Enhancing EFL Writing Through Online Peer Feedback: A Systematic Review of Higher Education Studies

Ting Liang[1], Charanjit Kaur Swaran Singh[1,*], Dodi Mulyadi[2] & Tarsame Singh Masa Singh[3]

[1]Faculty of Languages and Communication, Universiti Pendidikan Sultan Idris, Tanjong Malim, Perak, Malaysia

[2]Faculty of Foreign Language and Culture, University of Muhammadiyah, Semarang, Indonesia

[3]Language Department, Institute of Teacher Education, Pulau Pinang, Malaysia

*Correspondence: Faculty of Languages and Communication, Universiti Pendidikan Sultan Idris, 35900 Tanjong Malim, Perak, Malaysia. E-mail: charanjit@fbk.upsi.edu.my

Received: November 5, 2024 Accepted: December 28, 2024 Online Published: February 7, 2025

doi:10.5430/jct.v14n1p184 URL: https://doi.org/10.5430/jct.v14n1p184

Abstract

This systematic review critically examines the effectiveness of online peer feedback (OPF) for enhancing EFL writing skills in higher education. Analysis of 24 empirical studies reveals consistently positive impacts of OPF on writing outcomes, with effect sizes ranging from small to extremely large. Key principles for effective OPF implementation include adopting a formative approach, providing structured guidance, incorporating comprehensive training, and facilitating multiple revision opportunities. The review also highlights the benefits of asynchronous interactions and integrating diverse feedback sources. However, methodological limitations in many studies, such as small sample sizes and potential biases, necessitate a cautious interpretation of results. The findings underscore OPF's potential to transform EFL writing instruction while emphasizing the need for more rigorous, large-scale investigations. Future research should employ more stringent experimental designs, explore diverse OPF configurations, and examine the underlying mechanisms driving OPF effectiveness. This review contributes to the growing body of knowledge on technology-enhanced language learning, offering valuable insights for educators and researchers in the field of EFL writing instruction.

Keywords: online peer feedback, EFL writing, higher education, technology-enhanced learning, writing instruction

1. Introduction

Writing proficiency in English is a critical skill for students in higher education contexts where English is learned as a foreign language (EFL). Developing strong written communication abilities allows EFL students to effectively convey ideas, critically analyze texts, and demonstrate mastery of course content across disciplines. However, attaining writing expertise poses significant challenges for many EFL learners due to a number of factors, such as the linguistic and cultural differences between English and their mother tongues, limited exposure to as well as practice in authentic writing contexts, and a lack of effective feedback on their writing. Among these factors, feedback, as an integral part of the writing process, plays a crucial role, because receiving feedback, especially of the constructive kind, identifies areas for improvement and provides guidance for refinement in subsequent revisions.

Traditionally, writing feedback has been provided by instructors in either an oral or written form in a classroom setting, but the rise of digital technologies has enabled alternative approaches like online peer feedback (hereafter OPF). An array of instructional and social-cognitive benefits that students can gain from engaging in OPF for L2 writing have been well documented in previous research, such as authentic and relevant writing experience through engaging in OPF, and ample opportunities offered by peer assessment platforms for meaning negotiation and interaction in the target language (Chang, 2016; Saeed, Ghazali, & Aljaberi, 2018; Yu & Lee, 2016). However, there remains a lack of systematic understanding specifically about the effectiveness of OPF approaches for EFL students' writing development in higher education settings. Existing studies have reported conflicting findings about whether and how OPF enhances EFL writing outcomes. For example, by comparing OPF to face-to-face peer feedback (hereafter FFPF), Liu and Sadler (2003) found a higher number of revision-oriented comments were generated through FtFPF. In contrast, Song and Usaha (2009) reported that the computer-mediated communication (hereafter

CMC) group provided more comments focused on revisions than the face-to-face group, leading to more revisions in subsequent drafts. Additionally, these studies have implemented diverse OPF models with varying study details (e.g., study design, purposes, sample size etc.) and peer assessment (hereafter PA) design elements (e.g., technology, writing assignment, frequency, constellation etc.). This inconsistency across study designs and mixed results signals a need for a critical synthesis of the evidence on the effect of OPF for EFL writing.

Some previous reviews (Cao et al., 2022; Chang, 2016; Chen, 2016; Cuocci et al., 2023; Saeed, Ghazali, & Aljaberi, 2018; Yu & Lee, 2016) include all types of studies in their selection criteria ranging from qualitative, quantitative, to mixed methods, whereas studies examining the effects of an intervention compared to a control group (i.e., experimental design) should be prioritized for inclusion in order to answer questions concerning effectiveness (Petticrew & Roberts, 2008). Only quasi-experiments can be regarded as the second-best evidence where experiments are unavailable. There are also meta-analyses synthesizing the results of either implementing online feedback of varied sources (Lv et al., 2021), or conducting peer feedback in the broad context of L2 writing (Vuogan & Li, 2022), or conflating different learning outcomes (i.e., translation, speaking, and writing) of EFL students engaging in peer feedback activities (Wu et al., 2022). As a result, we believe that there is a need to look at these studies from the perspective of the EFL context and via the lens of writing assessment. Our review was based on primary studies examining OPF as an independent variable, enabling us to make more direct comparisons between the intervention and standard practice or baseline in terms of feedback modality and/or source, as well as PA design elements.

1.1 OPF Approaches to L2 Writing

Peer feedback is used to describe both the process and the result of providing qualitative comments on a student's work without assigning a numerical score or grade (Panadero et al., 2018) in contrast to peer scoring which refers to the situation of merely providing a score by peers. For the purposes of this review, peer feedback and peer scoring are subsumed under the broad category of peer assessment, which refers to students taking on the role of assessors by providing qualitative feedback as well as assigning a score or grade to their peers' work (Liu & Carless, 2006). Online peer feedback (OPF), also known as web-based or electronic peer feedback, is a technology-mediated activity widely carried out in writing courses where students reflect and provide comments to one another with the aim of enhancing writing quality (Breuch, 2004). In terms of feedback mode, OPF can be produced in either written or oral form. As for feedback contact, OPF can be executed synchronously or asynchronously. Besides mode and contact, OPF can be studied in terms of a wide range of miscellaneous elements influencing its effect on performance, which are discussed in detail in the section of design elements. Two major theoretical frameworks underpinning peer feedback practices are collaborative learning theory and writing process theory. Collaborative learning supports peer feedback through meaningful peer interaction (Villamil & de Guerrero, 2006), cognitive involvement in giving and receiving feedback (Rouhi et al., 2020), and revision opportunities from multiple peers (Dochy et al., 1999). Within the multi-step writing process framework (Chenoweth & Hayes, 2001; Flower & Hayes, 1981; Hayes, 1996), peer response enhances L2 writing by deepening students' understanding of themselves as writers and of readers' needs (Rollinson, 2005; Yu & Hu, 2017), providing more feedback, and fostering meaningful peer interaction.

Numerous studies across global higher education contexts have investigated the implementation of OPF approaches to EFL students' writing over the past decade. This evidence has been obtained on several different levels of comparison conditions, for example, comparison of the effects of OPF treatment with other feedback treatments, on overall as well as specific aspects of writing development. Research comparing OPF to FFPF has produced mixed results regarding their impact on overall writing performance. Some studies found no significant difference between the two methods (Vaezi & Abbaspour, 2015), while others reported advantages for traditional FFPF (Braine, 2001). Conversely, several researchers observed superior writing outcomes with OPF, particularly in technology-enhanced environments such as Facebook-based groups (Wahyudin, 2018) and CMC contexts (Mellati & Khademi, 2014). These conflicting findings suggest that the effectiveness of OPF versus FFPF may depend on various factors and contexts. Exploring whether OPF leads to greater changes in revision than face-to-face teacher feedback (here after FFTF), some earlier pre-experimental EFL research found that the OPF group seemed to have made more meaning-related changes than the FFTF group, though a larger percentage of FFTF than OPF was incorporated in the revision (Yang et al., 2006; Zhao, 2010). Some quasi-experimental research showed that the OPF plus online teacher feedback (here after OTF) group outperformed the control group receiving solely OTF in writing performance (Tai et al., 2015), though this could be attributed to increased overall feedback. However, Pham (2021) found no significant differences between the effects of OPF and OTF on student revisions. Additionally, when comparing OPF to self-feedback (here after SF), Kayacan and Razı (2017) observed no significant differences between the treatment and the control groups in terms of final writing task scores. Besides, specific aspects of writing skills were also

considered, for instance, global issues (e.g., ideational quality and rhetorical structure) and local concerns (e.g., grammar, vocabulary, and mechanics) (Ferris, 2012). Some previous studies suggested that increased interactions in peer editing led to both more local and global revisions (Yang & Wu, 2011), while the majority of the EFL students involved in other OPF studies focused on local revisions (Colpitts & Past, 2019; Saeed, Ghazali, Sahuri, et al., 2018). Compared to other sources of feedback, such as automated corrective feedback, OPF also shows more strengths in promoting local-level revision (Shang, 2022). Nevertheless, these studies are all pre-experimental, necessitating caution from researchers when drawing conclusions about treatment effects (DeCarlo, 2018).

1.2 Purpose of the Review

The literature on OPF for EFL writing presents conflicting evidence, reflecting the complexity of this multifaceted approach. This implies the need to adequately understand these varied results by comparing multiple condition levels, isolating online-specific factors, and comprehensively evaluating study quality in terms of research design, sample size, attrition rates, and outcome measure validity. This review addresses these gaps by analyzing varying OPF levels, considering unique online aspects, and rigorously assessing study quality. Its main aim is to comprehensively examine the effectiveness of online peer feedback on EFL students' writing outcomes in higher education, as well as to identify the principles for its effective implementation. This aim, derived from the gaps in the current literature identified in the background section, is operationalized through the following four research questions:

RQ1) What is the evidence of the impact of online peer feedback on EFL students' writing outcomes in higher education?

RQ2) How strong is the evidence of the impact for different approaches to online peer feedback?

RQ3) To what extent do these studies differ in terms of the online peer feedback contextual elements and design elements?

RQ4) What do the evidence and design elements suggest about the principles for effective online peer feedback and its implementation in EFL writing contexts?

2. Method

2.1 Identification of Studies

We conducted a search in December 2021 in the following electronic databases: Web of Science Core Collection, ScienceDirect, EBSCOhost (including ERIC), Scopus, SpringerLink, JSTOR, and Wiley. These databases were searched to ensure that studies in varied education-related fields (e.g. computer science, engineering) were covered, given that online peer feedback could be carried out in a wide range of educational contexts. The following search terms were used: (EFL) AND (Writing) AND ("Peer Feedback" OR "Peer Review" OR "Peer Response" OR "Peer Editing" OR "Peer Interaction") AND (Online OR "Web-Based" OR "Computer-Mediated" OR "Computer-Assisted" OR "Technology-Supported") in the title, abstract, or keywords search fields. Various word combinations and different terms were used to describe the two focus areas: "Peer Feedback" and online. The "AND" operator was used between focus areas, and the "OR" operators were used inside each focus area to gather phrases and words with similar meanings. The search was restricted to peer-reviewed journal articles published in English between 2010 and 2021. Grey literature, such as dissertations, book chapters, and conference proceedings, was excluded due to concerns about the lack of detailed design information, unclear explanations of the online peer feedback intervention, anecdotal reports without data to support the impact or usefulness of the intervention, and the feasibility of analyzing a large volume of reports given the detailed data extraction required for the review. Only studies that use experimental and quasi-experimental designs were included, while surveys, case studies, correlational studies, though useful, were excluded since they could not determine causality due to the uncontrolled confounders (Christensen et al., 2011). The database search was updated using the same method, but the search was narrowed to include only articles published after December 2021. An additional literature search was performed using the Google Scholar database to identify any relevant publications that were not indexed in the previously searched databases. This was done to minimize the risk of publication bias. Furthermore, the reference lists of the studies identified through the database searches were examined to locate any additional pertinent literature.

2.2 Data Cleaning

The initial database search conducted in December 2021 yielded a total of 2,055 studies from the seven databases and Google Scholar. This number was reduced to 1,884 studies after being imported to Endnote 21 and screened for duplicates. The updated database search carried out in February 2024 yielded 897 studies, which was reduced to 553

studies after the removal of duplicates. A total of 14, 900 results were yielded from the initial search using the same search terms in Google Scholar. To avoid an unnecessarily large quantity of screening workload, the references shown on page 1 to page 10, a total of 1000 results (10 results per page), were exported to Endnote, as results shown after page 10 were vaguely relevant. The deduplication process further reduced the results to 464. The large number of duplicates observed is not unexpected, as there is significant overlap in the journals covered by the different databases. To be eligible for inclusion in this review, studies had to meet the following criteria: (a) the study investigated the implementation of web-based peer feedback and its impact, (b) the dependent variable was writing performance, competence, achievements, or improvements, as measured by writing test or task scores, (c) the peer feedback practice was conducted in higher education contexts, (d) the study was empirical and published in a peer-reviewed journal, (e) the study collected data on students' writing result and included a comparison of data before and during/after the web-based peer feedback process, (f) the study applied statistical analysis of the data and reported effect sizes, rather than just percentages or frequencies, and (g) the study was published in the English language. Both quantitative and mixed-methods studies conducted at the higher education level, including postgraduate, were eligible for inclusion. Only articles published from 2010 until February 2024 were included in the analyses for this review.

Studies were excluded for the following reasons: (a) reviews, meta-analyses of peer assessment research, editorial comments/reports, discussions or opinion articles, exploratory analysis, simple surveys, theoretical articles, narrative accounts of researchers' experience, (b) studies that investigate the peer feedback practice implemented in non-computer-assisted environments, (c) studies that focused on ESL learners' or native English learners' peer feedback, (d) studies on the effect of peer feedback on proficiency or performance related to other linguistic skills (e.g., speaking) or skills in other disciplines (e.g., mathematics), (e) studies on students' feedback on teaching, (f) studies on teachers' feedback on their own or fellow teachers' teaching, (g) studies on peer feedback conducted in the workplace or other work-related training activities, (h) studies on peer review between researchers (i.e., journal articles' peer review), (i) studies that investigate peer grading without qualitative peer feedback provided for writing, (j) studies that included peer feedback but did not explicitly investigate its implementation or its impact (e.g., learners' or teachers' perceptions and beliefs about peer assessment), (k) studies that measured the effect of peer feedback on writing anxiety or autonomy instead of writing outcome or performance per se, and (l) studies on the effect of teachers' instruction on peer review skills.

The study selection process involved two phases, as illustrated the PRISMA flow diagram (in Figure 1). Initially, we reviewed titles and abstracts to identify potentially relevant studies. Subsequently, we conducted a thorough examination of the full texts of these preliminarily chosen articles. The primary reviewer (first author) carried out the initial assessment of each record. When uncertainties arose during the full-text review stage, the first and second authors engaged in discussions to jointly determine whether to include or exclude the articles in question. This process resulted in a total of 20 studies and a further citation searching of these studies generated 13 results, of which only 4 met the inclusion criteria. After the two-stage screening process of studies from both databases and citation searching, only 24 studies met the inclusion criteria and were selected for data extraction. The PRISMA flowchart (Page et al., 2021) demonstrates the number of records identified and the number of included and excluded studies at each stage.

2.3 Data Extraction

Our coding scheme consists of three parts: study details, peer assessment design elements, and quality appraisal. To begin with, the following study details were extracted from the 24 articles included in this literature review after ensuring an adequate level of inter-rater reliability was achieved: names of authors, year of publication, type of study (quantitative, qualitative, mixed methods), study aims, study design (pre-experimental, quasi-experimental, experimental), reference group comparison (if an experiment), study variables (dependent and independent variables), sample size and characteristics (e.g., demographics), pre- and post-test, statistical method, and outcome. Table 1 illustrates the main study details of the 24 studies included in this review. In cases where study details were ambiguous or inaccurately reported (e.g., experimental vs. quasi-experimental designs), a minimum of 70% agreement among two trained research assistants as raters was deemed sufficient for each category of the study characteristics as well as peer assessment design elements.

Of the series of typologies compiled to address the difficulties posed by the great diversity of peer assessment (Alqassab et al., 2023; Gielen et al., 2011; Topping, 1998; Van Den Berg et al., 2006), the present review adopted Alqassab et al. (2023)'s coding scheme to extract the related information from the selected articles due to two main reasons. First, their typology drew from Topping's seminal review. Second, their refinement of some design

elements based on contemporary reviews (e.g., Strijbos et al., 2009; van Zundert et al., 2010) achieves more inclusiveness in terms of peer assessment designs. Their coding scheme categorized elements into the groups of Context, Instructional Design, Outcomes, and Moderators/Mediators. Most original elements under these categories were retained for this review, including setting (formal/informal), requirement (compulsory/voluntary), alignment (aligned to program curriculum or not), purpose (summative, formative, or both), output, relation to teacher feedback, official weight, frequency, group constellation, unit of assessment, privacy, contact, scope of involvement, matching, reward, format, training, revision, and moderators/mediators. Additionally, four other elements—technology, writing genre, course setting, and duration—were incorporated for this critical review to conduct a comprehensive evaluation of all the potentially significant factors influencing OPF for EFL writing.

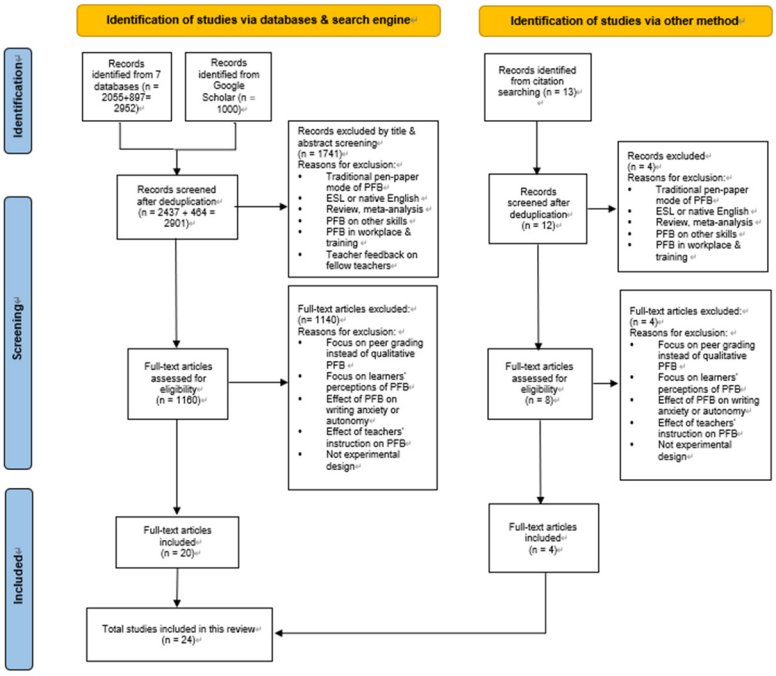

Figure 1. PRISMA Flow Diagram for the Identification, Screening, and Inclusion of Research Studies in This Review (N = 24 studies)

A crucial aspect of the review involved evaluating the quality of each individual study to determine how much confidence could be placed in their findings (Petticrew & Roberts, 2008). This step ensures the trustworthiness of the evidence, distinguishing this review from typical systematic reviews that focus solely on results without differentiating between low-quality and credible evidence. We judged the quality of evidence for the 24 included studies by applying the "Sieve" quality assessment tool (Gorard et al., 2017). Using a four-star rating system with 4 being the most robust and reliable, each study was evaluated from left to right and top to bottom based on six criteria proposed by Gorard et al. (i.e., research design, sample size, attrition rate, outcome measurement methods, fidelity, and threats to validity). For example, randomized controlled trials would receive 4 stars at the beginning but be reduced to 2 stars if they had a small sample size. Similarly, small studies involving randomization of classes to treatment or control groups receive lower ratings as well. Since the two classes may differ on student and teacher characteristics, any observed impact might be affected by these differences rather than the sole result of the intervention (Shadish et al., 2002). In the case of randomizing intact classes, each class is regarded as a case and a large number of clusters is required. To ensure inter-rater reliability, two raters independently rated the studies, and any disparities were resolved after they were discussed to reach agreement (Cofie et al., 2022).

A large number of studies included in this review did not report participant attrition rates, and one study lacked sufficient data to calculate effect sizes. Incomplete or inadequate reporting practices reflect poor research quality, leading to low ratings for those studies. None of the reviewed studies calculated effect sizes, making it difficult to determine the presence or magnitude of positive impacts. Therefore, effect sizes were calculated using the available data from the original papers by conducting between-group comparisons. Effect size is crucial for quantifying

intervention effectiveness, emphasizing the magnitude of effects rather than statistical significance, and promoting a more scientific approach to knowledge accumulation (Coe, 2002). Specifically, Hedges' g effect size was employed here, calculated by taking the difference between the experimental and control groups' post-test means and dividing by the pooled standard deviation. After rating each study's quality, the ratings were presented in Table 2.

Table 1. An Overview of the Study Details

Study details	Category	Frequency	%
Study Type	Quantitative	15	62.5
	Mixed methods	9	37.5
Study design	Experimental	8	33.3
	Quasi-experimental	16	66.7
Reference group comparison	OPF vs. FFPF	12	50
	OPF vs. FFTF	3	12.5
	OPF vs. NF	3	12.5
	SF vs. OPF vs. AF vs. OPF + AF	1	4
	SPF vs. UPFB	1	4
	SPF vs. USPF vs. GPF	1	4
	OPF vs. OTF vs. FFTF	1	4
	Different sequences of SF, FFPF, OPF	1	4
	ASPF + SD vs. ASPF	1	4
Sample size	Under 15	0	0
	15 - 49	10	41.6
	50 - 99	10	41.6
	100+	4	16.7
Education level	Preparatory	1	4
	Undergraduate	15	62.5
	Postgraduate	1	4
	Undergraduate & Postgraduate	1	4
	Language school students	5	20.8
	Details not provided	1	4

(In the column of category for reference group comparison, OPF = online peer feedback, OTF = online teacher feedback, FFPF = face-to-face peer feedback, FFTF = face-to-face teacher feedback, NF = no feedback, SF = self-feedback, AF = automated feedback, SPF = scripted peer feedback, USPF = unscripted peer feedback, GPF = guided peer feedback, ASPF = asynchronous peer feedback, SD = synchronous discussion)

Table 2. An Overview of the Quality Ratings

Author(s) + Year	Scale	Attrition	Effect size	Threats	Rating
Acarol (2024)	S	Not reported	g = 1	Unblinded raters	0*
AbuSeileek (2013)	S	Not reported	g = 3.03 (TC) g = 1.42 (WP) g = 2.45 (TC+WP)	Possible diffusion of treatment	1*
AbuSeileek & Abualshar' (2014)	S	Some attrition	g = 6.16 (TC) g = 4.16 (RC) g = 1.71 (ML)	Possible diffusion of treatment	1*
Arslan (2014)	S	Not reported	g = 1.25	Possible diffusion of treatment	0*
Awada & Diab (2021）	S	7 out of 129	g = 1.02	Possible teacher effect	1.5*

		(5%)			
Ciftci & Kocoglu (2012)	S	Not reported	g = 0.96	Possible diffusion of treatment	0*
Ebadi & Rahimi (2017)	S	Not reported	g = 1.39	Moderate assurance of quality	Low 1*
Ebadi & Alizadeh (2021）	S	Not reported	g = 6.10	Moderate assurance of quality	Low 1*
Ghadi & Khodabakhshzadeh (2016)	S	Not reported	g = 1.05	Teacher effect	1.5*
Ho, et al. (2020)	S	Not reported	g = 0.597	Moderate assurance of quality	1*
Kioumarsi et al. (2018)	VS	Not reported	g = 5.236	Teacher effect	0*
Latifi & Noroozi (2021)	S	Not reported	g = 2.604	Unblinded raters	1.5*
Latifi et al. (2021)	S	Not reported	SPF vs USPF g = 3.27	Unblinded raters	1*
Li (2023)	S	Not reported	g = 0.525	Moderate assurance of quality	1.5*
Ly (2023)	S	Not reported	g = 0.95	Unblinded raters	Low 1*
Motallebzadeh & Amirabadi (2011)	S	Not reported	OPF vs FFTF g = 1.40 OPF vs OTF g = 0.81	Unclear allocation method	1.5*
Sayed (2010)	VS	7	g = 1.75	Experimental group had more chance for revision	0*
Tan et al. (2023)	S	Not reported	OPF vs SF g = 0.55 OPF+AF vs SF g = 0.75	Same topic in pre & post-test; unblinded raters	1.5*
Vahedipour & Rezvani (2017)	S	Not reported	g = 1.41	Unclear allocation method	0*
Wanchid (2015)	S	Not reported	g = 0.92	Possible diffusion of treatment	1.5*
Wang (2015)	S	Not reported	g = 1.15	Possible diffusion of treatment; unblinded raters	1.5*
Yang (2015)	VS	Not reported	g = 3.4	Unblinded raters	1*
Zhang & McEneaney (2019)	M	Not reported	g = 0.45	Teacher effect	2*
Zheng et al. (2018)	S	Not reported	g = 0.583	Unclear allocation & rating method	Low 1*

(In the column of scale, S = small, VS = very small, M= medium; In the column of effect size, TC = Track Changes, WP = Word Processor, Recast = RC, Metalinguistic = ML. See notes under Table 1 for the full names for the other abbreviations of multiple comparison groups)

3. Results

RQ1) What is the evidence of impact of OPF on EFL students' writing outcomes in higher education?

An overview of the characteristics of the 24 eligible studies is provided in Table 1. Table 2 provides an overview of the quality ratings and the criteria used to determine these. Following this, we go on to present a narrative synthesis

of these studies. The studies reviewed provide considerable evidence that OPF interventions have a positive impact on EFL students' writing outcomes in higher education contexts. All of the 24 studies examined reported statistically significant positive effects of OPF compared to control conditions, such as traditional face-to-face peer feedback, teacher feedback, or self-feedback alone, on overall writing performance. The effect sizes ranged from small to extremely large, with 9 studies reporting a large or very large effect size ($g > 0.8$) and 5 studies reporting an extremely large effect size ($g > 2.0$) favouring OPF groups over FFPF groups. For example, extremely large effects were found in Ebadi and Alizadeh (2021) with $g = 6.10$ and Kioumarsi et al. (2018) with $g = 5.236$, AbuSeileek and Abualsha'r (2014) with $g = 6.16$ for track changes OPF and $g = 4.16$ for recast OPF, and Latifi and Noroozi (2021) with $g = 2.604$ for scripted OPF. Large or very large effects were reported in studies, such as Awada and Diab (2021) with $g = 1.02$, Ghadi and Khodabakhshzadeh (2016) with $g = 1.05$, Ciftci and Kocoglu (2012) with $g = 0.96$, Vahedipour and Rezvani (2017) with $g = 1.41$, and Sayed (2010) with $g = 1.75$. A few studies reported small to medium effects, such as Li (2023) with $g = 0.525$ for OPF vs. FFPF, Motallebzadeh and Amirabadi (2011) with $g = 0.59$ for OPF vs. FFTF, and Tan et al. (2023) with $g = 0.55$ for OPF compared with SF.

Regarding the OPF effects on specific aspects of writing development, some studies have shown that students achieve more gains in all the examined aspects of writing performance through OPF than FFPF. For example, Ebadi and Rahimi (2017) showed that the OPF group of Iranian adult students outperformed the FFPF group in all four dimensions of academic writing (i.e., task achievement, coherence and cohesion, lexicon, and grammatical accuracy). Yang (2015) also found that Taiwanese graduate students in the OPF group made more revision related to both local and global issues than the paper-based FFPF group. Comparatively, Wang (2015) observed that FFPF resulted in a significantly higher number of comments addressing global issues, while computer-mediated feedback produced more comments focused on local concerns. Conversely, other studies have failed to find measurable impacts, particularly for lower-order writing dimensions like mechanics (e.g., Zheng et al., 2018). Investigating the effect of OPF in comparison with feedback of other sources, or their combination, Tan et al. (2023) reported that the combination of automated written corrective feedback (AWCF) and asynchronous computer-mediated peer feedback (ACMC) produced the highest impact on writing based on mostly lexical and syntactic measures (i.e., complexity, accuracy, and fluency) among the three treatment groups (AWCF, ACMC, AWCF+ACMC) and the control group (SF).

While the studies reviewed consistently report positive effects of online peer feedback interventions on EFL students' writing outcomes in higher education, the reliability and validity of these findings must be carefully evaluated. Therefore, a thorough examination of the study quality, with the research design, sample size, attrition, data quality, and potential biases of the reviewed studies considered, has been carried out to determine the robustness of the evidence. Addressing RQ2, the result of this examination is provided in the next section.

RQ2) How strong is the evidence of impact for different OPF approaches?

The evidence suggests that different configurations and approaches to delivering online peer feedback can lead to varying degrees of impact on writing outcomes (refer to Quality Appraisal in the supplementary materials for details). Overall, while the evidence base remains limited, the findings suggest that carefully structured, technology-mediated OPF approaches can have a substantial positive impact on EFL students' writing skills in higher education contexts. Unstructured approaches appear less effective, though still showing small-to-medium positive effects in some cases.

Of the 24 eligible studies, all reported positive effects of online peer feedback on EFL writing in higher education. However, the quality of these studies varied considerably, with ratings ranging from 2* to 0* based on our quality appraisal criteria. This review focuses primarily on studies rated 2*, 1.5*, 1*, and low 1*, while excluding 0-rated studies due to their limited relevance to the research questions. The highest-rated study in this review, Zhang and McEneaney (2020), received a 2* rating, followed by 1.5* rating studies ($n = 8$), 1* rating studies ($n = 5$), low 1* rating studies ($n = 4$), and 0* rating studies ($n = 6$). This quasi-experimental study involved 198 sophomore English majors from 8 intact classes in an EFL context and compared OPF with traditional FFTF in argumentative essay writing. Using cluster randomization, 4 classes were assigned to each group. The study employed a pretest-posttest design, with both tests consisting of a 45-minute, 200-word argumentative essay task. Three blind raters assessed the essays using a predefined rubric. The OPF group showed statistically significant improvements in writing performance compared to the FFTF group, with a medium effect size ($g = 0.45$). While the study demonstrates several strengths, including a larger sample size, blind rating, and grounding in multiple theoretical frameworks, it has limitations such as potential teacher effects, cluster effects, and unreported attrition rates. The use of non-equivalent groups in the quasi-experimental design may introduce selection bias. Despite these limitations, the study provides evidence for the effectiveness of online peer feedback in improving EFL students' writing

performance in higher education contexts.

The evidence from the 1.5* studies for online peer feedback (OPF) as a strategy to improve EFL students' writing skills is moderate. Eight studies evaluated the use of OPF in higher education EFL writing contexts, all of which showed positive effects favoring OPF, with effect sizes ranging from medium to very large. However, all studies were rated 1.5* due to various methodological limitations. All of these studies employed quasi-experimental designs with pre-test and post-test measures, comparing OPF with traditional FtFPF (e.g., Wang, 2015) or different OPF approaches (e.g., Latifi & Noroozi, 2021). Some studies employed more complex designs, comparing multiple feedback conditions or sequences (e.g., Tan et al., 2023; Wanchid, 2015). Compared to Zhang and McEneaney (2020), their sample sizes were relatively small, ranging from 42 to 122 participants. Most studies focused on university-level EFL learners, with a few targeting language school students. Common strengths across these studies included the use of multiple assessment points (Awada & Diab, 2021), examination of various aspects of writing skills (e.g., Li, 2023; Tan et al., 2023), and grounding in relevant theoretical frameworks (e.g., Latifi & Noroozi, 2021; Motallebzadeh & Amirabadi, 2011). Some studies used standardized tests or blinded rating, enhancing outcome quality (e.g., Ghadi & Khodabakhshzadeh, 2016; Li, 2023). Shared limitations included small sample sizes, lack of true randomization, potential teacher effects, and often unreported attrition rates. Some studies used researcher-developed assessment tools, which may affect result reliability (Ghadi & Khodabakhshzadeh, 2016). Effect sizes reported in these studies ranged from medium to very large ($g = 0.525$ to 2.604), consistently favoring OPF over traditional methods or showing benefits of structured OPF approaches. However, these large effect sizes should be interpreted cautiously due to some aforementioned methodological limitations.

Nine studies were rated 1* or low 1*, indicating weak and mixed evidence, with most showing positive effects but suffering from significant methodological limitations. Among the 1* studies, AbuSeileek (2013) and AbuSeileek and Abualsha'r (2014) investigated computer-mediated corrective feedback using MS Word tools. Both studies reported large to extremely large effect sizes favoring the experimental groups, particularly for track changes feedback. However, these studies had small sample sizes ($n = 16$ per group) and potential diffusion of treatment effects. Ho et al. (2020) and Latifi et al. (2021) examined Facebook-based and scripted online peer feedback respectively, both reporting medium to large effect sizes. However, these studies were limited by small sample sizes ($n = 32$ and $n = 16$ per group), lack of randomization, and unclear rating procedures. Yang (2015) and Ly (2023) focused on OPF in academic writing contexts, reporting large effect sizes. However, both studies suffered from very small or small sample sizes ($n = 11$ and $n = 32$ per group) and potential validity issues related to rating process. The low 1* studies by Ebadi and Rahimi (2017) and Ebadi and Alizadeh (2021) investigated learner-driven online peer editing using Google Docs. Both reported significant improvements in writing skills for the experimental groups but had very small or small sample sizes ($n = 10$ and $n = 20$ per group) and potential diffusion of treatment effects. Zheng et al. (2018) examined synchronous online discussion in peer assessment, reporting significant improvements in writing performance. However, the study had a small sample size ($n = 32$ per group) and unclear randomization procedures. Common limitations across these studies include small sample sizes (ranging from 16 to 64 participants), lack of true randomization, unreported attrition rates, and potential diffusion of treatment effects, and other issues with validity, such as convenience sampling and unclear blinding procedures for raters.

In conclusion, the variety of OPF approaches and contexts studied provides a broad view of OPF applications but limits direct comparability across studies. While these studies generally suggest positive effects of OPF on EFL writing performance, the evidence is weak due to the following methodological limitations: (1) Small sample sizes. The majority of studies had fewer than 30 participants per condition, limiting their statistical power and generalizability. (2) Lack of randomization. Many studies used intact classes or convenience sampling, potentially introducing selection bias. (3) Short intervention durations. Several studies had brief interventions, which may not reflect long-term effects of online peer feedback. (4) Potential diffusion of treatment. In some studies, all groups were taught by the same instructor, possibly leading to contamination between conditions. (5) Unclear or unreported attrition. Many studies did not report dropout rates, which could affect the validity of their results. Given the small samples and potential biases, the consistently large, some extremely large effect sizes reported, such as $g = 6.16$ in AbuSeileek and Abualsha'r (2014), raise concerns about their reliability and should be interpreted cautiously.

RQ3) To what extent do these studies differ in terms of the online peer feedback contextual elements and design elements?

In terms of the settings where OPF interventions were conducted for EFL writing, 8 studies were contextualized in Iran, the largest number among the 24 studies reviewed, followed by China. The studies predominantly utilized various online platforms and technologies for peer feedback, including popular tools like Google Docs, Wikis, blogs,

online forum, and custom-built systems, which fall into six different categories according to the taxonomy of Golonka et al. (2014). The writing genres covered were diverse, with essay writing being the most common (including general, narrative, and argumentative types), followed by paragraph writing and specific formats like business letters or IELTS tasks. Nearly all studies were conducted in formal educational settings and aligned with curriculum objectives. Figure 2 details the number of the studies that implement OPF in a specific context and its relative size in that contextual category.

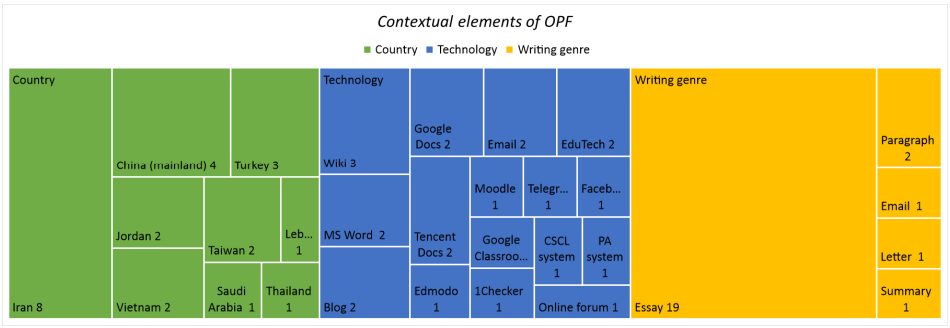

Figure 2. Distribution of Studies by Contextual Elements of OPF in EFL Writing

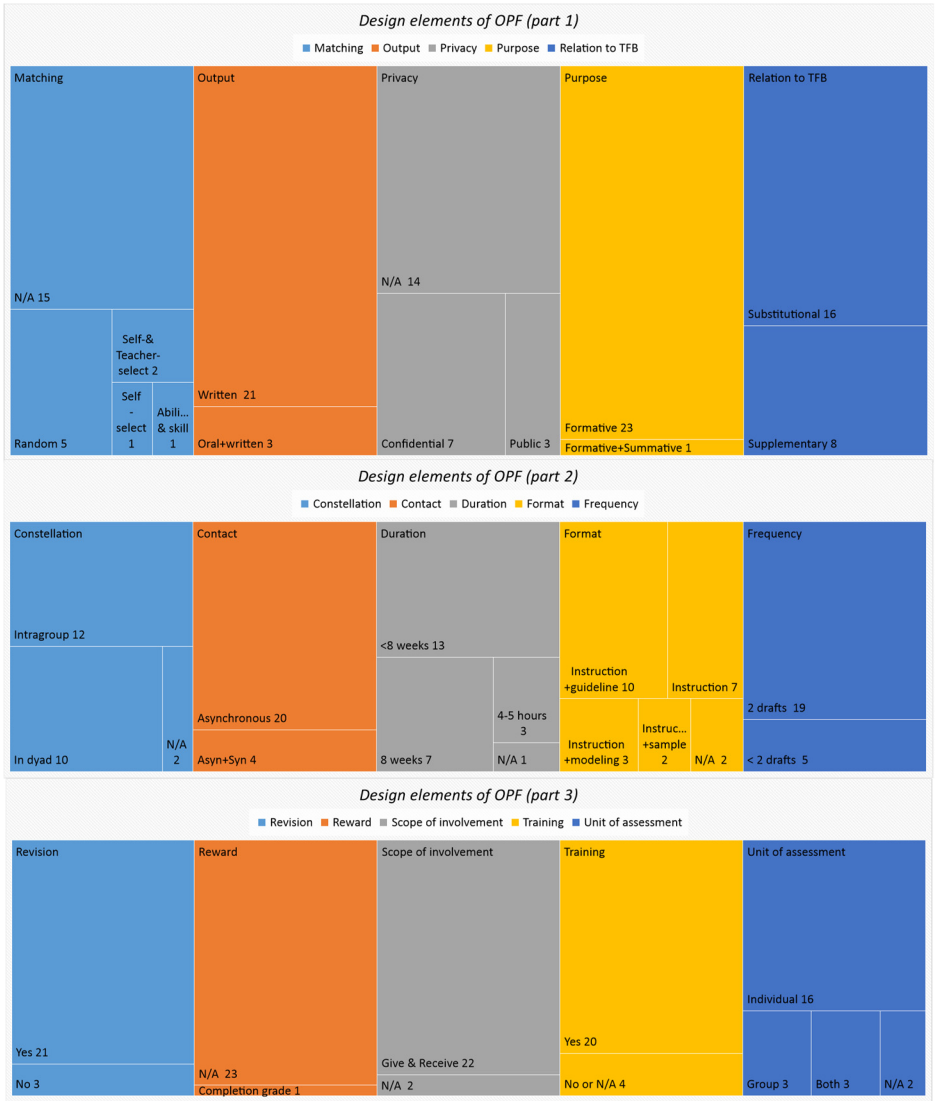

Figure 3. Distribution of Studies by Design Elements of OPF in EFL Writing

As regards the design elements of online peer feedback in EFL writing contexts, while there are some common trends, such as the prevalence of written feedback, two-draft processes, and the inclusion of training, there is significant variation in how these elements are implemented across studies. This diversity reflects the adaptability of peer feedback to different educational contexts and objectives, as well as the ongoing exploration of optimal design features for enhancing EFL writing skills through online peer feedback. The purpose of OPF was predominantly formative, with one study using it for both formative and summative purposes. The output was predominantly written comments, with a few studies involving a combination of oral and written comments. The relation to teacher feedback varied, with some studies using peer feedback as a substitute for teacher feedback, while others used it as a supplement. The frequency of OPF implementation varied, with most studies using a two-draft process, and a few including multiple drafts or cycles of feedback. The constellation of peer feedback was typically organized in dyads or small intragroup settings, with one study utilizing a large group configuration. The unit of assessment was most commonly individual feedback, though some studies implemented group assessment processes or followed individual feedback with group discussion. Privacy approaches varied, ranging from confidential to public sharing of feedback. Contact was predominantly asynchronous, with a few studies including synchronous elements or face-to-face discussions. The scope of involvement was broad, with most studies having participants both give and receive feedback. The feedback format was almost always guided, with most studies providing guided instruction accompanied by rubrics, prompts, or specific guidelines. Training was included in most studies, ranging from brief introductory sessions to extensive multi-week training or semester-long preparations. Revision stages following peer feedback were incorporated in almost all studies. The duration of OPF interventions varied considerably, from single sessions to semester-long interventions. Figure 3 shows the number of the studies that feature each design element subcategory of OPF and its relative size in a category.

RQ4) What do the evidence and design elements suggest about the principles for effective online peer feedback and its implementation in EFL writing contexts?

To address this question, we employed a rigorous thematic analysis approach. This process involved thoroughly reviewing all included studies, with a focus on reported effective practices, implementation strategies, and outcomes. We then systematically coded relevant information related to effective online peer feedback practices across all studies. These codes were subsequently grouped into initial themes, which were reviewed and refined to ensure distinctiveness and coherence. Finally, we clearly defined each theme and named it as a principle for effective online peer feedback, cross-checking against the original studies to ensure accurate representation of the data. Through this comprehensive analytical process, seven key principles emerged for effective online peer feedback and its implementation in EFL writing contexts. It is important to note that these principles are derived from the design elements and study details reported in the provided documents and should be considered in conjunction with the specific contexts, objectives, and constraints of individual EFL writing instruction scenarios.

(1) Adopt a formative approach. When used to support learning and skill development rather than for summative assessment, implementing OPF for formative purposes is most effective.

(2) Provide structured guidance. Incorporating structured guidance in OPF, including guided instruction, peer-editing forms, worksheets, rubrics, or prompts, is highly important to ensure meaningful and effective feedback.

(3) Offer training. Providing training for online peer feedback is of great significance as it equips students with the necessary skills and knowledge for giving and receiving feedback effectively.

(4) Facilitate revision opportunities. Allowing for revision based on the feedback received provides opportunities for incorporating feedback and revising work, thus it is crucial for the improvement of learning results.

(5) Leverage technology wisely. The choice of technology should align with the specific needs, context, and affordances required for effective online peer feedback.

(6) Encourage asynchronous interactions. Utilizing asynchronous online peer feedback might be more suitable for EFL students to foster reflective and thoughtful feedback exchange.

(7) Utilize a combination of feedback sources and modalities. Adopting a balanced approach by combining additional feedback sources (e.g., teacher feedback, self-feedback) with OPF, or oral and written feedback may be beneficial for learning and performance.

4. Discussion

This systematic review critically examined the effectiveness of Online Peer Feedback (OPF) approaches for

enhancing EFL writing skills in higher education contexts. By synthesizing evidence from 24 eligible studies, we addressed gaps in previous research regarding the impact of OPF on EFL writing outcomes, the strength of evidence for different OPF approaches, and the variability in contextual and design elements across studies. Our findings provide insights into the potential benefits and limitations of OPF for EFL writing instruction, as well as principles for its effective implementation.

Our results suggest that OPF interventions generally have a positive impact on EFL students' writing outcomes in higher education settings. All 24 studies reported statistically significant positive effects of OPF compared to control conditions, with effect sizes ranging from small to extremely large. This consistent trend indicates that OPF can effectively improve various aspects of EFL writing performance, including overall quality, specific writing dimensions (e.g., content, organization, language, and mechanics), and revision behaviors, though findings remain inconclusive on OPF having an advantage in improving lower-level dimensions of writing. The effectiveness of OPF aligns with theoretical perspectives on collaborative learning and the writing process. The online environment appears to facilitate meaningful peer interactions, cognitive engagement, and opportunities for revision, which are key components of effective writing development (Rouhi et al., 2020; Villamil & de Guerrero, 2019). The digital medium may also provide affordances that enhance the feedback process, such as increased accessibility, flexibility in timing, and the potential for more detailed and reflective comments (Breuch, 2004).

While the substantial effect sizes underscore the potential of well-designed OPF approaches, they also point to the influence of specific design elements and configurations in moderating the impact of OPF interventions. It is important to note that the strength of evidence varies considerably across studies. According to our quality assessment, the highest-rated study (Zhang & McEneaney, 2020) is of medium quality (2*), with the majority of studies rated as 1.5* or lower due to methodological limitations. This variability in study quality highlights the need for caution when interpreting the reported effect sizes, particularly those in the extremely large range (e.g., $g > 4.0$). Our findings both support and diverge from previous reviews. We corroborate the observations of Chang (2016), Chen (2016), and Saeed, Ghazali and Aljaberi (2018) regarding the overall positive impact of technology-mediated peer feedback on L2 writing. However, unlike some earlier reviews that found mixed or inconclusive evidence (Yu & Lee, 2016), our synthesis suggests a more consistently positive trend. This difference may be attributed to advancements in OPF technologies and pedagogical approaches over time, as well as our focus on experimental and quasi-experimental studies.

The variability in effect sizes across studies can be explained by differences in OPF design elements and implementation contexts. Studies employing structured guidance, comprehensive training, and multiple revision opportunities tended to report larger effect sizes (Awada & Diab, 2021). This observation aligns with Rotsaert et al. (2018)'s emphasis on the importance of scaffolding in peer assessment processes. Our analysis revealed significant diversity in OPF implementations across studies, demonstrating the flexibility of OPF approaches while presenting challenges in directly comparing outcomes. When compared to the broader peer assessment landscape described by Alqassab et al. (2023), we found both similarities and differences, including the predominant use of OPF for formative purposes and the prevalence of written assignments as the object of peer assessment. Key design elements in our review align with two core themes identified by Alqassab et al., namely implementation decisions (providing structured guidance, incorporating training, and allowing for revision based on received feedback) and assessment constellation and peer interactions (employing intragroup or dyad configurations and utilizing asynchronous OPF). These alignments suggest that effective OPF practices in EFL writing contexts share commonalities with broader peer assessment approaches while adapting to the specific needs of language learners. The emphasis on scaffolding may be particularly crucial in EFL contexts, where students are simultaneously developing language skills and learning to provide constructive feedback (Min, 2006, 2016).

Our review highlighted the diverse range of technologies employed for OPF, including classroom-based technologies, social computing platforms, and cloud-based word processors. This variety reflects the growing integration of technology in educational settings and necessitates flexible OPF implementations that can adapt to different technological contexts. A notable finding is the utilization of combined feedback modalities, such as automated corrective feedback with technology-mediated peer feedback (Tan et al., 2023), both oral and written peer feedback supported by technology (Zhang & McEneaney, 2020), and multiple feedback sources, including teacher feedback and self-feedback (Motallebzadeh & Amirabadi, 2011). This aligns with the notion of a balanced approach to feedback, where learners can benefit from diverse perspectives and sources of information to support their writing development (Hyland & Hyland, 2019).

5. Conclusion and Limitations

This critical review of 24 empirical studies demonstrates the positive impact of online peer feedback (OPF) on EFL students' writing skills in higher education, complementing Alqassab et al. (2023)'s findings. Our analysis reveals improvements in both global and local aspects of writing, highlighting key principles for effective OPF implementation: adopting a formative approach, providing structured guidance, incorporating comprehensive training, facilitating multiple revision opportunities, leveraging technology for asynchronous interactions, and integrating diverse feedback sources and modalities. These findings suggest that OPF can create dynamic writing ecosystems that enhance linguistic skills while fostering critical thinking, metacognitive awareness, and digital literacy.

Our review also underscores crucial areas for future research and practice. There is a pressing need for more rigorous, large-scale investigations with stringent experimental designs to address current methodological limitations and advance understanding of OPF effectiveness and its underlying mechanisms. Future studies should explore diverse OPF configurations, investigate long-term effects on EFL writing development, and examine the integration of emerging technologies like AI and VR in multicultural learning environments. Practically, our findings emphasize the importance of ongoing professional development for educators to effectively leverage OPF technologies (Luo et al., 2024). From a policy perspective, OPF integration presents a potentially cost-effective strategy for improving writing outcomes at scale, though it may require investment in technological infrastructure and faculty training (Mulyadi et al., 2024; Xu et al., 2024). By addressing these research gaps and leveraging our insights, we can develop more effective, evidence-based strategies for OPF implementation, working towards more inclusive and technologically enhanced writing education for EFL learners.

While this systematic review provides valuable insights into the effectiveness of online peer feedback for EFL writing in higher education, it is important to acknowledge several limitations. First and foremost, merely 24 studies of experimental or quasi-experimental design were eligible for review. The scarcity of such studies, despite repeated calls for more rigorous research designs (Strijbos et al., 2009; Topping, 1998; van Zundert et al., 2010), suggests a persistent gap in the literature. This limitation may affect the generalizability of our findings and highlights the need for more experimental research in the field of online peer feedback for EFL writing. Next, the review relied heavily on effect size calculations based on the available data reported in the original studies. As none of the studies provided effect size estimates, the review team had to calculate effect sizes using reported means and standard deviations. This process may have introduced inaccuracies or inconsistencies in the effect size calculations, potentially affecting the interpretation of the findings. Lastly, the review could not comprehensively investigate the influence of specific design elements or combinations thereof due to the heterogeneity of the studies and the lack of detailed reporting on those elements in which the included studies varied considerably.

References

AbuSeileek, A. F. (2013). Using track changes and word processor to provide corrective feedback to learners in writing. *Journal of Computer Assisted Learning*, *29*(4), 319-333. https://doi.org/10.1111/jcal.12004

AbuSeileek, A., & Abualsha'r, A. (2014). Using peer computer-mediated corrective feedback to support EFL learners' writing. *Language Learning & Technology*, *18*(1), 76-95. Retrieved from http://llt.msu.edu/issues/february2014/abuseileekabualshar.pdf

Alqassab, M., Strijbos, J. W., Panadero, E., Ruiz, J. F., Warrens, M., & To, J. (2023). A Systematic Review of Peer Assessment Design Elements. *Educational Psychology Review*, *35*(1), 18. https://doi.org/10.1007/s10648-023-09723-7

Awada, G. M., & Diab, N. M. (2021). Effect of online peer review versus face-to-Face peer review on argumentative writing achievement of EFL learners. *Computer Assisted Language Learning*, *36*(1-2), 238-256. https://doi.org/10.1080/09588221.2021.1912104

Braine, G. (2001). A study of English as a foreign language (EFL) writers on a local-area network (LAN) and in traditional classes. *Computers and Composition*, *18*(3), 275-292. https://doi.org/10.1016/S8755-4615(01)00056-1

Breuch, L. A. K. (2004). *Virtual Peer Review: Teaching and Learning about Writing in Online Environments*. State University of New York Press.

Cao, S., Zhou, S., Luo, Y., Wang, T., Zhou, T., & Xu, Y. (2022). A review of the ESL/EFL learners' gains from

online peer feedback on English writing. *Frontiers in Psychology*, *13*, Article 13:1035803. https://doi.org/10.3389/fpsyg.2022.1035803

Chang, C. Y. H. (2016). Two Decades of Research in L2 Peer Review. *Journal of Writing Research*, *8*(1), 81-117. https://doi.org/10.17239/jowr-2016.08.01.03

Chen, T. (2016). Technology-supported peer feedback in ESL/EFL writing classes: A research synthesis. *Computer Assisted Language Learning*, *29*(2), 365-397. https://doi.org/10.1080/09588221.2014.960942

Chenoweth, N. A., & Hayes, J. R. (2001). Fluency in writing: Generating text in L1 and L2. *Written Communication*, *18*(1), 80-98. https://doi.org/10.1177/0741088301018001004

Christensen, L. B., Johnson, R. B., & A., T. L. (2011). *Research Methods, Design, and Analysis* (12th ed.). Pearson.

Ciftci, H., & Kocoglu, Z. (2012). Effects of Peer E-Feedback on Turkish EFL Students' Writing Performance. *Journal of Educational Computing Research*, *46*(1), 61-84. https://doi.org/10.2190/EC.46.1.c

Coe, R. (2002, September 12-14). *It's the effect size, stupid* [Paper presentation]. British Educational Research Association Annual Conference, Exeter, United Kindom.

Cofie, N., Braund, H., & Dalgarno, N. (2022). Eight ways to get a grip on intercoder reliability using qualitative-based measures. *Canadian Medical Education Journal*, *13*(2), 73-76. https://doi.org/10.36834/cmej.72504

Colpitts, B. D., & Past, T. H. (2019). A Comparison of Computer-Mediated Peer Corrective Feedback between High and Low-Proficiency Learners in a Japanese EFL Writing Classroom. *JALT CALL Journal*, *15*(2), 23-39. https://eric.ed.gov/?id=EJ1232877

Cuocci, S., Fattahi Marnani, P., Khan, I., & Roberts, S. (2023). A Meta-Synthesis of Technology-Supported Peer Feedback in ESL/EFL Writing Classes Research: A Replication of Chen's Study. *Languages*, *8*(2), 114. https://doi.org/10.3390/languages8020114

DeCarlo, M. (2018). *Scientific inquiry in social work*. Open Social Work Education.

Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education*, *24*(3), 331-350. https://doi.org/https://doi.org/10.1080/03075079912331379935

Ebadi, S., & Alizadeh, A. (2021). The effects of online learner-driven feedback on IELTS writing skills via Google Docs. *Teaching English with Technology*, *21*(3), 42-66. Retrieved from https://www.ceeol.com/search/article-detail?id=1063456

Ebadi, S., & Rahimi, M. (2017). Exploring the impact of online peer-editing using Google Docs on EFL learners' academic writing skills: a mixed methods study. *Computer Assisted Language Learning*, *30*(8), 787-815. https://doi.org/10.1080/09588221.2017.1363056

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

Flower, L., & Hayes, J. R. (1981). A Cognitive Process Theory of Writing. *College Composition & Communication*, *32*(4), 365-387. https://doi.org/10.58680/ccc198115885

Ghadi, S. A., & Khodabakhshzadeh, H. (2016). The effect of employing electronic peer assessment on Iranian EFL learners' writing ability and autonomy. *Theory and Practice in Language Studies*, *6*(12), 2272. https://doi.org/10.17507/tpls.0612.06

Gielen, S., Dochy, F., & Onghena, P. (2011). An inventory of peer assessment diversity. *Assessment & Evaluation in Higher Education*, *36*(2), 137-155. https://doi.org/10.1080/02602930903221444

Golonka, E. M., Bowles, A. R., Frank, V. M., Richardson, D. L., & Freynik, S. (2014). Technologies for foreign language learning: A review of technology types and their effectiveness. *Computer Assisted Language Learning*, *27*(1), 70-105. https://doi.org/10.1080/09588221.2012.700315

Gorard, S., See, B. H., & Siddiqui, N. (2017). *The trials of evidence-based education: The promises, opportunities and problems of trials in education*. Routledge.

Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications.* (pp. 1-27).

Lawrence Erlbaum Associates, Inc. https://doi.org/10.4324/9780203811122

Ho, P. V. P., Phung, L. T. K., Oanh, T. T. T., & Giao, N. Q. (2020). Should Peer E-Comments Replace Traditional Peer Comments? *International Journal of Instruction*, *13*(1), 295-314. https://doi.org/10.29333/iji.2020.13120a

Hyland, K., & Hyland, F. (2019). Contexts and Issues in Feedback on L2 Writing. In F. Hyland & K. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (2 ed., pp. 1-22). Cambridge University Press. https://doi.org/10.1017/9781108635547.003

Kayacan, A., & Razı, S. (2017). Digital self-review and anonymous peer feedback in Turkish high school EFL writing. *Journal of Language and Linguistic Studies*, *13*(2), 561-577. https://dergipark.org.tr/en/pub/jlls/issue/36120/405630

Latifi, S., & Noroozi, O. (2021). Supporting argumentative essay writing through an online supported peer-review script. *Innovations in Education and Teaching International*, *58*(5), 501-511. https://doi.org/10.1080/14703297.2021.1961097

Latifi, S., Noroozi, O., Hatami, J., & Biemans, H. J. A. (2021). How does online peer feedback improve argumentative essay writing and learning? *Innovations in Education and Teaching International*, *58*(2), 195-206. https://doi.org/10.1080/14703297.2019.1687005

Li, Y. (2023). The effect of online collaborative writing instruction on enhancing writing performance, writing motivation, and writing self-efficacy of Chinese EFL learners. *Frontiers in Psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1165221

Liu, J., & Sadler, R. W. (2003). The effect and affect of peer review in electronic versus traditional modes on L2 writing. *Journal of English for Academic Purposes*, *2*(3), 193-227. https://doi.org/10.1016/S1475-1585(03)00025-0

Liu, N. F., & Carless, D. (2006). Peer feedback: the learning element of peer assessment. *Teaching in Higher Education*, *11*(3), 279-290. https://doi.org/10.1080/13562510600680582

Luo, J., Singh, C. K. S., Ong, E. T., Li, Y., Wang, Z., Guo, Z., & Qian, P. (2024). Effects of Curriculum Alignment, Professional Development, and Access and Resources on Teaching Strategies for Museum Education. *Asian Journal of University Education*, *20*(2), 248-272. https://doi.org/10.24191/ajue.v20i2.17168

Lv, X., Ren, W., & Xie, Y. (2021). The Effects of Online Feedback on ESL/EFL Writing: A Meta-Analysis. *The Asia-Pacific Education Researcher*, *30*(6), 643-653. https://doi.org/10.1007/s40299-021-00594-6

Ly, N. N. (2023). Applying online peer feedback via a peer review application in English academic writing. *HCMCOUJS-Social Sciences*, *13*(1), 89-105. https://doi.org/10.46223/HCMCOUJS.soci.en.13.1.2300.2023

Mellati, M., & Khademi, M. (2014). Peer evaluation in CMC learning environment and writing skill. *International Journal of Applied Linguistics and English Literature*, *3*(5), 220-228. https://doi.org/10.7575/aiac.ijalel.v.3n.5p.220

Min, H. T. (2006). The effects of trained peer review on EFL students' revision types and writing quality. *Journal of Second Language Writing*, *15*(2), 118-141. https://doi.org/10.1016/j.jslw.2006.01.003

Min, H. T. (2016). Effect of teacher modeling and feedback on EFL students' peer review skills in peer review training. *Journal of Second Language Writing*, *31*, 43-57. https://doi.org/10.1016/j.jslw.2016.01.004

Motallebzadeh, K., & Amirabadi, S. (2011). Online interactional feedback in second language writing: Through peer or tutor. *Theory and Practice in Language Studies*, *1*(5), 534-540. https://doi.org/10.4304/tpls.1.5.534-540

Mulyadi, D., Arifani, Y., Abdurrosyid, A., Mufaridah, F., & Singh, C. K. S. (2024). Fulfillment of Basic Psychological Needs of English for Specific Purpose Learners during Online Learning: Lecturers' and Learners' Perspectives. *Language Related Research*, *15*(3), 49-78. https://doi.org/10.29252/LRR.15.3.3

Panadero, E. O., Jonsson, A., & Alqassab, M. (2018). Providing formative peer feedback: What do we know? In *The Cambridge handbook of instructional feedback.* (pp. 409-431). Cambridge University Press. https://doi.org/10.1017/9781316832134.020

Petticrew, M., & Roberts, H. (2008). *Systematic reviews in the social sciences: A practical guide*. John Wiley & Sons.

Pham, V. P. H. (2021). The effects of lecturer's model e-comments on graduate students' peer e-comments and writing revision. *Computer Assisted Language Learning*, *34*(3), 324-357.

https://doi.org/10.1080/09588221.2019.1609521

Rollinson, P. (2005). Using peer feedback in the ESL writing class. *ELT Journal*, *59*(1), 23-30. https://doi.org/10.1093/elt/cci003

Rotsaert, T., Panadero, E., Schellens, T., & Raes, A. (2018). "Now you know what you're doing right and wrong!" Peer feedback quality in synchronous peer assessment in secondary education. *European Journal of Psychology of Education*, *33*(2), 255-275. https://doi.org/10.1007/s10212-017-0329-x

Rouhi, A., Dibah, M., & Mohebbi, H. (2020). Assessing the effect of giving and receiving written corrective feedback on improving L2 writing accuracy: does giving and receiving feedback have fair mutual benefit? *Asian-Pacific Journal of Second and Foreign Language Education*, *5*(1), 11. https://doi.org/10.1186/s40862-020-00093-z

Saeed, M. A., Ghazali, K., & Aljaberi, M. A. (2018). A review of previous studies on ESL/EFL learners' interactional feedback exchanges in face-to-face and computer-assisted peer review of writing. *International Journal of Educational Technology in Higher Education*, *15*, Article 6. https://doi.org/10.1186/s41239-017-0084-8

Saeed, M. A., Ghazali, K., Sahuri, S. S., & Abdulrab, M. (2018). Engaging EFL learners in online peer feedback on writing: What does it tell us? *Journal of Information Technology Education: Research*, *17*. https://doi.org/10.28945/3980

Sayed, H. O. (2010). Developing business management students' persuasive writing through blog-based peer feedback. *English Language Teaching*, *3*(3), 54-66. Retrieved from https://eric.ed.gov/?id=EJ1081812

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton, Mifflin and Company.

Shang, H. F. (2022). Exploring online peer feedback and automated corrective feedback on EFL writing performance. *Interactive Learning Environments*, *30*(1), 4-16. https://doi.org/10.1080/10494820.2019.1629601

Song, W., & Usaha, S. (2009). How EFL university students use electronic peer response into revisions. *Suranaree Journal of Science & Technology*, *16*(3). Retrieved from http://sutir.sut.ac.th:8080/jspui/bitstream/123456789/3630/1/131676.pdf

Strijbos, J. W., Ochoa, T. A., Sluijsmans, D. M., Segers, M. S., & Tillema, H. H. (2009). Fostering interactivity through formative peer assessment in (web-based) collaborative learning environments. In *Cognitive and emotional processes in web-based education: Integrating human factors and personalization* (pp. 375-395). IGI Global. https://doi.org/10.4018/978-1-60566-392-0.ch018

Tai, H. C., Lin, W. C., & Yang, S. C. (2015). Exploring the effects of peer review and teachers' corrective feedback on EFL students' online writing performance. *Journal of Educational Computing Research*, *53*(2), 284-309. https://doi.org/10.1177/0735633115597490

Tan, S., Cho, Y. W., & Xu, W. (2023). Exploring the effects of automated written corrective feedback, computer-mediated peer feedback and their combination mode on EFL learner's writing performance. *Interactive Learning Environments*, *31*(10), 7276-7286. https://doi.org/10.1080/10494820.2022.2066137

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, *68*(3), 249-276. https://doi.org/10.3102/00346543068003249

Vaezi, S., & Abbaspour, E. (2015). Asynchronous online peer written corrective feedback: Effects and affects. In *Handbook of research on individual differences in computer-assisted language learning* (pp. 271-297). IGI Global. https://doi.org/10.4018/978-1-4666-8519-2.ch012

Vahedipour, R., & Abbaspour, E. (2017). Impact of Wiki-based Feedback on Grammatical Accuracy of Iranian EFL learners writing skills. In M. Rahimi (Ed.), (pp. 27). https://doi.org/10.4018/978-1-4666-8519-2.ch012 (Please delete this reference from the reference list because it is a mix-up of the previous and the subsequent reference. The in-text citation is Vahedipour, Z., & Rezvani, E. (2017), which was marked in red on page 9)

Vahedipour, Z., & Rezvani, E. (2017). Impact of wiki-based feedback on grammatical accuracy of Iranian EFL learners' writing skill. *International journal of foreign language teaching & research*, *1*(20), 111-124. https://sanad.iau.ir/Journal/jfl/Article/990772

Van Den Berg, I., Admiraal, W., & Pilot, A. (2006). Design principles and outcomes of peer assessment in higher

education. *Studies in Higher Education*, *31*(3), 341-356. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-33646534004&doi=10.1080%2f03075070600680836&partnerID=40&md5=82f095582e8c3ea4c1124129868fe975

Van Zundert, M., Sluijsmans, D., & van Merriënboer, J. (2010). Effective peer assessment processes: Research findings and future directions. *Learning and Instruction*, *20*(4), 270-279. https://doi.org/10.1016/j.learninstruc.2009.08.004

Villamil, O. S., & de Guerrero, M. C. (2019). Sociocultural theory: A framework for understanding the social-cognitive dimensions of peer feedback. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 23-41). Cambridge University Press.

Villamil, O. S., & de Guerrero, M. C. M. (2006). Sociocultural theory: A framework for understanding the social-cognitive dimensions of peer feedback. In F. Hyland & K. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 23-41). Cambridge University Press. https://doi.org/10.1017/CBO9781139524742.004

Vuogan, A., & Li, S. (2022). Examining the Effectiveness of Peer Feedback in Second Language Writing: A Meta-Analysis. *TESOL QUARTERLY*. https://doi.org/10.1002/tesq.3178

Wanchid, R. (2015). Different sequences of feedback types: effectiveness, attitudes & preferences. *PASAA*, *50*, 31-64. https://doi.org/10.58837/CHULA.PASAA.50.1.2

Wang, Y. C. (2015). Promoting collaborative writing through wikis: a new approach for advancing innovative and active learning in an ESP context. *Computer Assisted Language Learning*, *28*(6), 499-512. https://doi.org/10.1080/09588221.2014.881386

Wu, J., Zhang, L., Lu, C., & Zhao, Y. (2022). The Influence of Peer Review on the Effectiveness of Foreign Language Teaching: A Meta-Analysis. *Modern Foreign Languages*, *45*(4), 539-551. Retrieved from https://kns.cnki.net/kcms/detail/44.1165.H.20220411.1813.024.html

Xu, Y., Singh, C. K. S., Singh, S. K. J. C. B., Richard Peter , Arivayagan, K., & Bamiro, N. B. (2024). The Digital Classroom: Systematic Review of Use of English as a Foreign Language (EFL) in Teaching in Collaborative Online Settings. *International Journal of Learning, Teaching and Educational Research*, *23*(10), 1-26. https://doi.org/10.26803/ijlter.23.10.1

Yang, M., Badger, R., & Yu, Z. (2006). A comparative study of peer and teacher feedback in a Chinese EFL writing class. *Journal of Second Language Writing*, *15*(3), 179-200. https://doi.org/10.1016/j.jslw.2006.09.004

Yang, Y. F. (2015). Transforming and constructing academic knowledge through online peer feedback in summary writing. *Computer Assisted Language Learning*, *29*(4), 683-702. https://doi.org/10.1080/09588221.2015.1016440

Yang, Y. F., & Wu, S.-P. (2011). A collective case study of online interaction patterns in text revisions. *Journal of Educational Technology & Society*, *14*(2), 1-15. Retrieved from https://www.jstor.org/stable/10.2307/jeductechsoci.14.2.1

Yu, S., & Hu, G. (2017). Understanding university students' peer feedback practices in EFL writing: Insights from a case study. *Assessing Writing*, *33*, 25-35. https://doi.org/10.1016/j.asw.2017.03.004

Yu, S., & Lee, I. (2016). Peer feedback in second language writing (2005–2014). *Language Teaching*, *49*(4), 461-493. https://doi.org/10.1017/S0261444816000161

Wahyudin, A. Y. (2018). The impact of online peer feedback on EFL students writing at tertiary level. *BAHTERA: Jurnal Pendidikan Bahasa dan Sastra*, *17*(1), 1-10. https://doi.org/10.21009/BAHTERA.171.1

Zhang, X., & McEneaney, J. E. (2020). What is the influence of peer feedback and author response on Chinese University students' English writing performance? *Reading Research Quarterly*, *55*(1), 123-146. https://doi.org/10.1002/rrq.259

Zhao, H. (2010). Investigating learners' use and understanding of peer and teacher feedback on writing: A comparative study in a Chinese English writing classroom. *Assessing Writing*, *15*(1), 3-17. https://doi.org/10.1016/j.asw.2010.01.002

Zheng, L., Cui, P., Li, X., & Huang, R. (2018). Synchronous discussion between assessors and assessees in web-based peer assessment: impact on writing performance, feedback quality, meta-cognitive awareness and

self-efficacy. *Assessment & Evaluation in Higher Education*, *43*(3), 500-514. https://doi.org/10.1080/02602938.2017.1370533

Acknowledgments

This research was supported by the Ministry of Higher Education (MoHE) Malaysia through the Fundamental Research Grant Scheme (FRGS/1/2024/SSI09/UPSI/02/19).

Authors contributions

Authors contributed equally to the study.

Funding

This research was supported by the Ministry of Higher Education (MoHE) Malaysia through the Fundamental Research Grant Scheme (FRGS/1/2024/SSI09/UPSI/02/19).

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Informed consent

Obtained.

Ethics approval

The Publication Ethics Committee of the Sciedu Press.

The journal's policies adhere to the Core Practices established by the Committee on Publication Ethics (COPE).

Provenance and peer review

Not commissioned; externally double-blind peer reviewed.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Data sharing statement

No additional data are available.