# Enhancing Trustworthiness in Qualitative Research: Integrating Quantitative Validation in English Language Teaching Studies

## Kyle Perkins

Professor, Department of Teaching and Learning, Florida International University.
*perkinsk@fiu.edu*

| Article info | Abstract |
|---|---|
| | I propose that quantitative research designs should be used to augment and to complement the validation of qualitative research in a mixed methods design for the following reasons. Combining quantitative and qualitative data allows the phenomenon under study to be explored more deeply and allows for perspectives that may otherwise be overlooked. Moreover, qualitative and quantitative research should meet credibility criteria which have different foci. Qualitative research quality criteria are in flux and are situated within local disciplinary contexts. As a result, there is an absence of universal acceptance and adoption of criteria for judging qualitative study quality. Qualitative and quantitative designs complement each other, and both approaches are necessary.<br><br>***Keywords:*** data coding and analysis, exploratory factor analysis, quality measures' criteria, qualitative research methodologies, quantitative research methodologies |

# 1. Introduction

This is a position paper in which I propose that quantitative designs should be used in conjunction with qualitative designs which will allow researchers to address a greater spectrum of research problems and to address any blind spots in current research. Mixed methods research includes collecting and analyzing both qualitative and quantitative data in a single study to understand a research problem (Busetto et al., 2020; Fitzpatrick et al., 2004). Researchers also point out that combining quantitative and qualitive allows the phenomenon under study to be explored more deeply and allows for different perspectives that may otherwise be overlooked (Carter et al. 2014). I think that qualitative and quantitative designs complement each other, and both approaches are necessary. The thesis of this paper is that quantitative procedures should be used to augment the validation process of qualitative research.

This paper consists of five major sections: methods of qualitative data collection, qualitative data analysis, qualitative validation criteria, pre-analysis quantitative data screening, and quantitative data analyses for validation.

# 2. Literature Review

## 2.1. Methods of Qualitative Data Collection

Document study, sampling, participant observation, semi-structured interviews, and focus groups are potential sources of qualitative data.

Document study, also known as document analysis, involves the review of written materials such as annual reports, archives, diaries, letters, guidelines, and policy guidelines (Busetto et al., 2020).

Qualitative researchers use snowball sampling and purposive or deliberate sampling. Snowball sampling involves having participants identify others who have knowledge which is relevant to the research question, which generally leads to a homogeneous sample. On the other hand, researchers may specify or pre-define the types of participants whom they need to include to cover all relevant variations based on theory, a review of published research, or previous findings and experience (Fossey et al., 2002). Authorities in the field are aware that sampling, data collection, analysis, and interpretation are cyclically and iteratively related to each other rather than a one-after-the-other stepwise approach. More data collection may be necessary to identify missing variants in the current sample. "This process continues until no new (relevant) information can be found and further sampling becomes redundant—which is called saturation" (Busetto et al., 2020, pp. 6-7).

Participant observations provide more insights into actual behavior in a certain setting than opinions or reported behavior. Two major operations are entailed by participant observation: the observers taking notes on everything or certain parts which were predetermined and transcribing the field notes into observation protocols which provide for the potential discovery of previously

overlooked topics which are relevant to the research questions (Busetto et al., 2020).

During semi-structured interviews, observers or the researchers themselves use an interview guide which may be a topic guide and open-ended questions. The purpose of these interviews is to gain insight into a person's affective domain, e.g., motivations, opinions, and subjective experiences (Busetto et al., 2020).

A focus group typically consists of six to eight persons who are led by an experienced moderator who follows a script or a topic guide. The intended purpose of a focus group is to explore a person's expertise related to the research question and experiences. Focus groups are considered to be useful for assembling homogenous groups.

### 2.2. Data Analysis

Coding, grounded theory, phenomenological analysis, thematic analysis, and computer-assisted analysis are important concepts in qualitative research data analysis.

Researchers transcribe field notes, audio recordings, observations, interviews and focus groups' data into protocols and transcripts. Researchers use qualitative data management software such as NVivo, MaxQDA, and Atlas.ti to organize, code, and manage the data to identify patterns, themes, and concepts in the data. The transcripts and protocols are coded with short descriptions. The next steps are to group, summarize, and categorize the codes. Ideally, at least two researchers should perform the coding process. Before the coding process begins, the coders should define a common approach, develop a usable coding list, and concur on a common meaning of individual codes. See Busetto et al. (2020) for additional details.

In a study of qualitative data coding and analysis, Riazi et al. (2023) point out that both qualitative and quantitative research need to meet credibility criteria which have different foci. In quantitative research, reliability, validity, generalizability, and objectivity are related to the accuracy or consistency of the measures utilized in the reported research. On the other hand, qualitative research focuses on trustworthiness the principles for which, according to Lincoln and Guba (1985) are credibility, transferability, dependability, and confirmability.

Researchers who elect to use quantitative approaches to augment qualitative analysis validation may construct a data matrix for input into various quantitative analyses to be presented later in this paper. "A data matrix is an organization of raw scores or data, where the rows represent participants, or cases and the columns represent variables" (Mertler & Vannatta, 2013, p. 3). The rows could represent the concepts, the themes, and/or the variables identified in the study. If the researcher chose a binary table, a participant who

endorsed a concept, a theme, or a variable in a column would be coded as a 1. A participant who did not endorse a variable, a concept, or a theme could be coded as 0. SPSS and an Excel spreadsheet could be used to record the data for quantitative analyses.

The purpose of grounded theory is to construct theory or to explicate theory from the data. It combines coding procedures and sampling procedures. A more complete description follows

> Grounded theory is a common approach to the analysis of qualitative data. It makes use of an iterative process, alternating inductive and deductive methods to explore for     patterns and confirm findings. An inductive, open approach is used as the first step in examining the data. The evaluation then builds some tentative hypotheses or theories about the data and then changes to a deductive approach, examining the data to see if it confirms the theories. This process of back-and-forth inductive and deductive examination of the data is used to allow the evaluator to become 'grounded' in the theory implicit in the qualitative data. (Fitzpatrick et al., 2004, p. 362)

Phenomenological analytic techniques involve within and between analyses of meaning in data. There are two levels of analysis: a search for recurrent themes of meaning within the data for each participant and a search for common themes and areas of divergence across participants (Fossey et al., 2002).

Thematic analysis involves "a progressive process of classifying, comparing, grouping and refining groupings of text segments to create and then clarify the definition of categories or themes, within the data" (Fossey et al., 2002, pp. 728-729). In sum, researchers derive categories directly from the data, not from a priori theory.

## 2.3. Validation Criteria

Tracy and Hinrichs (2017) made the point that we need criteria for qualitative research because "values for quality, like all social knowledge, are ever-changing and situated within local disciplinary contexts and current scholarly conversations" (p. 1). Tracy's (2010) "Eight 'big-tent' criteria for excellent qualitative research" indicate that high-quality qualitative research is marked by: (1) worthy topic, (2) rich rigor, (3) sincerity, (4) credibility, (5) resonance, (6) significant contribution, (7) ethics, and (8) meaningful coherence. "Each criterion of quality can be approached via a variety of means, paths, or crafts—the combination of which depends on the specific researcher, context, theoretical affiliation, and project" (Tracy & Hinrichs, 2017, p. 2).

Riazi et al. (2023) reported on what qualitative terminology (i.e., trustworthiness, credibility, dependability, transferability, confirmability) or conventional terminology (i.e., validity, etc.) that were used in published

articles in the Journal of Second Language Writing to refer to trustworthiness concerns. One reason for research of this nature is a move away from the traditional terminologies associated with quantitative research methodologies. A goal of the Riazi et al.'s study was "to see what sources informed the methodology of the studies as far as quality is concerned…Our results broadly indicated insufficient and inconsistent attention to the quality criteria (trustworthiness and its components) in L2 writing qualitative studies" (p. 9). Classical quality terminologies like validity seem to be lingering in the field and in many L2 scholars' studies.

Mirhosseini and Pearson (2025) offered meaningful insight into the lack of universal acceptance and adoption of criteria for judging qualitative study quality and credibility. "Qualitative research is characterized by its social constructivist epistemological essence realized through various methodological traditions and approaches. The issue of quality in qualitive inquiry has, therefore, been interconnected with diverse understandings of such epistemological and methodological perspectives (p. 100). Calcifying good practice into immovable criteria is considered fundamentally at odds with the guiding philosophy of qualitative research, which stressed creativity, exploration, conceptual flexibility, and freedom of spirit" (p. 100). As a result, the list of criteria for judging qualitative study quality grows and includes a sampler of the following: quantitative (reliability, validity, generalizability, objectivity) and qualitative (trustworthiness, credibility, transferability, dependability, confirmability, replicability, reflexivity, perspicacity, emotional vulnerability, resonance).

A literature review of validation criteria contains many suggestions, some of which are adherence to checklists, reflexivity, piloting, co-coding, member checking, stakeholder involvement, rival or competing theories, negative cases, triangulation, credibility, transferability, dependability, confirmability, and the FAIR principles for research data management. A sampler of suggested validation criteria follows.

Various methods are used to confirm the findings from qualitative analysis, including

- Weighting the evidence to give stronger data more weight;
- Looking for and examining rival or competing themes;
- Trying to find negative cases that do not conform to the theory or hypotheses and considering whether they approve or refine the theory;
- Considering an attempt to replicate the findings with another case;
- Looking for triangulation of findings across different methods;
- Using another evaluator or a selected informant to provide feedback

on your analysis and to act as a devil's advocate against your theory. (Fitzpatrick et al., 2004, pp. 362-363)

Other researchers have suggested criteria for validation, some of which overlap and appear in different lists.  For example, Carter et al. (2014) offer the following strategies to ensure data dependability and credibility such as debriefing, member checking, triangulation or use of a reflective journal.  In a similar vein, Busetto et al. (2020) offer for consideration adherence to checklists, reflexivity, sampling strategy, piloting, co-coding, member checking, and stakeholder involvement as evaluation criteria.  And finally, Barker et al. (2022) provide research data guidelines to ensure that data are findable, accessible, inter-operative, and reusable. Researchers should also adhere to the Standards for Reporting Qualitative Research to ensure that all items are relevant for the research question under study which will contribute to transparency.  Many reports on validation criteria also mention credibility, transferability, confirmability, reliability, repeatability, and reproducibility.

Member checking, reflexivity, and triangulation are three recurrent themes in the qualitative research literature.

Member checking, also known as respondent validation, refers to returning the data and/or results to the participants so that they can check for accuracy, which should improve the credibility, validity, and transferability of a study (Birt et al., 2016, p. 1802).

Reflexivity refers to the relationship between the researcher and what is being studied, because the researcher as a person cannot be isolated from the research process. Reflexivity accounts for how subjectivity can shape a researcher's inquiry and research questions in term of how the researcher's role influenced the construction of the research findings.  A researcher must control the researcher's subjectivity and the context, because they may influence the research process.

Triangulation refers to the use of multiple methods or data sources in qualitative research to develop a comprehensive understanding of the phenomena related to the research question.  Triangulation offers insights into the qualitative research process such as a comparison and convergence of perspectives to identify corroborating and dissenting accounts, an examination of many aspects of the research issue, a research strategy to test validity through convergence of information from different resources, and the development of a comprehensive understanding of the phenomena related to the research question (Anderson et al., 1991; Carter et al., 2014; Fossey, Harvey et al., 2002).

There are four types of triangulation:

Method triangulation involves the use of multiple methods of data collection about the same phenomenon. This type of triangulation may include interviews, observation, and field notes. Investigator

triangulation involves the participation of two or more researchers in the same study to provide multiple observations and conclusions. This type of triangulation can bring both confirmation of findings and different perspectives, adding breadth to the phenomenon of interest. Theory triangulation uses different theories to analyze and interpret data. With this type of triangulation, different theories or hypotheses can assist the researcher in supporting or refuting findings. Data source triangulation involves the collection of data from different types of people, including individuals, groups, families, and communities to gain multiple perspectives and validation of data. (Carter et al., 2014, p. 545).

In summary, triangulation allows researchers to cross-verify information, identify discrepancies, and provide a more comprehensive understanding of the phenomenon under study. By leveraging different perspectives, it minimizes biases, reduces the likelihood of errors, strengthens the reliability of the results, enhances credibility, provides comprehensive insights, and improves rigor (Bryman, 2016; Cresswell & Plano Clark, 2010; Denzin, 1978; Flick, 2018; Hassan, 2024; Patton, 2015).

## 3. Method
### 3.1. Quantitative Analysis for Qualitative Research

In the introduction, I stated that, in my opinion, qualitative and quantitative designs complement each other, and both approaches are necessary in order to produce quality research. In addition, the thesis statement of the paper is that quantitative procedures can be used to augment the validation process of qualitative research. Reasons for combining methods include "triangulation for corroboration of findings, complementary for illustration and clarification of results, expansion to extend the breadth and range of the study, explanation of (unexpected) results generated with one method with the help of another, or offsetting the weakness of one method with the strength of another" (Busetto et al., 2020, p. 5). I provided a brief description of the steps involved in conducting qualitative research in order to provide the context for how and where quantitative analyses can be utilized to augment the validation process. Pre-analysis quantitative data screening and quantitative analyses for validation are presented below.

### 3.2. Pre-analysis Quantitative Data Screening

Before submitting data for analysis, the researcher should check the data for accuracy, identify missing data, identify outliers, i.e., extreme values, determine that the data set has a normal distribution, using skewness and kurtosis coefficients, and determine whether the data set manifests linearity and homoscedasticity. If a researcher decides that these constructs are important and need to be addressed, see Rezvani et al., (2024), Mertler and

Vannatta (2013), and Tabachnick and Fidell (2017) for alternative methods for handling these data

## 4. Quantitative Analyses for Validation

Quantitative procedures for establishing agreement, concordance, and correlation or the lack thereof, exploratory factor analysis, and the Guttman scale follow.

Some quantitative procedures for establishing agreement, concordance, and correlation or the lack thereof among concepts, variables, coders, participants, sampling strategies, observers, protocols, recurrent and common themes, categories, data sources, multiple methods, and corroborating and dissenting accounts include the following.

The Pearson product-moment coefficient is a parametric measure/estimate of the linear correlation between two sets of data; it is a measure of the strength of the linear relationship between two variables. For example, a significant correlation coefficient of 0.50 or higher for two researchers' coding of the same data would be a decent estimate of inter-rater reliability.

Spearman's rho is a non-parametric measure of rank order correlation between the rankings of two variables, which can be used to analyze ordinal level as well as continuous level data.

Analysis of variance compares variances across the means of different groups, a check if the means of two or more groups are significantly different from each other.

The Mann-Whitney U test is the nonparametric alternative to the independent sample t-test to compare two sample means that come from the same population and is used to test whether two sample means are equal or not equal.

The Student's t-test is used to determine whether the difference between the response of two groups is statistically significant or not. The procedure is also used to determine if there is a significant difference between the means of two groups and how they are related (or not).

The nonparametric Kruskal-Wallis test is used to compare two or more independent samples of equal or different sample size to determine if there are statistically significant differences between two or more groups of an independent variable on a continuous or ordinal dependent variable.

Kendall's tau is a nonparametric measure of the strength and direction of association that exists between two variables measured on at least an ordinal scale.

Cohen's kappa is a widely used quantitative measure of reliability for two raters who are rating the same thing. It has the advantage of correcting for how often the raters may agree by chance.

Regression analysis is a set of statistical methods that are used to estimate the relationships between a dependent variable and one or more independent variables. As such, these methods are used to assess the strength of the relationship between variables.

The chi-square test is a hypothesis testing method, which involves checking to ascertain if observed frequencies in one or more categories match the expected frequencies. The chi-square test can also provide estimates of goodness of fit and of test of independence.

The Bland-Altman plot is a useful display of the relationship between two paired variables using the same scale. It is identical to a Tukey mean-difference plot (Bland & Altman, 1986).

The Tukey mean difference plot is an adaption of the quantile-quantile (Q-Q) plot, which is a graphical data analysis technic for the comparison of the distributions of two data sets. For further information on these topics, see Hinkle et al. (1994), Little (2013), Roscoe (1975), Sprinthall (2003), and Thorkildsen (2005).

## 4.1. Exploratory Factor Analysis (EFA)

EFA is a data reduction technique that reduces a data set into a smaller set of variables. It also identifies the structure of the relationship between a variable and the respondents. In Mertler and Vannatta's words (2013), Factor analysis is a procedure used to determine the extent to which measurements overlap—that is, shared variance—exists among a set of variables. Its underlying purpose is to determine if measures for different variables are, in fact, measuring something in common" (p. 237).

Rezvani et al. (2024) described and evaluated methodological issues in EFA research practices and examined the relationship between study features and outcomes. EFA has been described as an exploratory method that is used to generate theory because researchers using EFA have no hypotheses about the number of factors that underlie the input data nor about the factor loadings. EFA has also been described as a method of data reduction, resulting in an economical description of correlational data. In summary, EFA is used to uncover the underlying structure of a large set of variables and to identify the underlying relationships between measured variables.

Rezvani et al. (2024) discuss statistical assumptions and practical considerations in EFA such as the ratio of participants to variables, missing data, normality of distribution, linearity, outliers, multicollinearity and singularity, extraction methods such as principal component analysis, principal axis factoring, alpha factoring, image factoring, maximum likelihood, and orthogonal and oblique rotation methods. The study concludes with 18 recommendations which are aligned with the best EFA practices for using EFA in L2 research.

There are two types or methods of EFA. The R-type factor analysis focuses on associations between variable measurements to provide an understanding of how the variables are related and group together. The Q-type factor analysis focuses on how persons group together using a correlation of persons across a range of variables. The R-type factor analysis might be useful in examining whether codes, themes, researchers, and the observers group together and are related. This approach might be useful in the identification and examination of reflexivity. Because qualitative researchers do not begin data analysis with set categories, perhaps factor analysis might be useful in helping to suggest patterns or themes to reduce subjectivity. The Q-type factor analysis might be useful in determining if participants, observers, and coders group together.

If a researcher wishes to identify the minimum number of factors which explain the maximum amount of variance in a variable, then principal component analysis is utilized. If, on the other hand, the researcher doesn't know the nature of the factor to be extracted, then common factor analysis might be considered.

Factor analysis researchers use eigenvalues, scree test criteria, and the percentage of variance explained in the selection of factors to be retained.

## 4.2. Guttman Scale

A Guttman scale is a deterministic model which uses the total score to predict which items were correctly and incorrectly answered. In the case of qualitative data, as mentioned earlier, a binary data set would record those persons who endorsed items, coded as 1, and those who did not endorse items as 0. A Guttman scale, also known as a scalogram, orders persons from the most able (Person A) to the least able (Person K), and the items are ordered from the least difficult (Item 1) to the most difficult item (Item 10), shown in the following perfect Guttman scale (Table 1).

Bond eta al. (2021) describe the Guttman scale as follows:

A Guttman scale consists of a unidimensional set of items that can be ranked in order of difficulty, such that any person's entire set of responses to all items can be determined from that person's total score on the test. For a true Guttman patten, the response pattern for a person scoring 6 on a 10-item Guttman scale will show success on items 1-6 and failure on items 7-10 when those responses are ordered by item difficulty, e.g., 1111110000. (p. 33).

**Table 1**

*An Example of Guttman Scale*

| Items | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Person Score |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Person A | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 10 |
| Person B | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 9 |
| Person C | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 8 |
| Person D | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 7 |
| Person E | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 6 |
| Person F | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 5 |
| Person G | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 4 |
| Person H | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 |
| Person I | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Person J | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Person K | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Item Score | 10 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | |

Although perfect Guttman scales do not exist in the real world, they do offer valuable insights, as noted by Thorkildensen (2005). "The simultaneous stimulus and person techniques offer scaling of items and persons in one administration" (p. 111). "The most obvious advantage of simultaneous stimulus and person scaling is the degree of precision in locating persons on scales" (p. 113). In qualitative research, Guttman scales might be utilized to identify outlying persons and items, which the researcher could investigate to ascertain whether bias, reflexivity, or an error in coding was responsible for the outliers.

## 5. Conclusion

Qualitative research has enjoyed a long history in academic fields such as education, psychology, sociology, social work, anthropology, history, political science and government, communication studies, nursing and psychiatry. It will continue to be used in these fields and in future fields yet to be developed for the academic portfolio. The purpose of this paper is to illustrate how quantitative approaches can be used to augment qualitative analysis validation. I hope that this paper will assist current researchers and graduate students who elect to use qualitative research in their dissertations and in their future research programs.

## Acknowledgements

# References

Anderson, N. J., Bachman, L., Perkins, K., & Cohen, A. (1991). An exploratory study into the construct validity of a reading comprehension test: triangulation of data sources. *Language Testing, 8*(1), 41-66.

Barker, M., et al. (2022). Introducing the FAIR principles for research software. *Scientific Data, 9*, 622.

Birt, L., Scott, S., Cavers, D., Campbell, D., & Walter, F. (2010). Member checking: a tool to enhance trustworthiness or merely a nod to validation? *Health Quality Research, 26*(13), 1802-1816.

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of a clinical measure. *Lancet, 1*, 307-310.

Bond, T. G., Yan, Z., & Heene, M. (2021) *Applying the Rasch model: fundamental measurement in the human sciences* (4th ed.). Routledge.

Bryman, A. (2016). *Social research methods.* Oxford University Press.

Busetto, L., Wick, W. E., & Gumbinger, C. (2020). How to use and assess qualitative research. *Neurological Research and Practice, 2*, 14.

Carter, N., Bryant-Lukosius, D., DiCenso, A., Blythe, J, & Neville, A. J. (2014). The use of triangulation in qualitative research. *Oncology Nursing Forum, 41*(5), 545-547.

Cresswell, J. W., & Plano Clark, V. L. (2017). *Designing and conducting mixed methods research.* Sage Publications.

Denzin, N. K. (1978). *The research act: A theoretical introduction to sociological methods.* McGraw-Hill.

Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2004). *Program evaluation. Alternative approaches and practical guidelines* (3rd ed.). Pearson Education, Inc.

Flick, U. (2018). *An introduction to qualitative research.* Sage Publications.

Fossey, D., Harvey, C., McDermott, F., & Davidson, L. (2002). Understanding and evaluating qualitative research. *Australian and New Zealand Journal of Psychiatry, 36*, 717-732.

Hassan, M. (2024). Triangulation in research-types, methods and guide.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (1994). *Applies statistics for the behavioral sciences* (3rd ed.). Houghton Mifflin Company.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry.* Sage.

Little, T. D. (Ed.) (2013). *The Oxford handbook of quantitative methods, Volume 2 statistical analysis.* Oxford University Press.

Mertler, C. A., & Vannatta, R. A. (2013). *Advanced and multivariate statistical methods* (5th ed.). Pyrczak Publishing.

Mirrosseini, S-A., & Pearson, W. S. (2025). How do language education researchers attend to quality in qualitative studies? *Language Teaching, 58*, 99-115.

Patton, M. Q. (1999). *Enhancing the quality and credibility of quantitative analysis. Sociological methods: A sourcebook.* McGraw-Hill.

Patton, M. Q. (2015). *Qualitative research & evaluation methods: Integrating theory and practice.* Sage Publications.

Perkins, K. (2023). English language assessment in a transnational world. A theory-into-practice approach. In O. Z. Barnawi, M. S. Alharbi, & A. Alzahrani (Eds.). *Transnational assessment practices in the age of metrics* (pp. 41-50). Routledge.

Rezvani, R., Ghanbar, H., & Perkins, K. (2024). Consideration and praxis of exploratory factor analysis: Implications for L2 research. *Teaching English as a Second Language Quarterly, 43*(1), 151-178.

Riazi, M., 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.

Riazi, M., Rezvani, R., & Ghanbar, H. (2023). Trustworthiness in L2 writing research: a review and analysis of qualitative articles in the Journal of Second Language Writing. *Research Methods in Applied Linguistics, 2,* Article e100065.

Roscoe, J. T. (1975). *Fundamental research statistics for the behavioral sciences* (2nd ed.). Holt, Rinehart and Winston, Inc.

Sprinthall, R. C. (2003). *Basic statistical analysis* (7th ed.). Pearson Education Group, Inc.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Allyn & Bacon.

Thorkildsen, T. A. (2005). *Fundamental measurement in applied research.* Pearson Education, Inc.

Tracy, S. A. (2010). Qualitative quality: Eight big-tent criteria for excellent qualitative research. *Qualitative Inquiry, 16*(10), 837-851.

Tracy, S. A., & Hinrichs, M. M. (2017). Big tent criteria for qualitative quality. In J. Matthes, C. S. Davis, & R. Potter (Eds.), *The international encyclopedia of communication research methods,* pp.1-10. John Wiley & Sons, Inc.