The findings accentuate the efficacy of the intervention in enhancing task response performance.

3.2.2 Task Response Score for Task 2

Table 18. Descriptive statistic (Mean, SD) for task response score in Task 2

TEST	Group	N	Mean	SD
pre test	Control	27	4.037	0.706
	Experimental	30	3.85	0.632
Post test	Control	27	5.519	0.802
	Experimental	30	6.133	0.706
Follow up	Control	27	5.389	0.813
	Experimental	30	6.117	0.611

Table 19. Task response mean: the difference between Experimental and Control Groups in pre, post and delayed posttest for Task2

	Test	Mean Difference	SE	P value	Partial Eta Squared
Control Vs Experimental	pretest	0.187	0.177	0.296	0.02
	Post test	-0.615*	0.2	0.003	0.147
	Follow up	-0.728*	0.189	<0.05	0.212

Table 20. The holistic mean score: the difference between tests in Experimental and Control Group for Task2

Group	(I) TEST	(J) TEST	Mean Difference	SE	P value	Partial Eta Squared
Control	Pretest	Post test	-1.481*	0.078	<0.05	
	Pretest	Follow up	-1.352*	0.095	<0.05	0.874
	Post test	Follow up	0.13	0.084	0.389	
Experimental	Pretest	Post test	-2.283*	0.074	<0.05	
	Pretest	Follow up	-2.267*	0.09	<0.05	0.950
	Post test	Follow up	0.017	0.08	1	

The data presented in the tables suggests a similarity in the procedural patterns leading to developments in Task 2, akin to those observed in Task 1. Consequently, the null hypothesis regarding the absence of improvement in the 'task response/achievement' sub-scale of the IELTS writing scoring rubric must be rejected. It is evident that both groups demonstrated improvements in the gain scores within the 'task response/achievement' sub-scale of the IELTS writing scoring rubric.

3.3 Question 2

Research Question 2 was directed toward discerning disparities in the 'task response/achievement' sub-scale of the IELTS writing scoring rubric between the experimental group (receiving model essays) and the control group across academic writing tasks, specifically Task 1 and Task 2.

The outcomes of the Bonferroni test revealed that the task response scores exhibited significant differences between the Control Group (CG) and the Experimental Group (EG) for both Task 1 and Task 2 in both the posttest and the delayed posttest when compared to the pretest. Notably, the degree of improvement in both tasks was more pronounced within the Experimental Group (EG). Consequently, it can be inferred that model essays served as a more efficacious feedback tool compared to mere error correction or reformulated text. The rationale behind the heightened level of significance in the "task response" sub-scale disparity between the EG and CG groups, concerning the two sub-scales of the IELTS writing rubric, stems from the EG group's adeptness in recognizing and integrating the descriptors outlined in the band descriptor, which were explicitly explained in their model texts as part of the written corrective feedback. Additionally, the model essays provided during the

treatment sessions facilitated a noticeable distinction of topic keywords, thereby contributing to the group's enhanced performance.

3.3.1 Effect of Both Interventions on Mean Score of Coherence and Cohesion: Task 1 & Task 2

The subsequent null hypothesis posited that the implementation of model essays would have no evident impact on students' gain scores within the 'cohesion & coherence' sub-scale of the IELTS writing scoring rubric, encompassing both Task 1 and Task 2. To assess its validity, a similar analytical procedure was applied as in the previous hypotheses. The results, once again, unequivocally demonstrated the presence of a statistically significant distinction between the two sets of scores for both Task 1 and Task 2. As anticipated, this null hypothesis was also rejected, affirming the efficacy of the instructional approach involving model essays. Comprehensive statistical analysis for Task 1 can be found in Tables 21, 22, and 23, while analogous results for Task 2 are available in Tables 24, 25, and 26.

Table 21. Descriptive Statistics (Mean, SD) of Coherence & Cohesion Scores for Task 1

TEST	Group	N	Mean	SD
Pre test	Control	27	3.759	0.594
	Experimental	30	3.717	0.449
Post test	Control	27	4.611	0.610
	Experimental	30	5.05	0.735
Follow up	Control	27	4.648	0.569
	Experimental	30	5.05	0.674

Table 22. Differences in Coherence & Cohesion Means Between Experimental and Control Groups in Pre, Post, and Delayed Posttest for Task 1

	Test	Mean Difference	SE	P value	Partial Eta Squared
Control Vs. Experimental	pre test	0.043	0.139	0.76	0.002
	Post test	-0.439*	0.18	0.018	0.097
	Follow up	-0.402*	0.166	0.019	0.096

Table 23. Differences in Coherence & Cohesion Mean Scores Between Tests in the Experimental and Control Groups for Task 1

Group	(I) TEST	(J) TEST	Mean Difference	SE	P value	Partial Eta Squared
Control	Pre test	Post test	-0.852*	0.11	<0.05	
	Pre test	Follow up	-0.889*	0.117	<0.05	0.568
	Post test	Follow up	-0.037	0.099	1	
Experimental	Pre test	Post test	-1.333*	0.105	<0.05	
	Pre test	Follow up	-1.333*	0.111	<0.05	0.775
	Post test	Follow up	0	0.094	1	

The findings of the post hoc test unveiled a noteworthy disparity in the coherence score between the pretest and posttest within the experimental group, demonstrating statistical significance ($p < 0.05$). Specifically, the mean coherence score within the experimental group exhibited a substantial increment of 1.33 units. However, it is notable that there was no remarkable variation in the coherence mean between the posttest and the subsequent delayed test ($p > 0.05$) within the same experimental group. A parallel pattern was observed in the control group, where the results similarly indicated statistical significance ($p < 0.05$). Here, the mean coherence score within the control group demonstrated an increase of 0.852 units. However, similar to the experimental group, there was no observable disparity in the coherence mean between the posttest and the delayed test ($p > 0.05$) within the control group (as illustrated in Table 24).

Table 24. Descriptive statistic (Mean, SD) of coherence & cohesion scores for Task2

TEST	Group	Mean	SD
Pre test	Control	3.741	0.447
	Experimental	3.65	0.575
Post test	Control	4.741	0.544
	Experimental	5.167	0.813
Follow up	Control	4.704	0.697
	Experimental	5.15	0.721

Table 25. Coherence & cohesion means: the difference between Experimental and Control Groups in pre, post and delayed posttest for Task 2

	Test	Mean Difference	SE	P value	Partial Squared	Eta
Control Vs. Experimental	pre test	0.091	0.137	0.512	0.008	
	Post test	-0.426*	0.185	0.025	0.088	
	Follow up	-0.446*	0.188	0.021	0.093	

Table 26. Coherence & cohesion means score difference between tests in Experimental and Control Group for Task 2

Group	(I) TEST	(J) TEST	Mean Difference	SE	P value	Partial Squared	Eta
Control	Pre test	Post test	-1.000*	0.087	<0.05		
	Pre test	Follow up	-0.963*	0.112	<0.05	0.745	
	Post test	Follow up	0.037	0.118	1		
Experimental	Pre test	Post test	-1.517*	0.083	<0.05		
	Pre test	Follow up	-1.500*	0.106	<0.05	0.883	
	Post test	Follow up	0.017	0.112	1		

3.4 Question 3

The results of the study affirm the anticipated outcome for the third research question, which centers on whether a significant discrepancy exists in the 'coherence and cohesion' sub-scale of the IELTS writing scoring rubric between the experimental group (receiving model essays) and the control group.

The outcomes of the post hoc test substantiated the presence of a significant difference between the pretest and post-test scores for 'coherence & cohesion' within both the Experimental Group (EG) and Control Group (CG). However, it's crucial to note that the magnitude of improvement was not uniform across the two groups and tasks.

In Task 1, the EG demonstrated a substantial improvement of 1.33 units in their 'coherence & cohesion' score, while the CG exhibited an improvement of 0.852 units. In Task 2, the EG recorded an improvement of 1.51 units, while the CG achieved an improvement of 1 unit.

Beyond the conventional interpretation of these variations, a remarkable observation pertains to the differing levels of improvement between Task 1 and Task 2. This can be attributed to the unique characteristics of Task 2, where the provision of model essays offered students a more frequent opportunity to recognize and practice discourse-related elements. In contrast, Task 1 primarily requires students to extract and present key trends from graphic data, which is inherently more focused on descriptive essay writing with fixed sentential structures, report verbs, and data analysis.

Consequently, students engaging in Task 2 essays naturally grapple with discourse-related aspects such as paragraph organization, inter-sentential relationships, and coherence and cohesion, given the specific demands of this task. This distinction in task requirements likely contributed to the varying levels of improvement observed between Task 1 and Task 2 in both the experimental and control groups.

3.4.1 Correlation between Holistic Score and 2 Sub-scales of IELTS Writing Rubric in Post-test

Based on the findings from Pearson's correlation coefficients (Table 27), a noteworthy positive relationship emerged between IELTS scores and both task response ($r= 0.836, p< 0.01$) and coherence ($r= 0.775, p<0.01$).

Table 27. Correlation between holistic score and two sub-scales of IELTS writing rubric

	IELTS	post-task response	post coherence
IELTS	1		
post task response	0.836**	1	
post coherence	0.775**	.639**	1

** Correlation is significant at the 0.01 level (2-tailed).

The results of additional analyses affirmed a significant relationship among the research variables, prompting the inclusion of all dimensions for hypothesis testing. The mediation analysis was conducted using Hayes and Preacher's SPSS PROCESS macro.

Table 28. Result of mediation analyses incorporating the model variables (post-test)

IV to Mediators (a paths)				
	Coefficient	SE	t	P value
Task response	0.5636	0.1419	3.9709	0.0002
Coherence	0.4302	0.1379	3.1189	0.0029
Direct Effects of Mediators on DV (b paths)				
Task response	0.2971	0.0713	4.1667	0.0001
Coherence	0.1498	0.072	2.0809	0.0425
Total Effect of IV on DV (c path)				
intervention	0.3852	0.1206	3.193	0.0023
Direct Effect of IV on DV (c' path)				
intervention	0.0516	0.0566	0.9122	0.3659

Table 29. Model Summary for DV Model

Model Summary for DV Model				
	R2	Adj R2	F	P value
model	0.8688	0.8559	67.5327	<0.01
Indirect Effects of IV on DV through Proposed Mediators (ab paths)				
	Data	Boot	Bias	SE
TOTAL	0.437	0.442	0.006	0.116
Task response	0.167	0.170	0.002	0.057
Coherence	0.064	0.062	-0.003	0.035

Dependent variable (Y) = IELTS SCORE

In the mediation analysis (Table 28 and 29), significant findings emerged regarding the effects of the intervention on two mediators. The intervention exhibited a significant positive impact on task response ($B= 0.563, p < 0.01$), coherence ($B= 0.430, p < 0.05$), lexical ($B= 0.401, p < 0.051$), and grammar ($B=0.392, p < 0.05$), as observed in the "a path." Additionally, both dimensions demonstrated a significant effect on IELTS scores, as evidenced in the "b paths."

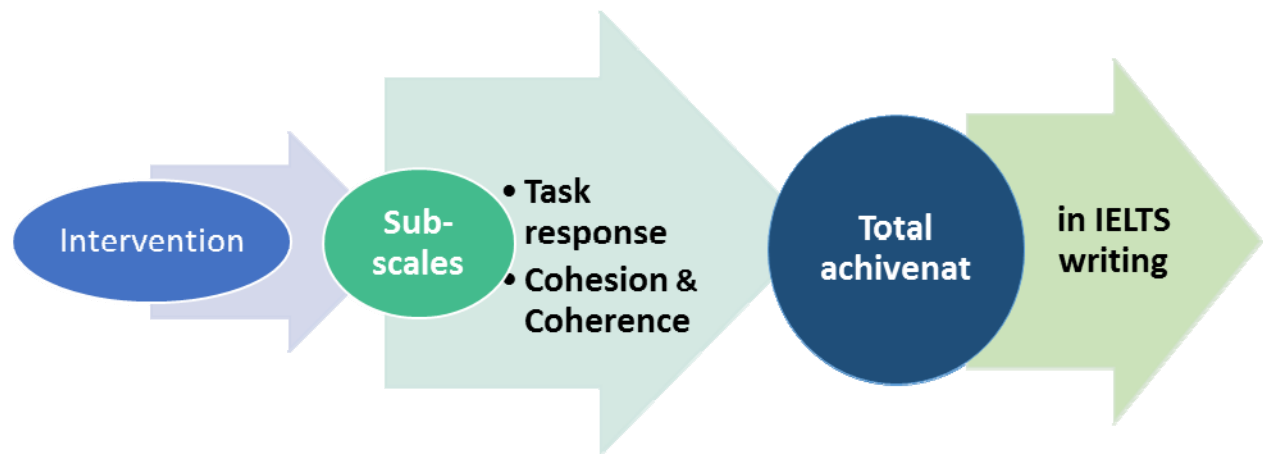


Figure 4. IELTS writing procedure.

4. Conclusion and Implications for Further Studies

The current experimental study delved into the effectiveness of model essays as a form of corrective feedback in the context of Iranian learners, shedding light on their potential to enhance language proficiency. The findings of this study confirmed that employing model essays as a pedagogical strategy facilitated the more effective language use by Iranian learners.

4.1 Intersecting Feedback and Second Language Acquisition (SLA)

The findings echo the growing body of research in Second Language Acquisition (SLA) that highlights the crucial role of feedback in language learning. Feedback, whether written or oral, serves as a bridge between learners' existing linguistic knowledge and their ability to apply it in real communication. It provides a mechanism for learners to reflect on their language production, identify errors or gaps in their proficiency, and make meaningful improvements.

Moreover, the positive correlation between IELTS scores and specific writing rubric sub-scales suggests that feedback interventions can influence broader language proficiency, impacting high-stakes language assessments like IELTS. This interplay between feedback and language proficiency underlines the interconnectedness of language learning and assessment.

4.2 Reflective Pedagogy and Self-Regulated Learning

The study's implications extend to pedagogical approaches that emphasize reflection and self-regulated learning. Model essays, by providing learners with clear examples of effective writing, enable them to engage in metacognitive processes. Learners can compare their own writing to the model essays, identify discrepancies, and iteratively improve their output. This self-monitoring and self-adjustment align with self-regulated learning theories, which emphasize learners' active participation in their own learning process.

4.3 Future Directions

While this study focused on model essays, the broader landscape of feedback in language pedagogy remains multifaceted and rich for exploration. Here are some avenues for future research:

- (1) **Comparative Analysis of Feedback Types:** Future studies can compare the effectiveness of different feedback types, including written feedback, oral feedback, peer feedback, and automated feedback systems. Such comparisons can clarify the relative advantages and disadvantages of each feedback modality in different language learning contexts.
- (2) **Longitudinal Studies:** Investigating the long-term impact of feedback interventions on language development and retention is crucial. Longitudinal studies can reveal whether the improvements observed in the short term persist over time or require ongoing support.
- (3) **Cultural and Contextual Variations:** The study primarily focused on Iranian learners. Future research can explore how the effectiveness of feedback interventions varies across different learner populations, cultural backgrounds, and educational contexts.
- (4) **Multimodal Feedback:** Integrating various feedback modalities, such as model essays, audio commentary,

and video exemplars, can provide a richer and more comprehensive feedback experience. This approach aligns with contemporary pedagogical trends that emphasize multimodal learning resources.

(5) Broader Examination of Scoring and Feedback Methods: As the study analyzed text according to the Task 2 band descriptor, future research can use various rubrics to examine the functionality of other scoring and feedback methods.

(6) Proficiency Tests with Global Recognition: Given the proliferation of proficiency tests with global recognition, future experiments can investigate the efficacy of feedback approaches for exams like CPE or domestic language proficiency assessments, providing valuable insights for the evolving landscape of language assessment and instruction.

In conclusion, the study emphasizes the transformative potential of feedback interventions in language pedagogy, connecting with broader concepts in SLA, pedagogical theory, and self-regulated learning. As we continue to unravel the intricacies of language acquisition and instruction, research in feedback strategies remains at the forefront of innovations in language education.

References

- Ahmadi, S., Riasati, M. J., & Bavali, M. (2019). A Comparison of Writing Performance of Iranian IELTS Candidates Facing Chart Topics vs. Table Topics in Academic Writing (Task1). *International Journal of Instruction*, 12(4), 17-34. <https://doi.org/10.29333/iji.2019.1242a>
- Alavi, S. M., Nemati, M., & Dorri Kafrani, J. (2020). Error gravity analysis of IELTS students' academic writing task 2. *International Journal of Language Studies*, 14(1).
- Anderson, J. (2023). Translanguaging in the ELT classroom. *Modern English Teacher*, 32(2), 46-50.
- Brown, H. D., & Lee, H. (2015). *Teaching principles*. P. Ed Australia.
- Burke, D., & Pieterick, J. (2010). *Giving students effective written feedback*. McGraw-Hill Education (UK).
- Chen, J. C., & Kent, S. (2020). Task engagement, learner motivation and avatar identities of struggling English language learners in the 3D virtual world. *System*, 88, 102168. <https://doi.org/https://doi.org/10.1016/j.system.2019.102168>
- Easton, C., Verdon, S., Brown, L., & Wilson, L. (2022). Building diversity in the speech-language pathology workforce through a blended online Master of Speech Pathology. *International Journal of Speech-Language Pathology*, 24(3), 307-319. <https://doi.org/https://doi.org/10.1080/17549507.2022.2055145>
- Ellis, R. (2009). Task-based language teaching: Sorting out the misunderstandings. *International journal of applied linguistics*, 19(3), 221-246. <https://doi.org/https://doi.org/10.1111/j.1473-4192.2009.00231.x>
- Evans, S., & Morrison, B. (2011). Meeting the challenges of English-medium higher education: The first-year experience in Hong Kong. *English for Specific Purposes*, 30(3), 198-208. <https://doi.org/https://doi.org/10.1016/j.esp.2011.01.001>
- Fenton-Smith, B., & Humphreys, P. (2017). Language specialists' views on the academic language and learning abilities of English as an additional language postgraduate coursework students: towards an adjunct tutorial model. *Higher Education Research & Development*, 36(2), 280-296.
- Ferris, D. (1999). The case for grammar correction in L2 writing classes: A response to truscott (1996). *Journal of second language writing*, 8(1), 1-11. [https://doi.org/https://doi.org/10.1016/S1060-3743\(99\)80110-6](https://doi.org/https://doi.org/10.1016/S1060-3743(99)80110-6)
- Green, A. (2007). Washback to learning outcomes: A comparative study of IELTS preparation and university pre-sessional language courses. *Assessment in Education*, 14(1), 75-97. <https://doi.org/https://doi.org/10.1080/09695940701272880>
- Hyland, F. (2011). The language learning potential of form-focused feedback on writing. *Learning-to-write and writing-to-learn in an additional language*. Philadelphia/Amsterdam: John Benjamins, 159-179.
- Jafary, M. (2014). *Effects of Model Essays on the Improvement of Iranian IELTS Candidates' Writing Ability* [Universiti Putra Malaysia].
- Jahin, J. H. (2012). The effect of peer reviewing on writing apprehension and essay writing ability of prospective EFL teachers. *Australian Journal of Teacher Education (Online)*, 37(11), 65-89.
- Jahin, J. H., & Idrees, M. W. (2012). EFL major student teachers' writing proficiency and attitudes towards learning English. *Umm Al-Qura University Journal of Educational & Psychologic Sciences*, 4(1), 10-72.

- Johnson, R. C., & Tweedie, M. G. (2021). "IELTS-out/TOEFL-out": Is the end of general English for academic purposes near? Tertiary student achievement across standardized tests and general EAP. *Interchange*, 52(1), 101-113. <https://doi.org/https://doi.org/10.1007/s10780-021-09416-6>
- Nueva, J. C. (2016). Genre-Based and Process-Based Approaches to Teaching News Articles. *Pertanika Journal of Social Sciences & Humanities*, 24(1).
- Sanavi, R. V., & Nemati, M. (2014). The effect of six different corrective feedback strategies on Iranian English language learners' IELTS writing task 2. *Sage Open*, 4(2), 2158244014538271. <https://doi.org/https://doi.org/10.1177/2158244014538271>
- Smith, M. M. (2015). Language Development of Individuals Who Require Aided Communication: Reflections on State of the Science and Future Research Directions. *Augment Altern Commun*, 31(3), 215-233. <https://doi.org/https://doi.org/10.3109/07434618.2015.1062553>
- Soleymanzadeh, L., & Gholami, J. (2014). Scoring argumentative essays based on thematic progression patterns and IELTS analytic scoring criteria. *Procedia-Social and Behavioral Sciences*, 98, 1811-1819. <https://doi.org/https://doi.org/10.1016/j.sbspro.2014.03.610>
- Sulaiman, T. (2012). A Review on IELTS Writing Test, Its Test Results and Inter Rater Reliability. *Theory & Practice in Language Studies*, 2(1). <https://doi.org/http://doi:10.4304/tpls.2.1.138-143>
- Taylor, S., & Asmundson, G. (2008). Internal and external validity in clinical research. *Handbook of research methods in abnormal and clinical psychology*, 23, 34.
- Truscott, J. (2007). The effect of error correction on learners' ability to write accurately. *Journal of second language writing*, 16(4), 255-272. <https://doi.org/https://doi.org/10.1016/j.jslw.2007.06.003>
- Weigle, S. C. (2002). *Assessing writing*. Ernst Klett Sprachen.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).