Research Methods in Applied Linguistics 2 (2023) 100065

Contents lists available at ScienceDirect

Research Methods in Applied Linguistics

journal homepage: www.elsevier.com/locate/rmal

Trustworthiness in L2 writing research: A review and analysis of qualitative articles in the *Journal of Second Language Writing*

A. Mehdi Riazi [a,*], Reza Rezvani [b], Hessameddin Ghanbar [c]

[a] *Hamad Bin Khalifa University, Qatar*
[b] *Yasouj University, Iran*
[c] *Islamic Azad University, Fereshtegaan International Branch, Iran*

ARTICLE INFO

Keywords:
Research methodology
Qualitative research
L2 writing
Trustworthiness
Quality and rigor in research

ABSTRACT

Researchers in applied linguistics have long been concerned with issues of quality and rigor. Because of various paradigmatic tensions, many qualitative researchers have sought to move away from the entrenched concepts of *reliability, validity*, and *objectivity* traditionally associated with quantitative research methodologies. Given the significance of quality in research methodology and the wide variation in evaluative nomenclature, we set out to investigate how "trustworthiness" is addressed by L2 writing researchers in their published articles in the *Journal of Second Language Writing*. We first identified all empirical articles in the JSLW since its inception (June 1992) up to volume 56 (June 2022), assigning them to one of three research approaches: quantitative, qualitative, or mixed methods. We found 389 articles that used primary data for analysis. From this pool, 152 (39%) were coded as qualitative in terms of their data and methods. Using AntConc (Anthony, 2014), 43 articles were found to have used quality-related terms or techniques. We then developed a coding and annotation scheme to code and annotate all 43 articles. Findings show that although researchers addressed research quality in different ways, there is a widespread lack of systematic attention to quality criteria in L2 writing qualitative studies. We hope the study findings and discussion of this review advance Applied Linguistics qualitative research and understandings of trustworthiness. In particular, we hope both early career and experienced researchers will find the insights generated by this study useful when designing, conducting, and publishing qualitative research.

1. Introduction

In her critical survey, Lazaraton (1995) provided a variety of evidence to support the assertion that "qualitative research has made significant gains in terms of visibility and credibility in recent years" (p. 456) in applied linguistics (AL). The rising popularity of qualitative research in AL has been reported by other researchers since Lazaraton's review. For example, Riazi, Shi, and Haggerty (2018) reported that in the *Journal of Second Language Writing* (JSLW), qualitative research articles outnumbered quantitative research reports over the journal's lifecycle. Pelaez-Morales (2017), Canagarajah (2016), Richards (2009), and Benson et al. (2009), among others, similarly reported that qualitative articles were predominant in the journals of *JSLW, Applied Linguistics*, and *TESOL Quarterly*, respectively.

Notwithstanding its apparent trendy growth, qualitative research is not without criticisms. The main criticism coming from the opposite camp concerns the relationship between the researcher (i.e., knower) and the object of investigation (i.e., un/known).

* Corresponding author.
E-mail address: ariazi@hbku.edu.qa (A.M. Riazi).

https://doi.org/10.1016/j.rmal.2023.100065
Received 21 January 2023; Received in revised form 25 June 2023; Accepted 25 June 2023

Epistemologically, then, the relationship is described as dependent and based on subjective observations and interpretations. Such observations provoked "allegations that qualitative research will be identified as undisciplined [and] sloppy research" (Lewis, 2009, p. 2). Hence, a continuing discussion centers on how to define and evaluate the quality of qualitative research.

There have been various attempts to answer the question of quality and to move away from the traditional terminologies of *reliability, validity*, and *objectivity* associated with quantitative research methodologies. One of the main reformulations of quality and quality criteria that sought to move away from a positivistic paradigm (Lazaraton, 2003; Shenton, 2004) and its epistemic criteria is credited to Lincoln and Guba, who in *Naturalistic Inquiry* (1985) set forth the concept of *trustworthiness* and a set of alternative and method-appropriate criteria: *credibility, dependability, transferability,* and *confirmability* (Bazeley, 2013; Flick, 2007; Glesne, 2016).

The impetus for the present study came from our meetings to discuss another project, where we analyzed and annotated a wide array of themes concerning data coding and analysis in second language (L2) writing research (see Riazi, Ghanbar, & Rezvani, 2023). When discussing quality measures and criteria considered and reported in the corpus, we noted varied nomenclature and procedures. Such concerns also grew from our experiences teaching postgraduate research methods courses. Interestingly, we each faced similar challenges when addressing the topic of quality in qualitative research and how it is established (for similar experiences, see also Delyser, 2008; Stallings, 1995). In our respective courses, we found ourselves going to great lengths to explain concepts such as "confirmability." Difficulties were also exacerbated when we sought to help our students understand how AL researchers approached and reported on quality-related matters. AL articles we used in our classes to showcase various issues of interest in qualitative research usually fell short of disclosing how the quality or trustworthiness of the research and research findings were established. Some illustrative articles utilized conventional quantitative research quality terminologies such as "validity" that the students were encouraged to avoid in qualitative inquiry. Given the significance of quality in research methodology and the definitional and terminological confusion, we then set out to investigate how trustworthiness is addressed by L2 writing researchers[1] in their published articles in the *Journal of Second Language Writing*. This was based on our shared sense of an urgent need in L2 writing for systematic and professional scrutiny of the methodological rigor of research pertaining to trustworthiness. We, therefore, intend to take the methodological robustness in this important field to the next level by scrutinizing how previous recommendations and frameworks relating to the quality of qualitative research have been put into practice. We also seek to help early career L2 writing scholars and, more broadly, AL researchers enhance the quality of their qualitative research studies. To achieve these objectives, the following research questions guided this critical review:

1. What qualitative terminology (i.e., *trustworthiness, credibility, dependability, transferability, confirmability*) or conventional terminology (i.e., *validity*, etc.) are used in the published articles to refer to trustworthiness-related concerns?
2. What strategies (i.e., different triangulation types, member checks, etc.) are used to ensure trustworthiness?
3. What sources (books, articles, etc.) do L2 writing researchers rely on to define and address trustworthiness in their qualitative studies?

Although our intended audience is emergent and early-career researchers, we hope experienced researchers also will benefit from the insights generated by the findings and discussion of this study. We especially seek to contribute to ongoing conversations concerning quality and rigor in qualitative research in L2 writing and across the interdisciplinary field of applied linguistics.

2. Background to the study

Issues of quality and rigor—that is, what constitutes "good" qualitative research—have captured the attention of researchers from the inception of this approach. In contrast to outsiders, who are obsessed with more positivistic approaches and unwavering oppositions to qualitative research, insiders, as qualitative-inclined researchers and theorists, have sought to establish consensual criteria and procedures for ascertaining quality in qualitative research. Broadly, there have been three positions concerning criteria to judge the quality of qualitative research: *exploiting traditional quantitative research criteria, reformulating the conventional criteria*, and *developing alternative criteria*. Some writers have also questioned the viability of using or developing any predetermined generic criteria, which we might categorize as *rejecting the need to develop predetermined generic criteria*.

Advocates of the first stance (see, for example, Cypress, 2017; Mores, 2015) suggest adopting the classical concepts and terminologies of quantitative research (Morse et al., 2002) as alternatives are not dissimilar to the conventional notions of reliability and validity in closer analysis (Long & Johnson, 2000). Reliability is linked to the stability and consistency of data and the procedures and findings in repeated data collection. However, given the basic assumption in the qualitative paradigm that reality is multiple and constructed, this pursuit of repeatability and reliability may be considered imposing and unhelpful in assessing research quality (Sandelowski, 1993). Validity is conceptualized dichotomously as internal and external validity. Internal validity in classical interpretation in quantitative research aims to show that the achieved results in a quantitative study are due to the manipulation of the study's independent variables and not the effect of any extraneous variables. This entails exerting control over the context and standardization, which are at odds with the qualitative research paradigm (Kirk & Miller, 1986). External validity is also linked to the representativeness of the sample and the potential generalizability of the findings to the target population. As with reliability, the concepts of representativeness and generalizability, as conventionally understood in the quantitative paradigm, have proven difficult aims or measures to adopt in qualitative research.

[1] When we refer to L2 writing researchers, we mean those whose papers were published in the *Journal of Second Language Writing* and analyzed for this article.

Several attempts have been made to reconceptualize reliability and validity to make them more suitable for qualitative researchers. Reformulations of reliability mainly targeted the data collection and analysis procedures in the interest of transparency (Flick, 2007) by developing procedural guidelines: for example, how to conduct interviews. Similar efforts were made to reformulate the concept of validity (see Kvale, 2007), more of which are specific to variants of qualitative research methods and techniques (e.g., Legewie, 1987). Johnson (1997) and Johnson and Christensen (2019) also developed a more comprehensive set of evaluative criteria, maintaining the concept and term of validity. In their elaboration, validity comes in three types—descriptive, interpretive, and theoretical—and is promoted by strategies such as triangulation and peer review.

On the other hand, insiders, including qualitative researchers, have sought to ensure the "goodness" of their qualitative investigations by providing evidence for the quality of their research. Qualitative research methodologists (e.g., Charmaz, 2006; Tracy, 2010) also sought to develop alternative evaluative criteria on the grounds that traditional ones fail to account for the specific features of qualitative research (Healy & Perry, 2000). Concerned with how a researcher can "persuade his or her audiences that the research findings of an inquiry are worth paying attention to," Lincoln and Guba (1985, p. 290) offered a framework centered on the key concept of trustworthiness. Lincoln and Guba popularized the concept of trustworthiness in qualitative research and suggested some strategies and techniques that qualitative researchers could use to ensure it. They also introduced four quality criteria as evidence for the trustworthiness of their results: *credibility, transferability, dependability*, and *confirmability*. These criteria might be viewed as analogously parallel to similar notions like reliability and validity in quantitative research, that is, credibility for internal validity, transferability for external validity, dependability for reliability, and confirmability for objectivity.

Lincoln and Guba's (1985) trustworthiness framework has endured for decades as a means of evaluating quality in qualitative inquiry (Morse, 2015; Staffileno et al., 2021) across a wide range of disciplines. It is thus worthy of further explanation. Lincoln and Guba (1985) outlined some techniques for each trustworthiness criterion. For example, to achieve credibility, qualitative researchers may exploit data triangulation, method triangulation, or investigator triangulation (Denzin, 2015). Through one of two or even three triangulation techniques, researchers could show that what they report as a finding is viable, since different data sources, methods, and investigators generated similar results. Another technique is member-check or participant validation, through which the study participants will judge the accuracy of the data and the researcher's interpretations.

Regarding transferability, the researcher needs to help readers understand if the interpretations and insights of the study are transferable to similar contexts. Whereas in quantitative research, generalizability is evidenced by the researcher through statistical tests of significance; in qualitative research, as noted by Lincoln and Guba (1985), it refers to providing sufficient details and evidence so that readers may judge the transferability of the results to other contexts. Dependability is roughly the qualitative term for reliability in quantitative research. In quantitative research, reliability applies to the consistency of the data collection instruments. It also refers to providing adequate details for the study's replicability. In qualitative research, dependability refers to the inquirer's responsibility to document the procedures for generating and analyzing the results. In a sense, it is the same as the second meaning of reliability, that is, providing adequate and detailed documentation and explanations of coding and data analysis. This way, the readers will be able to reach the same inferences as those of the researcher. Finally, confirmability refers to the extent to which the researcher has reported the results objectively. Researchers may achieve confirmability by reflexivity strategies through which they can disclose and examine their beliefs, judgments, and practices in their role as the researcher and how these might have influenced the research process and results (Olmos-Vega et al., 2022).

It is worth noting that there is another position that questions attempts to develop any predetermined inclusive evaluative criteria (Garratt & Hodkinson, 1998; Hope & Waterman, 2003; Schwandt, 1996; Sparkes & Smith, 2009). One problem is that the application of blanket criteria, "however defined, is not clear, and confusion exists as to how judgments should be made about whether or not a standard has been reached" (Hope & Waterman, 2003, p. 123). Some also argue that without any unified qualitative paradigm, it would be questionable to judge the quality of qualitative inquiries using generic criteria (Rolfe, 2006; Sandelowski & Barroso, 2002). Rolfe (2006, p. 304) asserts that "the commonly perceived quantitative–qualitative dichotomy is, in fact, a continuum which requires a continuum of quality criteria" and thus, "producing frameworks and predetermined criteria for assessing the quality of research studies is futile. In Flick's (2007) perspective, efforts to advance and assure standards are incompatible with the non-standardized nature of qualitative research. Given the diversity of qualitative research, applying predetermined generic criteria to assess quality is inherently exclusionary and narrows the perception of what constitutes legitimate research (Sparkes & Smith, 2009); as a result, some researchers argue that such criteria "should be abandoned in favor of individual judgments of individual studies" (Rolfe, 2006, p. 309).

Since Lincoln and Guba's (1985) initial formulation of trustworthiness and strategies to achieve it, many qualitative scholars have discussed this and other quality aspects of research. For example, Angen (2000) traced and problematized the concept of "trustworthiness" through the different lenses of quantitative and qualitative inquiry. She argued that although "validity" has been mainly construed within the quantitative paradigm, arising from a realist ontology and a foundationalist epistemology, it has since been transformed in qualitative research to encompass a lifeworld ontology and an interpretive epistemology, which are utterly non-foundationalist. She further argued that the concept of validity in qualitative research remains controversial, and, hence, applying quantitative research methodological requirements and standards of validity to qualitative research would be counterproductive and even unethical. She further contended that there is a need for a reconfiguration of the concept of validity in qualitative research to better cater to the interpretive underpinnings of this research orientation, which hinges on the capabilities of qualitative researchers through processes of negotiation (i.e., *validation* rather than validity). She also posed the fundamental question of "What does it mean to do good qualitative research?" (p. 380). In line with this question and drawing on several studies such as Creswell (1998), she mentioned that numerous techniques such as prolonged engagement, triangulation, peer review or debriefing, negative case

analysis, clarifying researcher bias, member checks, thick description, and external audits could be employed to assess the validity of a qualitative research study, yet she discussed the drawbacks pertaining to each of these validation techniques.

In another study, Davis and Dodd (2002) sought to disentangle the concept of "rigor" in qualitative research. Like Angen (2000), they believed that applying the concept of rigor in qualitative research in the same way as it has been exploited in the quantitative orientation is a contradiction in terms. They brought this issue to the fore by focusing on the problematic use of the quantitative concept of rigor to gauge the methodological rigor of qualitative research. Wolcott (1990, as cited in Angen, 2000) argued that objectivity, as propounded in a quantitative paradigm, is non-existent given the multiplicity of reality in the interpretive paradigm (see Angen, 2000, for more discussion on subtle realism and other related polemics). Davis and Dodd (2002) further asserted that what is paramount in a qualitative study is for it to "be valid, reliable, and faithful in providing an account of the social world" (p. 288). Re-evaluating the concept of rigor in the qualitative research orientation, they propounded that investigating rigor in qualitative research revolves around the construct of ethics, which itself is composed of several sub-constructs such as attentiveness, empathy, carefulness, sensitivity, respect, honesty, reflection, conscientiousness, awareness, engagement, openness, and context. They reckoned that considering rigor in light of the aforementioned terms (e.g., attentiveness to research practice), provides a more fruitful way to assess validity in qualitative research.

In line with the above-mentioned studies, Cutcliffe and McKenna (1999) re-examined the representativeness and credibility of qualitative research. They criticized the convention of utilizing the criteria of assessing the quality of quantitative research for qualitative studies and called for establishing a specific set of rubrics for establishing the credibility of qualitative research. They pointed out that, given the abstract and complex issues targeted by qualitative research studies, it is unlikely that their credibility can be judged by the same criteria used for quantitative studies. This study, in addition, discussed the challenges and hurdles of exploiting other experts besides the researcher in coding and analyzing the data as a way of evaluating the credibility of findings in qualitative research. They also argued that since coding, categorizing, and thematizing of qualitative data would encompass the subjective judgment of a researcher, it is highly unlikely that two experts will come up with a similar set of ideas. Several techniques, such as memoing, triangulation, and audit trail, were proposed in this study to boost the credibility of the results on the basis of recommendations in state-of-the-art articles like Guba and Lincoln (1989), Guba and Lincoln (1981), and Glaser and Strauss (1967). Some other precautions were referred to regarding credibility, such as researchers' previous theoretical knowledge and its interference with the participants' world (i.e., researchers' empathy) and also the lack of methodological transparency (i.e., clarity and appropriateness of the exploited qualitative approaches).

Creswell and Miller (2000) suitably delve into the concept of validity in qualitative research and decipher its complexity and labyrinthine nature. They asserted that qualitative researchers have several methodological sources at hand (e.g., Guba & Lincoln, 1989; Lincoln & Guba, 1985; Maxwell, 1996; Merriam, 1998), within which they find a perplexing set of terms for validity (e.g., trustworthiness, credibility, plausibility, validation), typologies (e.g., Maxwell, 1992; Schwandt, 1996) and perspectives (e.g., Miles & Huberman, 1994). They also emphasized that qualitative researchers have now been urged to substantiate that their studies are credible, and, in response to this need, some scholars such as Lincoln and Guba (1985) and Maxwell (1996) advanced a set of procedures to help researchers establish the validity of their studies. This resulted in the introduction of several terms such as *member checking, triangulation, thick description, peer reviews*, and *external audits* in published qualitative works (for an analysis and critique of these procedures, see Morse, 2015). Nonetheless, they reasonably pointed out that the Achilles heel of these sets of recommendations is that they barely provided guidance for researchers concerning the rationale behind these validation procedures and the strategies for implementing them. Responding to such critiques, Creswell and Miller (2000) advanced a two-dimensional framework, in addition to nine procedures, to better assist researchers in selecting an appropriate validation procedure.

As seen from the above review, the initial framework proposed by Lincoln and Guba (1985) has remained stable over time. Subsequently, the discussion of quality in qualitative research has focused mainly on the techniques for achieving trustworthiness.

Also, an important consideration is the feeder disciplines of L2 writing and how these might affect L2 writing researchers' methodological orientations and quality criteria. Silva and Leki (2004) describe "composition studies and applied linguistics, and their parent disciplines, rhetoric and linguistics" (p. 1), as the feeder disciplines of L2 writing. Methodologically speaking, these two disciplines and their parent disciplines are shaped by different theoretical and philosophical underpinnings that can be broadly linked to (post)positivism and interpretivism. This orientation may be reflected in L2 writing research and publications in one way or another. Knowing that L2 writing researchers have used qualitative approaches extensively, their approach, particularly their reference to qualitative criteria, might be different. It would be interesting to monitor L2 writing researchers' reference to the paradigms in our findings.

In the following sections, we present the methods and procedures, findings, and discussion of the findings of our study.

3. Methods and procedures

To prepare our corpus for analysis, we first identified all empirical articles in the JSLW since its inception (June 1992) up to volume 56 (June 2022). That is, we identified all those articles that employed one of the three research approaches: quantitative, qualitative, or mixed-methods research. We found 389 empirical articles that used primary data for analysis. From this pool of articles, 152 (39%) were coded as pure qualitative articles. They used qualitative data and analysis to address L2 writing issues.

We used AntConc (Anthony, 2014) and searched each of the 152 articles for a wide range of keywords relating to the concept of trustworthiness based on the related literature on qualitative research methodological issues, as reviewed in the previous section. The search terms included "trustworthiness," "credibility," "transferability," "confirmability," "dependability," and "bias." Because some qualitative researchers still refer to "validity" instead of alternative qualitative terminology, we also searched for "valid," "validity,"

"validate," "validation," and "rigor" to ensure we did not miss any article that had strived for trustworthiness. In addition, we used terms related to techniques for achieving trustworthiness. These terms included "triangulation," "member check," "thick description," "reflexivity," "peer debriefing," and "prolonged engagement." We also found terms such as "peer check" and "persistent observation" in some articles. We included these terms in their proxy categories (e.g., *peer check* in *member checks, persistent observation* in *prolonged observation*). A total of 43 articles were found to have used such quality-related terms or a technique for achieving it. We included all 43 articles in our corpus for review and analysis. Subsequently, the contexts of all returned hits were copied and included in the coding and annotation sheet.

In most cases, when a trustworthiness- or validity-related term was found in the article, there were also mentions or descriptions of the techniques used. This allowed us to also record the techniques that were not on our list. Some examples of the techniques L2 writing researchers used but were not on our list were "constant comparative analysis" and "critical friends' perspectives" (linked with trustworthiness), "multiple perspectives" (linked to validity), and "inter-coder reliability" (linked to bias). On the other hand, the search for the techniques revealed that some researchers only reported the techniques without using trustworthiness- or validity-related terms.

3.1. Coding and annotation procedure

We developed a coding and annotation scheme to code and annotate all the 43 articles (see appendix). The coding and annotation scheme included the reference to the article, the section in which the term was used, the exact trustworthiness- or validity-related term used, the technique mentioned to achieve it, the context in which the terms and/or techniques were mentioned, and remarks by the coder.

The first author coded and annotated five articles as a model and shared it with the second and third authors. The remaining articles (39) were then coded and annotated by the second and third authors. The first author then checked the coding and annotations to ensure the accuracy of the codings and annotations. Where there was ambiguity or uncertainty, the three coders met and discussed those ambiguities and uncertainties until a consensus was reached. The results of the review and analysis of the articles regarding how trustworthiness was addressed are presented in the next section.

4. Findings

4.1. The profile of the articles

As stated in the methods section, we identified and included 43 qualitative articles from the pool of JSLW empirical articles that had strived for trustworthiness by using one or more techniques. Table 1 presents the profile of the articles.

As seen in Table 1, 10 articles explicitly used the term "trustworthiness." Of these, one (Negretti & Kuteeva, 2011) used both "trustworthiness" and "validity," and one (Wan, 2014) used both "trustworthiness" and "credibility." Also, 15 articles used "validity," of which one (Canagarajah, 2015) used "validity" and "credibility" simultaneously, and three (Han & Hyland, 2015; Hyland, 1998; Hyland & Hyland, 2001) used "validity" and "bias" simultaneously. Three articles (Costino & Hyon, 2007; Sánchez-Martín & Seloni, 2019; Wette, 2010) used "bias," and one (Lee & Schallert, 2008) used "credibility." Surprisingly, 14 articles did not mention any trustworthiness- or validity-related terms, although they used some techniques like data triangulation (e.g., Bankier, 2022) or member check (e.g., Fujioka, 2014).

In the next section, we focus on the techniques used to achieve trustworthiness or validity, as mentioned by the L2 writing researchers and profiled in Table 1.

4.2. Techniques used to achieve trustworthiness

Table 2 presents the frequency and percentage of the techniques L2 writing researchers mentioned they used to achieve trustworthiness in their studies.

As seen in Table 2, there were six categories L2 writing researchers referred to when seeking evidence for trustworthiness or related quality elements. These six categories included "triangulation" with the highest percentage (43%), "member check" (30%), "seeking own and other views" (10%), "reflexivity" and "thick description" (5%), and "others" (7%). As can be seen, L2 writing researchers have relied heavily on just two techniques. A case in point is reflexivity, which implies thinking and critically reflecting on a researcher's role in a study and how the researcher has taken their stance into account in their research.

The most popular type of triangulation mentioned by L2 writing researchers was "data triangulation," with 21 mentions. Next to data triangulation was "member check," with 18 instances, followed by "seeking other views," with seven instances. Seeking other views included the conventional inter-coder reliability to check the codings by another coder, thus providing evidence for reliability, validity, and lack of bias. This category included other techniques close to inter-coder or inter-reflection reliability, as stated in the table. Two other categories were "reflexivity" and "thick description" that, while did not have a high frequency, were mentioned by some researchers. Finally, other techniques like "prolonged engagement," "constant comparative analysis," and the familiarity of the interviewer with the participants in the "other" category also had a few instances in the corpus.

Table 1
The profile of articles.

Article information	Key terms used	Techniques used to achieve trustworthiness
Bankier (2022)	None	Data triangulation
Britton & Leonard (2020)	Trustworthiness	Critical friends' perspectives
Canagarajah (2015)	Validity & Credibility	Triangulation
Cheng (2011)	None	Member check
Costino & Hyon (2007)	Bias	The interviewer being familiar to the participants to reduce bias
Fujioka (2014)	None	Data triangulation Member check
Hafner & Ho (2020)	Trustworthiness	Constant comparative analysis
Han & Hyland (2015)	Validity & Bias	Inter-coder agreement
Hyland & Hyland (2001)	Validity & Bias	Data triangulation (data from different sources) Method triangulation Member check
Hyland (1998)	Validity & Bias	Data triangulation Different perspectives Member check
Jiang (2018)	None	Member check
Junqueira & Payant (2015)	Validity	Multiple perspectives
Kibler (2010)	Validity	Data triangulation Member check
Lee & Schallert (2008)	Credibility	Prolonged engagement Data triangulation Method triangulation Member check
Lee & Yuan (2021)	Trustworthiness	Data triangulation Member check
Lee (2010)	Validity	Member check Data triangulation
Lee (2013)	Trustworthiness	Data triangulation Member check
Li & Kim (2016)	None	Data triangulation
Lim & Polio (2020)	None	Data triangulation
Liu & Tannacito (2013)	Trustworthiness	Member check
Liu (2011)	None	Data triangulation
Luo & Hyland (2016)	None	Data triangulation
Negretti & Kuteeva (2011)	Trustworthiness & Validity	Reflexivity
Negretti & McGrath (2018)	Trustworthiness	Recursive reflection on researchers' interpretations
Parks (2000)	None	Reflexivity
Sánchez-Martín & Seloni (2019)	Bias	Reflexivity
Satake (2020)	Validity	None
Seloni (2014)	None	Thick description
Sengupta (1999)	None	Data triangulation
Shin, Cimasko, & Yi (2020)	Validity	Data triangulation Investigator triangulation
Smith et al. (2017)	Trustworthiness	Data triangulation Method triangulation Member check
Wan (2014)	Trustworthiness & Credibility	Member check
Wang (2020)	None	Thick description
Weigle & Nelson (2004)	Validity	Member check
Wette (2010)	Bias	Inter-coder reliability
Worden (2018)	Trustworthiness	Data triangulation Member check
Wu (2020)	Validity	Thick description Data triangulation Member check
Yi (2007)	Validity	Member check
Yi (2010)	Validity	Data triangulation Prolonged engagement Member check
Yoon (2016)	None	Data triangulation
Yu (2020)	Validity	Member check
Yu (2021)	Trustworthiness	Data triangulation
Yu et al. (2020)	Validity	Data triangulation

Note: See the Appendix for the references of these articles.

Table 2
Frequency of the techniques mentioned in the articles.

Category	Technique	Frequency	Total frequency & percentage	Target quality element
Triangulation	Triangulation (in general)	1	26 (43%)	Bias (avoiding) Credibility Trustworthiness Validity None
	Data triangulation	21		
	Investigator triangulation	1		
	Method triangulation	3		
Member check	Member check	18	18 (30%)	Credibility Trustworthiness Validity
Seeking own and other views	Multiple perspectives	2	6 (10%)	Bias (avoiding)
	Inter-coder agreement	2		Validity
	Critical friends' perspectives	1		Trustworthiness
	Recursive reflection on one's own interpretation	1		
Reflexivity	Reflexivity	3	3 (5%)	Bias (avoiding) Trustworthiness Validity
Thick description	Thick description	3	3 (5%)	Validity
Other	Prolonged engagement	2	4 (7%)	Credibility
	Constant comparative analysis	1		Validity Trustworthiness
	The interviewer being familiar to the participants to reduce bias	1		Bias

4.3. Sample studies

More often than not, qualitative researchers in AL do not provide adequate accounts of their methodology (Riazi, Ghanbar & Rezvani, 2023). However, it is widely acknowledged (Seale, 2007) that research quality can be improved by reflecting on one's own and others' research practices and reports (see, e.g., Merriam & Grenier, 2019). In what follows, we showcase seven qualitative studies of L2 writing researchers to trace and illustrate the techniques they used to provide evidence for the trustworthiness of their methods and findings. They are intended to foster "moments of scholarly reflection" (Seale, 2007, p. 383) for readers to exercise their power of judgment to see if they can use the ideas in their own research practice. The following seven qualitative studies discussed in chronological order were selected because they (a) explicitly mentioned quality considerations (i.e., trustworthiness, validity, or credibility) and (b) used more than one technique to provide evidence to ensure quality. For emphasis, we bolded the noteworthy parts in the excerpts from the articles.

Hyland and Hyland (2001) used data triangulation techniques, different perspectives, and member check to provide evidence for the validity of their results and avoid researcher bias. Hyland and Hyland's study focused on two ESL teachers' written feedback to their students. The researchers provided a detailed text analysis of the teachers' written feedback. They analyzed the feedback regarding three functions: praise, criticism, and suggestions. In addition to text analysis, the researchers collected think-aloud protocols and interviewed the teachers on their motivation for mitigating their criticisms. They also examined cases where students failed to understand their teachers' feedback and comments due to their indirectness.

In the "Participants and data" section of their article, Hyland and Hyland (2001) state:

> To ensure content reliability and combat researcher bias, triangulation and respondent validation were included in the research design. Triangulation involved obtaining as many different perspectives on the data as possible. These different perspectives came first from the different sources of data: the teachers, the students, and the researcher and from a triangulation of methods, as data were collected through interviews, questionnaires, analysis of texts, observation of classes, and verbal reports. Respondent validation, or "member checking" (Lincoln & Guba, 1985), involved allowing participants in the research access to data and seeking their input and their evaluation of its authenticity to correct researcher bias. All observation notes were commented on by the two class teachers to help validate the interpretations and minimise misrepresentations. (p. 189, emphasis added)

As seen in the above quote, the researchers explicitly aimed to reduce researcher bias and improve the validity of interpretations by exploiting several techniques. These techniques included data triangulation, method triangulation, and member check, with details in the above quote.

The next showcase study belongs to Lee and Schallert (2008), who aimed to ensure credibility and used triangulation (of data and method), member checks, and prolonged engagement to achieve it. In the "Data sources and procedures" section, the researchers explain different data sources and different methods, as quoted below:

> Data in this study came from multiple sources, gathered with multiple methods. First, at the beginning of the semester, a semi-structured background interview with the teacher and each student was conducted. (p. 170, emphasis added)
>
> …. Fifth, every class session was observed, and notes taken in a researcher's journal, with particular attention to how the teacher interacted with the students, to her instructions about how they should take her written comments, and to the students' reactions to what the teacher was saying. Finally, the scheduled individual conferences between teacher and student were observed, audiotaped, and transcribed. In this way, prolonged engagement and persistent observation, two canons of qualitative research, were met. (p. 171, emphasis added)

As indicated in the above excerpt, they also state that "For this study, we addressed credibility in the following ways: (1) prolonged engagement in the field; (2) collecting data from various sources and methods; (3) member checking with the participants; and (4) peer debriefing with each other" (pp. 171-172).

The next study was done by Yi (2010), who referred to validity and used data triangulation, prolonged engagement, and member checks to provide evidence for the validity of the reported results. The researcher explains the data sources in the "Data Collection" section:

> The present study draws on ethnographic data gathered over the course of two academic years. Several data collection techniques were triangulated, including participant observations (through tutorials and in the ESL Resources classroom) and fieldnotes; interviews and transcripts; Jihee's literacy activity checklists; the collection of out-of-school written text artifacts, including her literacy autobiography, personal diary entries, scribbles, notes, emails, and online chatting with me; samples of her academic writing; and informal and formal oral interviews with her mother, ESL teachers, and her friends. (p. 21. emphasis added)

In addition to data triangulation, the researcher refers to prolonged engagement as another technique:

> Similarly, my study required me to have prolonged engagement with Jihee, building on both our mutual trust and respect and also her great curiosity about my conducting research on her literate life. During the early stage of the study, I simply asked her to show me the cover of the diary so that I could have an idea of what it looked like. (p. 21, emphasis added)

In the "Data analysis" section, Yi (2010) also refers to another technique, member check, and clarifies that these different techniques were used to enhance the validity of the study.

> To enhance validity, in addition to triangulation of data collection techniques, I engaged in member checks with Jihee at different stages throughout the study. Informal member checks constantly took place during our weekly interview sessions and through online communication (email, online chatting) and phone conversations. In addition, I conducted formal member checks in the final interview session; in these exchanges, she made comments, elaborated on certain points, or adjusted my interpretations and representation of her and her literate life. She also read an earlier draft of this paper and shared some of her thoughts. (p. 22, emphasis added)

One of the studies which explicitly targeted trustworthiness through data triangulation, method triangulation, and member check was that of Smith, Pacheco, and Almeida (2017). "Data from multiple sources were collected and triangulated to construct a multifaceted understanding of multimodal codemeshing processes" (p. 10, emphasis added). Further, the researchers state: "We sought to establish trustworthiness through triangulating different sources and methods, conducting member checks with participants, and seeking disconfirming evidence" (p.12, emphasis added).

In another study, Wu (2020) used the techniques of thick description, data triangulation, and member check to ensure the validity of the results. In the "Data analysis" section, the researcher states: "This research followed a multiple-case study approach and sought to provide 'thick description' and grounded interpretation" (Prior, 1995, p. 321) of the participants' digital literacy practices" (p. 4, emphasis added). In addition, on page 5, the researcher elaborates on data triangulation and member check intended to ensure the validity of the study:

> In the final stage, I conducted within-case analysis (Merriam, 2009) and cross-case analysis (Yin, 2009). For the within-case analysis, I triangulated the data (i.e., diaries, interviews, and resource grids) and composed narratives for each participant. The narratives focused on the connections between the products, processes, and perceptions of the participants' digital practices. Member-checking was conducted to ensure the validity of my interpretative accounts. For the cross-case analysis, I compared the narratives to generate thematic discussions relative to the three research questions (contextual relations, mediational relations, and shaping factors). (p. 5, emphasis added)

In addition to the above five studies in which trustworthiness and validity were explicitly referred to and addressed, we will present two other studies that did not mention any trustworthiness- or validity-related terms, whilst they referred to a technique related to trustworthiness. The two articles are Park (2000) and Wang (2020), in which reflexivity and thick description, respectively, were pointed to and used.

Park (2000) recognized the role of the researcher as an instrument in qualitative research and referred to reflexivity, as indicated in the following quote:

> With respect to my own involvement as a collaborator, it is important to note that within qualitative research, the researcher, a participant observer, is not regarded as having a neutral role. Researchers can and do influence participants' actions, what

Table 3
Details of the trustworthiness-related sources used in the 21 articles.

Source	Type (book, article, etc.)	Number of times used
Charmaz (2006)	Book	1
Creswell (1998)	Book	1
Creswell (2003)	Book	1
Creswell (2007)	Book	1
Denzin & Lincoln (2003)	Book	1
Duff (2008)	Book	1
Erlandson et al. (1993)	Book	2
Geertz (1973)	Book chapter	1
Guba & Lincoln (2005)	Book chapter	1
Hammersley (1983)	Book	1
Han & Hyland (2015)	Journal article	1
Hesse-Biber, & Leavy (2011)	Book	1
Jones & Arminio (2006)	Book	1
Lankshear & Knobel (2004)	Handbook	1
Lincoln & Guba (1985)	Book	5
Merriam (2009)	Book	2
Pinnegar & Hamilton (2009)	Book	1
Prior (1995)	Journal article	1
Ramanathan & Atkinson (1999)	Book	1
Richards (2003)	Book	1
Starfield (2013)	Book chapter	1
Webb & Danaher (2002)	Book	1

> Hammersley (1983) has referred to as the principle of reflexivity. Although I did not overtly invite questions about their writing, because of the ESL course, students knew me first and foremost in my role as an English teacher. (p. 111, emphasis added)

Similarly, Wang (2020) mentioned thick description without linking it to trustworthiness. It is stated that "Drawing on ethnographic methods, this study turns to a mobility framework to engage in thick descriptions of multilingual writers' translation practices as mediated, distributed, and negotiated language work" (p. 2, emphasis added).

4.4. Sources used to justify trustworthiness in qualitative studies

In this review, we also aimed to see what sources informed the methodology of the studies as far as quality is concerned. Our review of the 43 articles showed that in 23 (53%) articles, no sources related to trustworthiness or its constituents were cited. In the remaining 20 (47%) articles, an overall 28 sources related to trustworthiness or its elements were used. In addition, eighteen articles referred to one source to discuss trustworthiness or its elements; however, there were three articles (Fujioka, 2014; Lee & Yuan, 2021; Sánchez-Martín & Seloni, 2019) that cited two sources, and one article (Negretti & Kuteeva, 2011) that mentioned three sources. Table 3 provides some details about the sources mentioned in the 20 articles.

As seen in Table 1, L2 writing researchers referred to 16 books, one handbook, three book chapters, and two journal articles when they wanted to address trustworthiness or its elements in their articles. The most popular book was Lincoln and Guba (1985), which was mentioned five times, followed by Erlandson et al. (1993) and Merriam (2009), which was mentioned two times. All the other sources were mentioned only once.

5. Discussion

Our results broadly indicated insufficient and inconsistent attention to the quality criteria (trustworthiness and its components) in L2 writing qualitative studies. Since the corpus of the study included qualitative articles from the same JSLW, the common complaint of researchers that the demands of the journal or its editors to shorten the reports (Silverman, 2014) is ruled out as there were some articles which rather elaborately dealt with the trustworthiness of the findings. In addition, because the journal was launched in 1992, years after the inception of the discussions and proposals of quality criteria for qualitative research and ensuing clash of vocabularies as Smith (1997) noted, we expected to see rising use of trustworthiness terminologies and techniques in the articles of the corpus. However, we could not identify such a chronological pattern in the corpus. Conversely, the more frequent use of conventional terminologies like validity refutes the assumption that qualitative researchers more recently prefer terms like trustworthiness (Glesne, 2016; Patton, 2015) and rarely use traditional terms "because of their association with quantitative measurement" (Neuman 2006, p. 194).

In other words, we assume that the first stance outlined in the literature review, adhering to the classical concepts and terminologies of quantitative research (Morse et al., 2002), is still the strongest as far as our corpus of L2 writing research indicates. The finding thus suggests that classical quality terminologies like validity are lingering in the field and in many L2 scholars' studies. Since Lincoln and Guba's (1985) revolutionary perspective on quality in qualitative research, the delineation of goodness in qualitative research (see

for example Angen, 2000; Creswell & Poth, 2018; Eisner, 1991; Lather, 1991, Lincoln et al., 2011; Richardson & St. Pierre 2005; Tracy, 2010; Whittemore, et al., 2001; Wolcott, 1990, 1994) has been continually evolving resulting in a vast variation in both terminologies and their substantive underpinnings. The relentless quest for an agreed-upon understanding of quality in qualitative research and strategies to ensure it may induce qualitative researchers, including L2 writing scholars, to opt for more standardized and widely known nomenclature from quantitative research, although it is paradigmatically and methodologically different.

From another perspective, heavy reliance on quality criteria aligned to a (post)positivist paradigm may reflect the feeder disciplines in L2 writing. As discussed in the literature review, L2 writing has its roots in composition studies and applied linguistics (Silva & Leki, 2004). We may anecdotally (since we do not have solid data on the articles' underpinning discipline) say that reliance on such criteria may indicate that L2 qualitative researchers are predominantly applied linguists with a more scientific approach to research. Finally, according to Creswell and Poth (2018, p. 335), some authors even argue that using such positivist terminologies increases the chance of their paper's acceptance "in a traditionally focused quantitative world".

Regarding strategies, our analysis showed triangulation as the most frequently used strategy by L2 writing researchers to shore up the trustworthiness of their studies. Triangulation is the most well-known of the quality-ensuring strategies (Merriam & Grenier, 2019) used even before Denzin's (1970) systematic conceptualization (see Campbell & Fiske, 1959; Webb et al., 1966). Its use in qualitative research is recommended to ensure all the quality criteria except transferability (Lincoln & Guba, 1989). L2 researchers' frequent use of triangulation was thus sensible, given that it is reasonably easy and cost-effective (Creswell & Poth, 2018). Similarly, member checks are recommended for all the quality criteria except transferability (Guba & Lincoln, 1989). It is also supposed to be "the most crucial technique for establishing credibility" (Lincoln & Guba, 1985, p. 314), which was the most frequently and explicitly targeted quality criterion in our corpus. This might explain why it was the second most highly used strategy to ensure trustworthiness.

Another observation was the notorious absence of other quality criteria, namely, dependability, transferability, and confirmability, in the research reports of the corpus. With several theoretical articles on quality (e.g., Angen, 2000; Cutcliffe & McKenna, 1999; Davis & Dodd, 2002) and a few typologies (e.g., Maxwell, 1992; Lincoln & Guba, 1985; Schwandt, 1996; Tracy, 2010), our findings show that despite efforts to define quality in ways that diverge from the hegemony of positivist approaches, L2 writing researchers did not attend to such criteria. It is therefore imperative that future L2 writing researchers attend to these quality criteria when planning, implementing, and reporting studies.

More importantly, even when these terms have been mentioned, L2 writing researchers did not elaborate on the procedures they used to achieve them. The absence of elaboration and detail might be attributed to the fact that authors, although they might know these terms, are not cognizant of the rationale behind them and their implementation. As Creswell and Miller (2000) pointed out, utilizing each of those procedures should be based on "the lens researchers choose to validate their studies and researchers' paradigm assumptions" (p. 124). In our view, the reluctance or failure of some researchers to provide adequate and transparent explanations of measures taken to ensure the trustworthiness of the findings reflects what Silverman (2014) termed "anecdotalism" in qualitative research reports. This problem of anecdotalism, we argue, doesn't help to dispel the common misconception that qualitative research is a "fairly sloppy business involving little more than taking up a position and supplying a few well-chosen examples to lend it support" (Richards, 2003, p. 285). To keep away from such critiques, L2 writing researchers may consider narrative inquiry. Although not quantifiable, narrative inquiry based on detailed stories and narratives is considered a reputable technique in L2 education (see e.g., Barkhuizen, 2013; Bruner, 2010; Casanave, 2005; Mishler, 1990). A case in point is prolonged engagement, which is one of the techniques for establishing credibility in qualitative research. However, it was rarely reported in our corpus. This is a contingent issue regarding the constructivist paradigm of qualitative research. Techniques like triangulation or member checks were preferred over thick description, prolonged engagement, peer debriefing, and reflexivity in our study as well. This finding, especially the limited use of thick description, is in line with that of Tardy's (2021) finding in writing studies more broadly. One of the underlying causes for this finding can be reflected in sources used by L2 writing researchers to establish trustworthiness. As presented in the results section, only 21 out of 43 articles used a source relating to trustworthiness (or validity); in most cases, one source was cited. Arguably, the lack of elaboration and justification of the use of procedures to establish trustworthiness can be attributed to a lack of full understanding of the procedures. This argument is based on avoiding pertinent citations and researchers' perception that there are straightforward techniques to be described in only a few sentences. Even for member checking, which is one of the most well-known and frequently used procedures, as our corpus also suggested, it is imperative that qualitative researchers report which member checking was or was not used with examples of the participants' responses together with a description of the disconfirming cases and possible changes made (Brice, 2014). It stands to reason that such a lack of substantiation of methodological choices would result in overlooking the concept of trustworthiness of results and the procedures for attaining it in L2 writing research.

In addition, more than half of the reports did not point to any theoretical or methodological resources on quality in research, suggesting that they might have taken it as common knowledge in research. Interested researchers working in special research areas such as second language writing well know the leading journals of the field and their publications. Prospective authors intending to publish their research reports in JSLW search for and study theoretically and methodologically pertinent articles in the journal. If they are concerned about inclusion or how to report quality assurance measures and discussions, they are likely to notice the inconsistent attention, as our study indicated, and may not find it necessary to address or elaborate on the trustworthiness of their findings. However, as indicated earlier in this article, quality in qualitative research and procedures to ensure it are, far from consensus and common knowledge. Quality in qualitative research cannot be assumed (Whittemore et al., 2001) but rests on explication of "how we claim to know what we know" (Altheide & Johnson, 1994, p. 496). This entails researchers expressing their stances and positionalities and transparently describing the purposes intended and procedures used to show a correspondence between the researchers' understandings and reconstructions and reality. This explication will advantage both the researcher and the readers of the research

reports. One strategic technique is to use and cite resources which inform or guide the methodology, especially quality assurance. This presents and emphasizes both theory and practice and leads to more transparent, accountable research practices.

6. Conclusions and recommendations

As we noted in the introduction, the impetus for this research was twofold. First, even though the three of us have taught research methods in applied linguistics for many years at the graduate level, we found ourselves dissatisfied with discussions of quality measures, trustworthiness, evaluative criteria, and related matters reported in the published literature. Even the illustrative articles we used in our classes frequently failed to make explicit how the researchers conceptualized and ensured the quality of qualitative inquiry. Moreover, the divergent and overlapping terminologies used created confusion for our students and ourselves—often contradicting our own professional training and practices. Second, because second language writing is one of our areas of interest as L2 writers, researchers, and teachers, we wondered how researchers in this area deal with matters of quality or "goodness" in their published studies. Based on our systematic review of this literature, we offer the following two general conclusions and associated recommendations:

1. Of the four positions concerning criteria to judge the quality of qualitative research discussed in the background to the study, that is, *exploiting traditional quantitative research criteria, reformulating the conventional criteria, developing alternative criteria,* and *rejecting the need to develop predetermined generic criteria*, the first position was shown to be predominant in L2 writing research. That is, L2 writing researchers preferred to use traditional quantitative research criteria like "validity." Further, when L2 writing researchers approached the second position, that is, using alternative criteria, except for "trustworthiness" and "credibility," none of the other criteria (dependability, transferability, or confirmability) were clearly approached and accounted for.

At the current stage of the relentless quest for the characterization of the construct of quality in qualitative inquiry, we cannot recommend any of the positions outlined in the literature review and depicted in the review of our corpus. Nor can we (or should we) encourage L2 researchers to use a single or standardized nomenclature. We, however, recommend that prospective researchers clarify their position and understanding of "goodness" in qualitative research, along with how they addressed it in their study. It is also essential that they make informed and consistent use of quality-associated terminologies with substantive underpinnings. It also follows then that there should be more sustained effort to delineate various quality-related concepts and strategies, particularly those (for example, dependability) correlated with alternative approaches to address goodness in qualitative inquiry. In order for researchers to consider and use such concepts and strategies in their research, they should be clearly characterized in the theory and literature of their qualitative research methodology.

2. Of the various strategies for substantiating trustworthiness in L2 writing research, triangulation was most frequently used. Similarly, member checking was another common strategy used by L2 writing researchers. However, even when researchers mentioned such strategies, they did not elaborate on the specific steps or procedures. Regarding cited sources, only fewer than half (21 out of 43) articles used a source relating to trustworthiness (or validity); in most cases, only one source was cited. Since reflections and discussions on research quality are central to advancing research methodology and related issues, it is recommended that there should be more explicit attention to the trustworthiness considerations in the process of planning and conducting research and not just as a post hoc step (Glesne, 2016; Morse et al., 2002; Porte & Richards, 2012). Similarly, researchers must provide clearer, more explicit accounts of strategies and techniques used to ensure quality. Given the variations in qualitative research practices and various approaches and strategies to address quality in qualitative research, as we sketched in this article, we recommend qualitative researchers implement at least two strategies in their inquiries (Creswell & Poth, 2018). At minimum, in their research reports authors should cite sources of relevance or guidelines for interested readers and researchers.

Creswell and Poth (2018, p. 334) pointed out that an understanding of both traditional and current perspectives of validation in qualitative research is "essential for informing the work of qualitative researchers and readers of qualitative research." We further argue that because "pure intentions do not guarantee trustworthy findings" (Robson & McCartan, 2016, p. 204), it is imperative that researchers and readerships be informed of the praxis of how quality in interpretive inquiries is actually characterized and reported. We hope that the findings of this review paper and our discussion will benefit diverse stakeholders in AL qualitative research and, particularly, L2 writing researchers. We also encourage other researchers in various areas of AL to replicate our study, which was intended to be informative rather than exhaustive.

Declaration of Competing Interest

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.

References

Altheide, D. L., & Johnson, J. M (1994). Criteria for assessing interpretive validity in qualitative research. In N. K. Denzin, & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 485–499). Sage.

Angen, M. J. (2000). Evaluating interpretive inquiry: Reviewing the validity debate and opening the dialogue. *Qualitative Health Research, 10*(3), 378–395. 10.1177/104973200129118516.

Anthony, L. (2014). *AntConc (Version 3.4.3) [Computer Software]*. Tokyo, Japan: Waseda University Available from http://www.laurenceanthony.net/ .

Bankier, J. (2022). Socialization into English academic writing practices through out-of-class -interaction in individual networks of practice. *Journal of Second Language Writing, 56*, Article 100889.
Barkhuizen, G. (Ed.). (2013). *Narrative research in applied linguistics*. Cambridge University Press.
Bazeley, P. (2013). *Qualitative data analysis: practical strategies*. Sage.
Benson, P., Chik, A., Gao, X., Huang, J., & Wang, W. (2009). Qualitative research in language teaching and learning journals, 1997–2006. *The Modern Language Journal, 93*(1), 79–90. 10.1111/j.1540-4781.2009.00829.x.
Brice, C. (2014). Coding data in qualitative research on L2 writing: Issues and implications. In P. K. Matsuda, & T. Silva (Eds.), *Second language writing research* (pp. 159–176). Routledge.
Britton, E. R., & Leonard, R. L. (2020). The social justice potential of critical reflection and critical language awareness pedagogies for L2 writers. *Journal of second language writing, 50*, Article 100776.
Bruner, J. (2010). Narrative, culture, and mind. In D. Schiffrin, A. de Fina, & A. Nylund (Eds.), *Telling stories; Language, narrative, and social life* (pp. 45–49). Georgetown University Press.
Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*(2), 81–105. 10.1037/h0046016.
Canagarajah, A. S. (2015). "Blessed in my own way:" Pedagogical affordances for dialogical voice construction in multilingual student writing. *Journal of Second Language Writing, 27*, 122–139.
Canagarajah, S. (2016). TESOL as a professional community: a half-century of pedagogy, research, and theory. *TESOL Quarterly, 50*(1), 7–41. 10.1002/tesq.275.
Casanave, C. P. (2005). Uses of narrative in L2 writing research. In P. K. Matsuda, & T. Silva (Eds.), *Second language writing research: Perspectives on the process of knowledge construction* (pp. 17–32). Lawrence Erlbaum Associates.
Charmaz, K. (2006). *Constructing grounded theory*. Sage.
Cheng, A. (2011). Language features as the pathways to genre: Students' attention to non-prototypical features and its implications. *Journal of Second Language Writing, 20*(1), 69–82.
Costino, K. A., & Hyon, S. (2007). "A Class for Students Like Me": Reconsidering relationships among identity labels, residency status, and students' preferences for mainstream or multilingual composition. *Journal of Second Language Writing, 16*(2), 63–81.
Creswell, J. W. (1998). *Qualitative inquiry and research design: Choosing among five traditions*. Sage.
Creswell, J. W. (2003). *Research design: Qualitative, quantitative, and mixed methods approach* (2nd ed.). Sage.
Creswell, J. W. (2007). *Qualitative inquiry and research design*. Sage.
Creswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory into practice, 39*(3), 124–130. 10.1207/s15430421tip3903_2.
Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). Sage.
Cutcliffe, J. R., & McKenna, H. P. (1999). Establishing the credibility of qualitative research findings: The plot thickens. *Journal of Advanced Nursing, 30*(2), 374–380. 10.1046/j.1365-2648.1999.01090.x.
Cypress, B. S. (2017). Rigor or reliability and validity in qualitative research: Perspectives, strategies, reconceptualization, and recommendations. *Dimensions of Critical Care nursing, 36*(4), 253–263. 10.1097/DCC.0000000000000253.
Davis, D., & Dodd, J. (2002). Qualitative research and the question of rigor. *Qualitative Health Research, 12*(2), 279–289. 10.1177/104973202129119793.
Denzin, N. K. (1970). *The research act: A theoretical introduction to sociological methods*. Aldine.
Denzin, N. K. (2015). Triangulation. In G. Ritzer (Ed.), *The Blackwell encyclopedia of sociology* (pp. 5083–5088). John Wiley & Sons.
Denzin, N. K., & Lincoln, Y. S (2003). *The landscape of qualitative research: Theories and issues* (2nd ed.). Sage.
Delyser, D. (2008). Teaching qualitative research. *Journal of Geography in Higher Education, 32*(2), 233–244. 10.1080/03098260701514074.
Duff, P. A. (2008). *Case study research in applied linguistics*. Routledge.
Eisner, E. W. (1991). *The enlightened eye: Qualitative inquiry and the enhancement of educational practice*. Macmillan.
Erlandson, D. A., Harris, E. L., Skipper, B. L., & Allen, S. D. (1993). *Doing naturalistic inquiry: A guide to methods*. Sage.
Flick, U. (2007). *Managing quality in qualitative research*. Sage.
Fujioka, M. (2014). L2 student–US professor interactions through disciplinary writing assignments: An activity theory perspective. *Journal of Second Language Writing, 25*, 40–58.
Garratt, D., & Hodkinson, P. (1998). Can there be criteria for selecting research criteria? A hermeneutical analysis of an inescapable dilemma. *Qualitative Inquiry, 4*, 515–539. 10.1177/107780049800400406.
Geertz, C. (1973). Thick description: Toward an interpretive theory of culture. The interpretation of cultures: Selected essays *(3-30)*. Basic Books.
Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Aldine Transaction.
Glesne, C. (2016). *Becoming qualitative researchers: An introduction* (5th ed.). Pearson Education.
Guba, E., & Lincoln, Y (2005). Paradigmatic controversies, contradictions, and emerging confluences. In N. K. Denzin, & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 191–215). Sage.
Guba, E. G., & Lincoln, Y. S. (1981). Effective Evaluation: Improving the Usefulness of Evaluation Results Through Responsive and Naturalistic Approaches. *Jossey-Bass*.
Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Sage.
Hafner, C. A., & Ho, W. Y. J. (2020). Assessing digital multimodal composing in second language writing: Towards a process-based model. *Journal of Second Language Writing, 47*, Article 100710.
Hammersley, M. (1983). *The ethnography of schooling: Methodological issues*. Nafferton Books.
Han, Y., & Hyland, F. (2015). Exploring learner engagement with written corrective feedback in a Chinese tertiary EFL classroom. *Journal of Second Language Writing, 30*, 31–44. 10.1016/j.jslw.2015.08.002.
Healy, M., & Perry, C. (2000). Comprehensive criteria to judge validity and reliability of qualitative research within the realism paradigm. *Qualitative Market Research, 3*(3), 118–126. 10.1108/13522750010333861.
Hesse-Biber, S. N., & Leavy, P. (2011). *The practice of qualitative research*. Sage.
Hope, K., & Waterman, H. (2003). Praiseworthy pragmatism? Validity and action research. *Journal of Advanced Nursing, 44*(2), 120–127. 10.1046/j.1365-2648.2003.02777.x.
Hyland, F. (1998). The impact of teacher written feedback on individual writers. *Journal of second language writing, 7*(3), 255–286.
Hyland, F., & Hyland, K. (2001). Sugaring the pill: Praise and criticism in written feedback. *Journal of Second Language Writing, 10*(3), 185–212. 10.1016/S1060-3743(01)00038-8.
Johnson, R. B. (1997). Examining the validity structure of qualitative research. *Education, 118*, 282–292.
Jiang, L. (2018). Digital multimodal composing and investment change in learners' writing in English as a foreign language. *Journal of Second Language Writing, 40*, 60–72.
Johnson, R. B., & Christensen, L. (2019). Educational Research: Quantitative, Qualitative, and Mixed Approaches. Sage.
Jones, S. R., Torres, V., & Arminio, J. (2006). *Negotiating the complexities of qualitative research in higher education: Fundamental elements and issues*. Routledge.
Junqueira, L., & Payant, C. (2015). I just want to do it right, but it's so hard": A novice teacher's written feedback beliefs and practices. *Journal of Second Language Writing, 27*, 19–36.
Kibler, A. (2010). Writing through two languages: First language expertise in a language minority classroom. *Journal of Second Language Writing, 19*(3), 121–142.
Kirk, J. L., & Miller, M. (1986). *Reliability and validity in qualitative research*. Sage.
Kvale, S. (2007). *Doing interviews*. Sage.
Lankshear, C., & Knobel, M. (2004). *A handbook for teacher research*. McGraw Hill.
Lather, P. (1991). *Getting smart: Feminist research and pedagogy with/in the postmodern*. Routledge.

Lazaraton, A. (2003). Evaluative criteria for qualitative research in applied linguistics: Whose criteria and whose research? *The Modern Language Journal, 87*, 1–12. 10.1111/1540-4781.00175.
Lazaraton, A. (1995). Qualitative research in Applied Linguistics: A progress report. *TESOL Quarterly, 29*(3), 455–472. 10.2307/3588071.
Lee, I. (2010). Writing teacher education and teacher learning: Testimonies of four EFL teachers. *Journal of Second Language Writing, 19*(3), 143–157.
Lee, I. (2013). Becoming a writing teacher: Using "identity" as an analytic lens to understand EFL writing teachers' development. *Journal of Second Language Writing, 22*(3), 330–345.
Lee, G., & Schallert, D. L. (2008). Meeting in the margins: Effects of the teacher–student relationship on revision processes of EFL college students taking a composition course. *Journal of Second Language Writing, 17*(3), 165–182. 10.1016/j.jslw.2007.11.002.
Lee, I., & Yuan, R. E. (2021). Understanding L2 writing teacher expertise. *Journal of Second Language Writing, 52*, Article 100755. 10.1016/j.jslw.2020.100755.
Legewie, H. (1987). Interpretation und validierung biographischer interviews. In G. Jüttemann, & H. Thomae (Eds.), *Biographie und psychologie* (pp. 138–150). Springer.
Lewis, J. (2009). Redefining qualitative methods: Believability in the fifth moment. *International Journal of Qualitative Methods, 8*(2), 1–14. 10.1177/160940690900800201.
Li, M., & Kim, D. (2016). One wiki, two groups: Dynamic interactions across ESL collaborative writing tasks. *Journal of second language writing, 31*, 25–42.
Lim, J., & Polio, C. (2020). Multimodal assignments in higher education: Implications for multimodal writing tasks for L2 writers. *Journal of Second Language Writing, 47*, Article 100713.
Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Sage.
Lincoln, Y. S., Lynham, S. A., & Guba, E. G (2011). Paradigmatic controversies, contradictions, and emerging confluences. In N. K. Denzin, & Y. S. Lincoln (Eds.), *The SAGE handbook of qualitative research* (pp. 97–128). Sage.
Liu, P. H. E., & Tannacito, D. J. (2013). Resistance by L2 writers: The role of racial and language ideology in imagined community and identity investment. *Journal of Second Language Writing, 22*(4), 355–373.
Liu, Y. (2011). Power perceptions and negotiations in a cross-national email writing activity. *Journal of Second Language Writing, 20*(4), 257–270.
Long, T., & Johnson, M. (2000). Rigour, reliability and validity research. *Clinical Effectiveness in Nursing, 4*(1), 30–37. 10.1054/cein.2000.0106.
Luo, N., & Hyland, K. (2016). Chinese academics writing for publication: English teachers as text mediators. *Journal of Second Language Writing, 33*, 43–55.
Maxwell, J. A. (1992). Understanding and validity in qualitative research. *Harvard Educational Review, 62*, 279–300. 10.4135/9781412986274.n2.
Maxwell, J. A. (1996). *Qualitative Research Design: An Interactive Approach*. Sage.
Merriam, S. B. (1998). *Qualitative research and case study applications in education*. Jossey-Bass/Wiley.
Merriam, S. B. (2009). *Qualitative research: A guide to design and implementation*. Jossey-Bass/Wiley.
Merriam, S. B., & Grenier, R. S. (2019). *Qualitative research in practice: Examples for discussion and analysis*. Jossey-Bass/Wiley.
Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Sage.
Mishler, E. G. (1990). Validation in inquiry-guided research: The role of exemplars in narrative studies. *Harvard Educational Review, 60*(4), 415–442. 10.17763/haer.60.4.n4405243p6635752.
Morse, J. M. (2015). Critical analysis of strategies for determining rigor in qualitative inquiry. *Qualitative Health Research, 25*(9), 1212–1222. 10.1177/1049732315588501.
Morse, J. M., Barrett, M., Mayan, M., Olson, K., & Spiers, J. (2002). Verification strategies for establishing reliability and validity in qualitative research. *International Journal of Qualitative Methods, 1*(2), 13–22. 10.1177/160940690200100202.
Negretti, R., & Kuteeva, M. (2011). Fostering metacognitive genre awareness in L2 academic reading and writing: A case study of pre-service English teachers. *Journal of Second Language Writing, 20*(2), 95–110. 10.1016/j.jslw.2011.02.002.
Negretti, R., & McGrath, L. (2018). Scaffolding genre knowledge and metacognition: Insights from an L2 doctoral research writing course. *Journal of Second Language Writing, 40*, 12–31.
Neuman, W. L. (2006). *Social research methods qualitative and quantitative approach* (6th ed.). Pearson.
Olmos-Vega, F. M., Stalmeijer, R. E., Varpio, L., & Kahlke, R. (2022). A practical guide to reflexivity in qualitative research: AMEE Guide No. 149. *Medical Teacher*, 1–11. 10.1080/0142159X.2022.2057287.
Parks, S. (2000). Professional writing and the role of incidental collaboration: Evidence from a medical setting. *Journal of Second Language Writing, 9*(2), 101–122.
Patton, M. Q. (2015). *Qualitative research and evaluation methods* (4th ed.). Sage.
Pelaez-Morales, C. (2017). Short communication: L2 writing scholarship in JSLW: An updated report of research published between 1992-2015. *Journal of Second Language Writing, 38*, 9–19. 10.1016/j.jslw.2017.09.001.
Pinnegar, S., & Hamilton, M. L. (2009). *Self-study of practice as a genre of qualitative research*. Springer.
Porte, G., & Richards, K. (2012). Focus article: Replication in second language writing research. *Journal of Second Language Writing, 21*, 284–293. 10.1016/j.jslw.2012.05.002.
Prior, P. (1995). Tracing authoritative and internally persuasive discourses: A case study of response, revision, and disciplinary enculturation. *Research in the Teaching of English, 29*(3), 288–325. 10.1080/14623943.2021.1952972.
Ramanathan, V., & Atkinson, D. (1999). Ethnographic approaches and methods in L2 writing research: A critical guide and review. *Applied Linguistics, 20*(1), 44–70. 10.1093/applin/20.1.44.
Riazi, A., Ghanbar, H., & Rezvani, R. (2023). Qualitative data coding and analysis: A systematic review of the papers published in the Journal of Second Language Writing. *Iranian Journal of Language Teaching Research, 11*(1), 25–47. 10.30466/IJLTR.2023.121271.
Riazi, M., Shi, L., & Haggerty, J. (2018). Analysis of the empirical research in the journal of second language writing at its 25th year (1992–2016). *Journal of Second Language Writing, 41*, 41–54. 10.1016/j.jslw.2018.07.002.
Richards, K. (2009). Trends in qualitative research in language teaching since 2000. *Language Teaching, 42*, 147–180. 10.1017/S0261444808005612.
Richards, K. (2003). *Qualitative inquiry in TESOL*. Springer.
Richardson, L., Pierre, St., & A, E (2005). Writing: A method of inquiry. In N. K. Denzin, & Y. S. Lincoln (Eds.), *The SAGE handbook of qualitative research* (pp. 959–978). Sage.
Robson, C., & McCartan, C. (2016). *Real world research* (4th ed.). Wiley.
Rolfe, G. (2006). Validity, trustworthiness and rigour: Quality and the idea of qualitative research. *Methodological Issues in Nursing Research, 53*, 304–310. 10.1111/j.1365-2648.2006.03727.x.
Sánchez-Martín, C., & Seloni, L. (2019). Transdisciplinary becoming as a gendered activity: A reflexive study of dissertation mentoring. *Journal of Second Language Writing, 43*, 24–35. 10.1016/j.jslw.2018.06.006.
Sandelowski, M. (1993). Theory unmasked: The uses and guises of theory in qualitative research. *Research in Nursing and Health, 16*(3), 213–218. 10.1002/nur.4770160308.
Sandelowski, M., & Barroso, J. (2002). Reading qualitative studies. *International Journal of Qualitative Methods, 1*(1), 74–108. 10.1177/160940690200100107.
Satake, Y. (2020). How error types affect the accuracy of L2 error correction with corpus use. *Journal of Second Language Writing, 50*, Article 100757.
Schwandt, T. (1996). Farewell to criteriology. *Qualitative Inquiry, 2*, 58–72. 10.1177/107780049600200109.
(2007). Quality in qualitative research. In C. Seale, C. Seale, G. Gobo, J. Gubrium, & D. Silverman (Eds.), *Qualitative Research Practice* (pp. 379–389). Sage.
Seloni, L. (2014). I'm an artist and a scholar who is trying to find a middle point": A textographic analysis of a Colombian art historian's thesis writing. *Journal of Second Language Writing, 25*, 79–99.
Sengupta, S. (1999). Rhetorical consciousness raising in the L2 reading classroom. *Journal of Second Language Writing, 8*(3), 291–319.
Shenten, A. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information, 22*, 63–75. 10.3233/EFI-2004-22201.
Shin, D. S., Cimasko, T., & Yi, Y. (2020). Development of metalanguage for multimodal composing: A case study of an L2 writer's design of multimedia texts. *Journal of Second Language Writing, 47*, Article 100714.

Silva, T., & Leki, I. (2004). Family matters: The influence of applied linguistics and composition studies on second language writing studies—past, present, and future. *Modern Language Journal, 88*(1), 1–13. 10.1111/j.0026-7902.2004.00215.x.
Silverman, D. (2014). *Interpreting qualitative data. Methods for analyzing talk, text and interaction*. Sage.
Smith, J. (1997). The stories educational researchers tell about themselves. *Educational Researcher, 26*, 4–11. 10.3102/0013189x026005004.
Smith, B. E., Pacheco, M. B., & Almeida, C. (2017). Multimodal codemeshing: Bilingual adolescents' processes composing across modes and languages. *Journal of Second Language Writing, 36*, 6–22. 10.1016/j.jslw.2017.04.001.
Sparkes, A. C., & Smith, B. (2009). Judging the quality of qualitative inquiry: Criteriology and relativism in action. *Psychology of Sport and Exercise, 10*, 491–497. 10.1016/j.psychsport.2009.02.006.
Staffileno, B. A., Murphy, M. P., & Buchholz, S. W. (2021). *Research for advanced practice nurses: From evidence to practice* (4th ed.). Springer.
Stallings, W. M. (1995). Research news and comment: Confessions of a quantitative educational researcher trying to teach qualitative research. *Educational Researcher, 24*, 31–32. 10.3102/0013189X024003031.
Starfield, S. (2013). Researcher reflexivity. In C. A. Chappelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1–7). Wiley-Blackwell.
Tardy, C. M. (2021). What is (and could be) thick description in academic writing research?. In C. M. Tardy (Ed.), *Ethnographies of academic writing research: Theory, methods, and interpretation* (pp. 22–38). John Benjamin.
Tracy, S. (2010). Qualitative quality: Eight "Big-Tent" criteria for excellent qualitative research. *Qualitative Inquiry, 16*, 837–851. 10.1177/1077800410383121.
Wan, W. (2014). Constructing and developing ESL students' beliefs about writing through metaphor: An exploratory study. *Journal of Second Language Writing, 23*, 53–73.
Wang, X. (2020). Translation as Mobile practice. *Journal of Second Language Writing, 49*, Article 100733.
Webb, J., Schirato, T., & Danaher, G. (2002). *Understanding Bourdieu*. Sage.
Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. (1966). *Unobtrusive measures: Nonreactive research in the social sciences*. Rand Mcnally.
Weigle, S. C., & Nelson, G. L. (2004). Novice tutors and their ESL tutees: Three case studies of tutor roles and perceptions of tutorial success. *Journal of Second Language Writing, 13*(3), 203–225.
Wette, R. (2010). Evaluating student learning in a university-level EAP unit on writing using sources. *Journal of Second Language Writing, 19*(3), 158–177.
Whittemore, R., Chase, S. K., & Mandle, C. L. (2001). Validity in qualitative research. *Qualitative Health Research, 11*, 522–537. 10.1177/104973201129119299.
Wolcott, H. F. (1990). On seeking-and-rejecting validity in qualitative research. In E. W. Eisner, & A. Peshkin (Eds.), *Qualitative inquiry in education: The continuing debate* (pp. 121–152). Teachers College Press.
Wolcott, H. F. (1994). *Transforming qualitative data: Description, analysis, and interpretation*. Sage.
Worden, D. (2018). Balancing stability and flexibility in genre-based writing instruction: A case study of a novice L2 writing teacher. *Journal of Second Language Writing, 42*, 44–57.
Wu, Z. (2020). Tracing EFL writers' digital literacy practices in asynchronous communication: A multiple-case study. *Journal of Second Language Writing, 50*, Article 100754.
Yi, Y. (2007). Engaging literacy: A biliterate student's composing practices beyond school. *Journal of Second Language Writing, 16*(1), 23–39.
Yi, Y. (2010). Adolescent multilingual writers' transitions across in-and out-of-school writing contexts. *Journal of Second Language Writing, 19*(1), 17–32.
Yoon, C. (2016). Individual differences in online reference resource consultation: Case studies of Korean ESL graduate writers. *Journal of Second Language Writing, 32*, 67–80.
Yu, L. (2020). Investigating L2 writing through tutor-tutee interactions and revisions: A case study of a multilingual writer in EAP tutorials. *Journal of Second Language Writing, 48*, Article 100709.
Yu, S. (2021). Feedback-giving practice for L2 writing teachers: Friend or foe? *Journal of Second Language Writing, 52*, Article 100798.
Yu, S., Xu, H., Jiang, L., & Chan, I. K. I. (2020). Understanding Macau novice secondary teachers' beliefs and practices of EFL writing instruction: A complexity theory perspective. *Journal of Second Language Writing, 48*, Article 100728.